

TABLE OF CONTENTS

0. Executive Summary	12
1. Introduction	14
Definition and scope of sequencing.....	14
Historical aspects of sequencing	14
Basics of molecular biology.....	15
DNA	15
<i>DNA polymerases.....</i>	<i>15</i>
<i>Restriction endonucleases</i>	<i>15</i>
<i>DNA methylation.....</i>	<i>16</i>
RNA	17
<i>RNA polymerases.....</i>	<i>17</i>
<i>Non-coding RNAs</i>	<i>17</i>
DNA transcription.....	18
Chromosomes	18
Telomeres.....	18
Mitochondrial DNA.....	19
Genes.....	19
<i>The genetic code.....</i>	<i>20</i>
<i>Gene expression</i>	<i>20</i>
The human genome	20
Variations in the human genome	21
Variations in DNA sequences	21
Single nucleotide polymorphisms	22
Haplotyping.....	22
Complex chromosomal rearrangements	23
Insertions and deletions in the human genome	23
Large scale variation in human genome	24
Variation in copy number in the human genome	24
Structural variations in the human genome	25
Transposons.....	26
<i>Retrotransposon capture sequencing</i>	<i>26</i>
Mapping and sequencing of structural variation from human genomes	27
Impact of sequencing on healthcare	27
2. DNA Sequencing Technologies	30
Introduction.....	30
DNA extraction and sample preparation	31
<i>Apollo 300 System for next generation sequencing.....</i>	<i>31</i>
<i>Electrophoresis-based method</i>	<i>32</i>
<i>Ion OneTouch System</i>	<i>32</i>
<i>Microfluidics-based extraction and sample preparation.....</i>	<i>33</i>
<i>Pressure Cycling Technology.....</i>	<i>33</i>
<i>Selective immobilization of nucleic acids onto magnetic microparticles</i>	<i>33</i>
<i>Targeted and hybridization-based DNA capture.....</i>	<i>33</i>
Sanger-sequencing technology	34
<i>Dye-terminator sequencing</i>	<i>34</i>
<i>Large-scale sequencing.....</i>	<i>35</i>
<i>Automated DNA-sequencing</i>	<i>35</i>
<i>Enhancements of Sanger-sequencing</i>	<i>35</i>
<i>ABI PRISM® 310 Genetic Analyzer.....</i>	<i>36</i>
<i>Life Technologies' 3500 Dx genetic analyzer</i>	<i>36</i>
Limitations of sequencing methods and measures to remedy them	36
<i>High-throughput paired end transcriptome sequencing</i>	<i>37</i>
<i>Long vs short read lengths.....</i>	<i>37</i>
Emerging sequencing technologies.....	37
Chemical DNA sequencing.....	38
Second generation sequencers	39
4300 DNA analyzer	39
Apollo 100	40
Applied Biosystems 3500 series Genetic Analyzer.....	40
"Color blind" approach to DNA sequencing	41
Cyclic array sequencing	41
CEQ™ 8000.....	41
DeepCAGE sequencing.....	41
Electron microscope-based DNA sequencing	42
Genometrica™ sequencer.....	42

GS-FLEX system (Roche/454).....	43
<i>Background of sequencing technology</i>	43
454 sequencing	44
GS Junior System.....	46
IBS sequencing technology	47
Illumina Genome Analyzer System	47
Ion Torrent's sequencing technology	48
MegaBACE 500	49
Microdroplet-based PCR for large-scale targeted sequencing.....	49
Millikan sequencing	50
Multiplex amplification of human DNA sequences	50
Nanoscale sequencing	50
Polonator sequencer	51
RainStorm™ microdroplet technology	52
Sequential DEXAS	52
SOLiD system: sequencing by ligation	52
PCR-based DNA sequencing technologies	53
<i>Bridge amplification PCR system</i>	54
<i>COLD-PCR and sequencing</i>	54
<i>Digital PCR</i>	55
<i>Dual primer emulsion PCR</i>	55
<i>Emulsion PCR</i>	56
<i>Multiplex PCR</i>	56
Nucleic acid sequence-based amplification	56
Microarray-based DNA sequencing technologies	57
Arrayit's® H25K	57
<i>High-throughput array-based resequencing</i>	57
<i>Human exome microarrays</i>	58
<i>Sequencing by hybridization</i>	58
<i>SOLiD-System based ChIP-Sequencing</i>	59
<i>Companies developing whole genome chips/microarrays</i>	60
Next generation sequencing vs microarrays for gene expression profiling	60
Third generation sequencing.....	61
<i>SOLiD4 System</i>	61
<i>SOLiD PI System</i>	62
Detection of single molecules for sequencing	62
<i>DNA sequence by use of nanoparticles</i>	63
<i>Denaturation mapping of DNA in nanofluidic channels</i>	64
<i>Helicos™ Genetic Analysis System</i>	64
<i>Molecular Combing</i>	65
<i>Nanopore sequencing</i>	65
<i>Optical Mapping</i>	66
<i>PNA-based single-molecule detection of specific DNA sequences</i>	67
<i>Sequencing-by-synthesis for single-molecule sequencing</i>	68
<i>Single molecule DNA sequencing by use of carbon nanotubes</i>	69
<i>Single molecule sequencing using Qdot nanocrystals</i>	69
<i>Single-molecule DNA sequencing in a sTOP chip nanowell</i>	70
<i>Single-molecule real-time sequencing</i>	70
Single cell DNA sequencing	71
Haplotyping for whole genome sequencing	72
ImmunoSEQ technology	72
Role of bioinformatics in sequencing	73
Growth of the sequencing database.....	73
Data storage	73
Bioinformatics challenges of new sequencing technology.....	74
Bioinformatic tools for analysis of genomic sequencing data	74
<i>Commercially available software for DNA sequencing</i>	75
<i>ChIA-PET tool for analysis with paired-end tag sequencing</i>	76
<i>Detection of CNVs and gene duplications</i>	76
<i>Differential expression analysis for sequence count data</i>	77
<i>Expression profiling without genome sequence information</i>	77
VAAST.....	77
Sequencing data storage.....	78
Sequencing analysis tools at academic organizations	78
Accessing DNA sequence information.....	79
Analysis of genomic variation by sequencing of large populations.....	79
Fund of research into interpretation of sequencing data	79
Future challenges for managing sequencing data	80
3. Comparative Analysis of Sequencing Technologies	82
General findings of the study	82

Sanger versus second generation marketed sequencers.....	82
Common features and differences among second generation sequencers	82
Third generation large sequencers.....	83
SOLiD4 versus competing large sequencers.....	83
Illumina's HiSeq and MiSeq sequencers	84
Life Technologies' Benchtop Ion Proton™ Sequencer	85
The ideal small sequencer.....	85
SWOT analysis of small sequencers	86
Concluding remarks on SWOT analysis	88

4. Research Applications of Sequencing..... 90

Introduction..... 90

Applications in basic research..... 91

ChIA-PET technology for 3D study of the genome	91
ChIP-Seq for study of gene expression	91
Chromatin profiling by direct DNA sequencing	92
Discovery of immunoglobulin gene by pyrosequencing.....	92
Epigenetic modifications analyzed by next generation sequencing	92
Exome sequencing for study of human variation	92
Genome sequencing with combinatorial probe anchor ligation.....	93
GS-FLX sequencing for simultaneous detection of mutation and CNV	93
Identifying protein-coding genes in genomic sequences	93
mRNA sequencing	94
Mutation rate measured by direct sequencing	95
Protein-protein interactome network mapping	95
Sequencing the transcriptomes of stem cells	96
Sequencing and synthetic biology	96
<i>Synthetic sequence in a bacterial cell</i>	<i>96</i>
<i>Functional synthetic proteins</i>	<i>97</i>

Sequencing of human genomes 97

Whole genome sequencing.....	97
Whole-genome sequencing of methylome	97
Whole genome resequencing	98
Personal genome sequencing.....	98
Sequencing 1000 human genomes.....	99
Missing human genome sequences.....	99
Role of sequencing in identification of human remains	100
Sequence map of the human pan-genome	100
Sequencing of African genomes	101
Sequencing of Korean genomes	101
Sequencing mitochondrial genome	102
Sequencing of ancient genomes.....	102
<i>Saqqaq genome.....</i>	<i>102</i>
<i>Neandertal genome.....</i>	<i>103</i>
Future prospects of human genome sequencing.....	103

Sequencing genomes of non-human primates 103

Sequencing genomes of other organisms..... 104

Sequencing of the mouse genome.....	104
Sequencing of the dog genome.....	105
Sequencing the turkey genome.....	106
Sequencing of the frog genome	106
Sequencing the genome of body louse.....	107
Sequencing of the ant genome	107
Sequencing of the genome of water flea	108

Applications in drug discovery and development 108

Resequencing	108
RNA profiling	109
Transcriptome sequencing for mRNA Expression	110
RNA splice variants	110
Quantitative selection of aptamers through sequencing.....	111

Sequencing projects supported by US Government..... 111

NHGRI's sequencing initiatives.....	111
JGI's Community Sequencing Program.....	111
Approved medical sequencing projects	112

1000 Genomes Project..... 112

Findings of some studies of the 1000 genomes project	113
HapMap catalog as a foundation	114
Role of SOLiD™ System in 1000 Genomes Project.....	114
Concluding remarks.....	114

Human Variome Project 114

Academic centers conducting research on sequencing..... 115

Important academic collaborations.....	116
Scientific manpower for sequencing.....	117
Large-scale genomic sequencing projects at Joint Genome Institute	117
5. Applications of Sequencing in Healthcare	120
Introduction.....	120
Applications of sequencing in molecular diagnostics	120
Next generation sequencing for detection of solid organ transplant rejection	121
Applications of sequencing in oncology	121
A project to assess sequencing technologies for tumor DNA	122
Amplicon sequencing in cancer	122
Cancer Genome Atlas	122
Detection of cancer biomarkers.....	123
<i>Sequencing mitochondrial DNA to identify cancer biomarkers</i>	<i>123</i>
<i>Biomarkers for personalizing cancer treatment</i>	<i>124</i>
Digital proteomics for cancer profiling	125
Epigenome profiling.....	125
Gaining insights into mutational processes	125
Multiplexed cancer gene mutation analysis	126
NGS-based molecular profiling of cancer in FFPE specimens	126
Paired-end sequencing.....	127
Pathology tissue-ChIP.....	128
RNA-Seq to study cancer transcriptome	128
Sequencing cancer cell lines	128
Sequencing for studying chromothripsis in cancer	129
Sequencing of complex human cancer genomes.....	129
Sequencing single cells to study evolution of cancer	129
Sequencing for assessing resistance to anticancer therapy	130
Sequencing in various types of malignancies	130
<i>Brain tumors</i>	<i>130</i>
<i>Breast cancer</i>	<i>131</i>
<i>Hematological malignancies.....</i>	<i>133</i>
<i>Hepatocellular carcinoma</i>	<i>134</i>
<i>Melanoma</i>	<i>134</i>
<i>Ovarian cancer</i>	<i>135</i>
<i>Prostate cancer.....</i>	<i>135</i>
Sequencing in genetic disorders	136
Array-based whole-exome sequencing in Bartter syndrome.....	136
Detection of X-linked disorder due to N-terminal acetyltransferase deficiency	136
DNA sequencing of maternal plasma for detection of fetal aneuploidy	137
Exome sequencing for genetic disorders	138
<i>Discovery of the gene for Miller syndrome.....</i>	<i>138</i>
<i>Discovery of the gene for Kabuki syndrome.....</i>	<i>138</i>
<i>Familial combined hypolipidemia</i>	<i>139</i>
<i>Familial thoracic aortic aneurysm</i>	<i>139</i>
<i>Proteus syndrome</i>	<i>139</i>
Sequencing for study of transposons	139
Sequencing of whole genome in Charcot-Marie-Tooth disease.....	140
Sequencing in muscular dystrophy	140
Sequencing in Huntington's disease.....	140
Sequencing genomes of the newborn to screen for genetic disorders	141
Study of rare variants in pinpointing disease-causing genes	142
Whole genome sequencing for diagnosis of genetic disorders.....	142
Sequencing for HLA typing.....	143
Sequencing for study of the human immune system.....	143
Sequencing in neurological disorders	144
Sequencing for mutations associated with autism spectrum disorders	144
Sequencing in attention-deficit/hyperactivity disorder.....	144
Sequencing genomes of microbes	145
DNA sequencing for study of bacterial epidemics.....	145
Human Microbiome Project.....	147
Pyrosequencing of microbial flora in leg ulcers	148
Sequencing of gut microbes in obesity	148
Sequencing for mapping genomic variation in <i>Mycobacterium ulcerans</i>	149
Sequencing for mapping genetic interactions in bacteria	149
Sequencing for study of antibiotic resistance in bacteria.....	149
Sequencing of DNA from single cells of bacteria	150
Sequencing of the <i>Candida</i> genome.....	151
Sequencing of human salivary microbiome	151
Sequencing in the management of HIV/AIDS	152
Sequencing in the management of HBV	152

Sequencing in the management of HCV	153
Sequencing genome of Lassa fever virus	154
Surveillance of H1N1 influenza A virus using resequencing arrays	154
Population targeted sequencing studies	154
Sequencing in aging research	155
Next generation sequencing and pharmaceutical industry	155
Next generation sequencing and drug design and discovery	155
Next generation sequencing and drug safety	156
Next generation sequencing for antibacterial therapeutic discovery	156
Applications of human transcriptome array in clinical trials	156
Role of sequencing in personalized medicine	157
Whole genome sequencing and personalized medicine	158
Role of sequencing in personalized cancer management	158
Standardization of sequencing for personalized medicine	159
Future of sequencing and personalized medicine	159
Future prospects of next generation sequencing	160
Devices for next generation sequencing	161
Clinical applications of NGS	161
6. Markets for Sequencers	162
Introduction.....	162
Methods used for estimation of sequencer markets	162
Currently marketed sequencers	162
Academic and research markets for sequencing	162
Factors affecting future development of sequencing markets.....	163
Future needs and support of research.....	163
Bioinformatics in relation to sequencing	163
Reducing the cost of human genome sequencing	164
<i>US Government-supported research on sequencing.....</i>	<i>164</i>
<i>Contribution of American Recovery and Reinvestment Act</i>	<i>166</i>
<i>Genome X Prize Foundation.....</i>	<i>166</i>
<i>Innovations to reduce cost of whole genome sequencing.....</i>	<i>167</i>
<i>Commercial aspects of low cost genome sequencing.....</i>	<i>167</i>
Genome sequencing suitable for personalized medicine	167
The global sequencing market	168
Marketing potential for sequencers	169
Challenges to developing market for sequencers	170
Recommendations	170
7. Companies Involved in Sequencing	172
Introduction.....	172
Collaborations.....	281
8. References.....	286

Tables

Table 1-1: Historical landmarks in DNA sequencing	14
Table 1-2: Genetic variations in the human genome	21
Table 2-1: ChIP detection platforms for sequencing	60
Table 2-2: Companies developing whole genome chips/microarrays	60
Table 2-3: Systems for single molecule sequencing	63
Table 2-4: Companies providing DNA sequencing software.....	75
Table 3-1: Comparison of a generation I and generation II sequencers.....	82
Table 3-2: Similarities and differences between second generation sequencers.....	83
Table 3-3: SWOT of ABI 310	86
Table 3-4: SWOT of IBS sequencing	86
Table 3-5: SWOT of NABsys' Hybridization-Assisted Nanopore Sequencing.....	86
Table 3-6: SWOT of 4300 DNA Analysis System Li-Cor	86
Table 3-7: SWOT of Genometrica	86
Table 3-8: SWOT of Polonator	87
Table 3-9: SWOT of GS FLEX Junior	87
Table 3-10: SWOT of Oxford Nanopore Technologies' nanopore DNA sequencing	87
Table 3-11: SWOT of Ion Torrent™ Personal Genome Machine	87
Table 3-12: SWOT of Pacific BioSciences' single-molecule real-time sequencing.....	88
Table 4-1: Number of genes in organisms with fully sequenced genomes.....	90
Table 4-2: Approved medical sequencing projects	112
Table 4-3: Academic centers conducting research on DNA sequencing	115
Table 4-4: Distribution of scientific manpower for sequencing.....	117

Table 5-1: Companies involved in application of sequencing in molecular diagnostics	120
Table 6-1: Marketed next generation sequencers	162
Table 6-2: De novo sequencing vs resequencing markets	163
Table 6-3: Global markets for sequencing services according to geographical regions.....	168
Table 6-4: Global markets for sequencing services according to applications	168
Table 6-5: Global markets for sequencers from 2011 to 2021.....	170
Table 7-1: Companies developing sequencing technologies and instruments	172
Table 7-2: Companies that provide sequencing services	173
Table 7-3: Companies that provide bioinformatics support for sequencing.....	174
Table 7-4: Selected collaborations for DNA sequencing	281

Figures

Figure 2-1: DNA sequencing process	31
Figure 2-2: Comparison of traditional sequencing and next generation sequencing.....	38
Figure 2-3: Watson-Crick base pairing	39
Figure 2-4: Genome Sequencer FLX system(Roche/454).....	44
Figure 2-5: Workflow of Genome Sequencer FLX system.....	45
Figure 2-6: Sequencing by ligation.....	53
Figure 2-7: Construction of SOLiD fragment library using DNA enrichment by ChIP	59
Figure 2-8: Single molecule, realtime DNA sequencing	71
Figure 5-1: Role of sequencing in the development of personalized medicine	157
Figure 6-1: Global markets for sequencing services according to applications	169