

Contents

Contributors.....	xi
-------------------	----

SECTION I Privacy challenges in genomic data sharing

CHAPTER 1 Criticality of data sharing in genomic research and public views of genomic data sharing3

Gamze Gürsoy, PhD

1. Introduction	3
2. Advancing research and scientific knowledge.....	4
3. Importance of genomic data sharing in curing diseases.....	5
3.1 Rare disease perspective.....	5
3.2 Cancer perspective.....	6
3.3 Genome-wide association studies perspective	9
4. Impact of large-scale data sharing on basic research and discovery	11
4.1 The International HapMap Project.....	11
4.2 1000 Genomes Project.....	12
4.3 The Cancer Genome Atlas project	13
4.4 Encyclopedia of DNA Elements Project.....	13
4.5 Genotype-Tissue Expression Project.....	14
5. Reproducibility	15
6. Patient/public perspective	15
Acknowledgments.....	16
References.....	16

CHAPTER 2 Genomic data access policy models..... 19

Stephanie O.M. Dyke, PhD

1. Data-sharing policy developments.....	19
2. Open-access policy model	20
3. Controlled-access policy model	22
4. Registered-access policy model	25
5. Ongoing concerns and developments	28
5.1 Maintaining consent: Consent Codes.....	28
5.2 Data-sharing risk assessment: choosing the right access level.....	29
References.....	30

CHAPTER 3	Information leaks in genomic data: inference attacks.....	33
	<i>Erman Ayday, PhD</i>	
1.	Inference attacks on statistical genomic databases.....	33
2.	Inference attacks on genomic data-sharing beacons.....	34
3.	Inference attacks on kin genomic privacy.....	38
4.	Inference attacks using genotype–phenotype associations.....	43
5.	Conclusions.....	46
	Acknowledgments.....	47
	References.....	47
CHAPTER 4	Genealogical search using whole-genome genotype profiles.....	51
	<i>Yuan Wei, Ryan Lewis, Ardalan Naseri, PhD, Shaojie Zhang, PhD and Degui Zhi, PhD</i>	
1.	Introduction.....	51
2.	History of personal genetic data.....	51
2.1	HapMap.....	52
2.2	1000 genomes Project.....	52
2.3	UK Biobank and beyond.....	52
3.	Direct-to-consumer genetic companies.....	53
3.1	Early days.....	53
3.2	Growth.....	54
3.3	Current trends.....	54
3.4	GEDMatch and others.....	55
4.	How to encode genotype information at the genome scale.....	56
4.1	Genotype.....	56
4.2	Haplotype.....	58
4.3	Phasing.....	58
4.4	File format.....	60
5.	Identity-by-descent segment and familial relatedness.....	68
5.1	Genetic distance.....	68
5.2	What is IBD.....	70
5.3	Cousin nomenclature.....	71
5.4	How IBD is related to family relationships.....	72
5.5	Expected IBD family sharing.....	74
6.	Genealogical search.....	76
6.1	What is a genealogy search.....	76
6.2	Genotype-based method.....	78

6.3	Haplotype-based method.....	79
6.4	Benchmarking of IBD detection: runtime, power, and accuracy.....	81
7.	Practical methods.....	81
7.1	Methods used by DTC companies.....	81
8.	Challenges and unmet needs.....	82
8.1	Ancestry bias.....	82
8.2	Phasing imperfection.....	84
8.3	Benchmarking of genealogical search.....	84
9.	Privacy concerns.....	84
10.	Conclusions.....	86
	Acknowledgments.....	86
	References.....	86

SECTION II Privacy-preserving techniques for responsible genomic data sharing

CHAPTER 5 Homomorphic encryption..... 97

Kim Laine, PhD

1.	Overview.....	97
1.1	Early ideas.....	98
1.2	Homomorphic encryption.....	98
1.3	Note about terminology.....	99
1.4	Implementations.....	99
1.5	Standardization.....	100
1.6	Applications.....	100
2.	Homomorphic encryption.....	101
2.1	What is encryption?.....	101
2.2	Partially homomorphic encryption.....	101
2.3	Mathematical background.....	102
2.4	(Ring) Learning With Errors.....	104
2.5	The Brakerski—Fan—Vercauteren scheme.....	106
2.6	Computing on encrypted integers.....	112
2.7	Batching.....	112
2.8	Approximate arithmetic on encrypted numbers.....	114
2.9	Other schemes.....	115
3.	Applications.....	116
3.1	Outsourced storage and computation.....	116
3.2	Private prediction.....	116

3.3 Private learning	117
3.4 PSI and PIR.....	117
3.5 Biomedical applications.....	118
4. Future Outlook	118
4.1 Complexity.....	118
4.2 Usability	119
4.3 Hardware	119
5. Conclusions.....	120
References.....	120
CHAPTER 6 Secure multi-party computation.....	123
<i>Yan Huang, PhD</i>	
1. A brief overview	123
2. Defining security.....	124
3. Protocols	126
3.1 Oblivious transfer	127
3.2 Multiplicative triples	127
3.3 Generic MPC in linear rounds.....	127
3.4 Generic MPC in constant rounds	128
3.5 Pool-based cut-and-choose	133
References.....	133
CHAPTER 7 Game theory for privacy-preserving sharing of genomic data	135
<i>Zhiyu Wan, Yevgeniy Vorobeychik, Ellen Wright Clayton, Murat Kantarcioglu and Bradley Malin</i>	
1. Introduction	135
2. Background.....	137
3. Membership-inference game	138
3.1 The game and its solutions	138
3.2 Limitations of the model.....	141
4. Beacon service game	142
4.1 The game and its solutions	142
4.2 Limitations of the model.....	149
5. Kinship game	150
5.1 The game and its solutions	150
5.2 Limitations of the model.....	152
6. Re-identification game	153
6.1 The game and its solutions	153
6.2 Limitations of the model.....	155

7. Discussion	155
8. Conclusions.....	157
Acknowledgments.....	157
References.....	157
CHAPTER 8 Trusted execution environment with Intel SGX.....	161
<i>Somnath Chakrabarti, Thomas Knauth, Dmitrii Kuvaiskii, Michael Steiner and Mona Vij</i>	
1. Introduction	161
2. Trusted execution environment.....	163
3. Intel Software Guard Extensions.....	166
3.1 Hardware architecture.....	168
3.2 Software development	175
3.3 Security properties	178
3.4 Performance properties	181
4. HW-MPC and SGX in cloud.....	182
4.1 TEEs in the cloud.....	183
4.2 Developing with SGX in the cloud	183
4.3 Deploying SGX applications.....	184
4.4 Open questions on SGX in the cloud.....	186
4.5 Putting it all together.....	187
5. Outlook.....	188
References.....	189
Index.....	191