

# ISO/IEC 23092-6:2023-11 (E)

## Information technology - Genomic information representation - Part 6: Coding of genomic annotations

---

<b>Contents</b>		<b>Page</b>
Foreword .....		vi
Introduction .....		vii
1	Scope .....	1
2	Normative references .....	1
3	Terms and definitions .....	1
4	Abbreviated terms .....	8
5	Conventions .....	8
5.1	General .....	8
5.2	Logical operators .....	8
5.3	Arithmetic operators .....	9
5.4	Relational operators .....	9
5.5	Bit-wise operators .....	9
5.6	Assignment operators .....	10
5.7	Range notation .....	10
5.8	Mathematical functions .....	10
5.9	Array and strings operation functions .....	11
5.10	Order of operation precedence .....	11
5.11	Variables, syntax elements and tables .....	12
5.12	Text description of logical operators .....	13
5.13	Processes .....	15
5.13.1	General .....	15
5.13.2	Process output operators .....	15
5.14	Method of specifying syntax in tabular form .....	16
5.15	Bit ordering .....	17
5.16	Specification of syntax functions and data types .....	17
5.17	Semantics .....	17
6	Data Structures .....	18
6.1	General .....	18
6.2	Data unit .....	18
6.3	Annotation parameter set .....	19
6.3.1	General .....	19
6.3.2	Tile configuration .....	20
6.3.3	Annotation encoding parameters .....	23
6.3.4	Descriptor configuration .....	24
6.3.5	Compressor parameter set .....	31
6.3.6	Attribute parameter set .....	32
6.4	Annotation access unit .....	34
6.4.1	General .....	34
6.4.2	Annotation access unit header .....	35
6.4.3	Annotation access unit types .....	36
6.4.4	Block .....	37
7	Descriptors and attributes semantics .....	46
7.1	General .....	46

7.2	Descriptors .....	48
7.2.1	General .....	48
7.2.2	Genomic intervals .....	48
7.2.3	Genomic variants .....	48
7.2.4	Functional annotations .....	48
7.2.5	Contact matrices .....	48
7.3	Attributes .....	48
7.4	Data types .....	48
7.4.1	General .....	48
7.4.2	Typed data .....	49
8	Decompression codecs .....	50
8.1	General .....	50
8.2	Inverse transformation algorithms .....	52
8.2.1	General .....	52
8.2.2	Lempel-Ziv-Welch transform .....	52
8.2.3	Binarization transform .....	53
8.2.4	Sparse transform .....	54
8.2.5	Delta transform .....	55
8.2.6	Run-Length Encoding transform .....	57
8.2.7	Serialization transform .....	57
8.3	Decompression algorithms .....	58
8.3.1	General .....	58
8.3.2	Context-Adaptive Binary Arithmetic Coding .....	59
8.3.3	Lempel-Ziv-Markov Chain Algorithm .....	59
8.3.4	Zstandard .....	59
8.3.5	JBIG .....	59
8.3.6	Block Sorting Coder .....	60
9	Decoding process .....	60
9.1	General .....	60
9.2	Access Units decoding process .....	60
9.2.1	General .....	60
9.2.2	Genomic variant access units .....	62
9.2.3	Functional annotation Access Units .....	64
9.2.4	Gene expression Access Units .....	65
9.2.5	Position-to-position contact intensity Access Units .....	66
9.2.6	Genome browser track Access Units .....	66
9.3	Descriptors decoding process .....	67
9.3.1	General .....	67
9.3.2	Common descriptors .....	68
9.3.3	Variant site information descriptors .....	70
9.3.4	Functional annotation descriptors .....	73
9.3.5	Genotype descriptor .....	75
9.3.6	Likelihood descriptor .....	84
9.3.7	Contact matrix descriptor .....	87
9.4	Attributes decoding process .....	102
9.5	Generic block payload decoding process .....	103
9.5.1	Descriptor payload decoding process .....	103
9.5.2	Attribute payload decoding process .....	104
10	Output format .....	107
10.1	Variant site record .....	107
10.1.1	General .....	107
10.1.2	Semantics .....	108
10.1.3	Initialization .....	110
10.2	Variant genotype record .....	111
10.2.1	General .....	111
10.2.2	Semantics .....	112
10.2.3	Initialization .....	113
10.3	Sample record .....	114

10.3.1	General .....	114
10.3.2	Semantics .....	114
10.3.3	Initialization .....	115
10.4	Functional annotation record .....	115
10.4.1	General .....	115
10.4.2	Semantics .....	116
10.4.3	Initialization .....	117
10.5	Track property record .....	118
10.5.1	General .....	118
10.5.2	Semantics .....	119
10.5.3	Initialization .....	119
10.6	Track data record .....	120
10.6.1	General .....	120
10.6.2	Semantics .....	121
10.6.3	Initialization .....	121
10.7	Expression record .....	122
10.7.1	General .....	122
10.7.2	Semantics .....	123
10.7.3	Initialization .....	123
10.8	Feature record .....	124
10.8.1	General .....	124
10.8.2	Semantics .....	125
10.8.3	Initialization .....	125
10.9	Contact matrix record .....	126
10.9.1	General .....	126
10.9.2	Semantics .....	127
10.9.3	Initialization .....	128
Bibliography .....		129