

ICS 07.080

Q 8499

团 体 标 准

T/SZAS 13—2019

基因组学数据集

2019 - 12 - 05 发布

2019 - 12 - 24 实施

深圳市标准化协会 发布

目 次

	II
1	1
2	1
3	1
4	3
	A	5
	B	10

前 言

GB/T 1.1-2009

基因组学数据集

范围

规范性引用文件

GB/T 35890-2018

GB/T 29859-2013

术语、定义和缩略语

术语和定义

VCF格式 the variant call format

单核苷酸多态性 single nucleotide polymorphism;SNP

插入缺失型变异 insertion and deletion;Indel

50 bp

结构性变异 structural variation

Copy

number variation

180

1倍测序深度 1X

1

4倍测序深度 4X

4

20倍测序深度 20X

20

平均测序深度

测序通道 lane

lane

FASTQ格式 FASTQ format

FASTQ

唯一下机序列 uniq reads

Q20

20

注:

20

99%

Q20 95%

95%

20

Q30

30

注:

30

99.9%

Q30 85%

85%

30

平均读长 read length

bp 150 bp

bp

50 bp 90 bp 100

缩略语

SNV (Single Nucleotide Variants)
 SNP (Single Nucleotide Polymorphism)
 SV Structural Variants
 INDEL / Insertions/Deletions
 CNV Copy Number Variants
 MD5 MD5 Message-Digest Algorithm
 S string
 L boolean
 N number
 D date
 DT datetime
 T time

数据元目录

数据元目录公用属性

1

表1 数据元目录公用属性

	V1.0

数据元目录专用属性

B

A

附 录
(资料性附录)

简介

建库测序信息

A.1

表A.1 建库测序信息

DE07.01.001.00					S	
DE07.01.002.00					S	
DE07.01.003.00					S	
DE07.01.004.00					S	
DE07.01.005.00					S	
DE07.01.006.00					S	B.1
DE07.01.007.00					DT	
DE07.01.008.00					DT	
DE07.01.009.00					S	B.2
DE07.01.010.00					S	
DE07.01.011.00					S	
DE07.01.012.00					S	
DE07.01.013.00					S	
DE07.01.014.00					S	
DE07.01.015.00					S	
DE07.01.016.00	FASTQ	FASTQ			S	
DE07.01.017.00					S	GB T22602013
DE07.01.018.00					S	

S B €4 表 建库测序信息 (续)

DE07.01.019.00					DT	
DE07.01.020.00					S	
DE07.01.021.00					S	
DE07.01.022.00		(GRCh37.p5)			S	
DE07.01.023.00	/	/			S	
DE07.01.024.00					S	B.3
DE07.01.025.00	S	B {4			S	B.4

表 生物信息分析 (续)

DE08.01.004.00		N			S	
DE08.01.005.00		N			S	
DE08.01.009.00		N			S	
DE08.01.010.00					S	
DE08.01.011.00					S	
DE08.01.012.00		186 392154 .6 (q425.28 475.88234510 13E1499.08475)32428730487.68 Tm99.73928624687			S	

DE08.01.013.00

表 生物信息分析 (续)

DE08.01.025.00	SV	SV			S	B.11 SV
DE08.01.026.00	SV	SV			S	
DE08.01.027.00	SV	SV			S	

质控信息

A.3

表A.3 质控信息

DE01.01.001.00					S	
DE01.01.002.00					S	
DE01.01.003.00					S	
DE01.01.004.00					S	
DE01.01.005.00					S	
DE01.01.006.00		ng/μL		ng/μL		
DE01.01.007.00		L		L		
DE01.01.008.00					S	
DE01.01.009.00				bp	N	
DE01.01.010.00				%	N	
DE01.01.011.00		Gb		Gb		
DE01.01.012.00				%	N	
DE01.01.013.00					N	
DE01.01.014.00				%	N	
DE01.01.015.00				%	N	
DE01.01.016.00				%	N	
DE01.01.017.00					N	
DE01.01.018.00					N	

表 质控信息 (续)

DE01.01.019.00	1X	1X		%	N	
DE01.01.020.00	4X	4X		%	N	
DE01.01.021.00	20X	20X		%	N	
DE01.01.022.00	Q20	3.1.12			S	
DE01.01.023.00	Q30	3.1.13			S	
DE01.01.024.00	1 Q20	1 20		%	N	
DE01.01.025.00	2 Q20	2 20		%	N	
DE01.01.026.00	1 Q30	1 30		%	S	
DE01.01.027.00	2 Q30	2 30		%	S	
DE01.01.028.00	1 AT%	1 AT		%	N	
DE01.01.029.00	2 AT%	2 AT		%	N	
DE01.01.030.00	1 GC%	1 GC		%	N	
DE01.01.031.00	2 GC%	2 GC		%	N	
DE01.01.032.00				bp	N	
DE01.01.033.00				bp	N	
DE07.01.034.00				bp	N	
DE01.01.035.00	SNP	SNP		M	N	
DE01.01.036.00	/	/			N	

附 录
(资料性附录)
数据元值域代码表

测序仪名称代码

2

“01”

B.1

表B.1 测序仪名称代码表

00	Roche 454	454 GS/GS 20/GS FLX/GS FLX Titanium/GS FLX+/GS Junior
01	ABI 310	310 /3130 /3130xl
02	ABI 3500	3500/3500xl
03	ABI 3730	3730xl, 3700
04	ABI 5500	5500 / 5500xl /5500x - WI
05	ABI Solid	SOLiD 3 Plus System/SOLiD 4 System/SOLiD 4hq System/SOLiD PI System/SOLiD System 1.0/SOLiD System 2.0/SOLiD System 3.0
06	CapitalBio BioelectronSeq 4000	BioelectronSeq 4000
07	Thermo Fisher Ion Torrent PGM	Ion Torrent PGM
08	Thermo Fisher Ion Torrent Proton	Ion Torrent Proton
09	Bionano Genomics BioNano	BioNano IRY\$SAPMR
10	Complete Genomics	Complete Genomics
11	DAAN GENE	DA8600
12	Helicos BioSciences Corporation	Helicos HeliScope
13	HYK Genetic	HYKPSTARIA
14	Illumina Genome Analyzer	Genome Analyzer/Genome Analyzer II/Genome Analyzer IIx
15	Illumina HiSeq	HSeq SQ/1000/1500/2000/2500/X Ten/X Five/3000/4000
16	Illumina MiSeq	MiSeq/MiSeq DxMiSeq FGx
17	Illumina NextSeq	NextSeq500/550
18	Illumina MiniSeq	MiniSeq
19	Illumina iSeq	iSeq 100
20	Illumina NovaSeq	NovaSeq 500/6000/TM
21	BGI BGISEQ	BGISEQ1000/50/100/500
22	BGI MGISEQ	MGISEQ200/2000
23	BGI DNBSEQ	DNBSEQ

表 测序仪名称代码表（续）

24	BGI BGISEQ	BGISEQ500RS
25	BGI BGISEQ	BGISEQ500CX
26	BGI MGISEQ	MGISEQ200RS/2000RS/2000C2000CX
27	BGI DNBSEQ	DNBSEQ50/G400/E
28	Oxford Nanopore MinION	MinION
29	Oxford Nanopore GridION	GridION
30	Berry Genomics NextSeq CN500	NextSeq CN500
31	PacBio SMRT PacBio	PacBio RS RS II/Sequel
99	Other	

测序平台代码

1

“1”

B.2

表B.2 测序平台代码表

1	LS454
101	454 GS
102	454 GS 20
103	454 GS FLX
104	454 GS FLX Titanium
105	454 GS FLX+
106	454 GS Junior
2	ILLUMINA
201	HiSeq X Five
202	HiSeq X Ten
203	Illumina Genome Analyzer
204	Illumina Genome Analyzer II
205	Illumina Genome Analyzer Iix
206	Illumina HiScanSQ
207	Illumina HiSeq 1000
208	Illumina HiSeq 1500
209	Illumina HiSeq 2000
210	Illumina HiSeq 2500
211	Illumina HiSeq 3000
212	Illumina HiSeq 4000
213	Illumina MiSeq
	Illumina MiSeq Dx

表 测序平台代码表 (续)

	Illumina MiSeq FGx
214	NextSeq 500
215	NextSeq 550
216	Illumina iSeq 100
217	Illumina NovaSeq 5000
218	Illumina NovaSeq 6000
219	Illumina NovaSeq TM
3	ABI_SOLID
301	AB 5500 Genetic Analyzer
302	AB 5500xl Genetic Analyzer
303	AB 5500xl- W Genetic Analysis System
304	AB SOLiD 3 Plus System
305	AB SOLiD 4 System
306	AB SOLiD 4hq System
307	AB SOLiD PI System
308	AB SOLiD System
309	AB SOLiD System 2.0
310	AB SOLiD System 3.0
4	COMPLETE_GENOMICS
401	Complete Genomics
5	PACBIO_SMRT
501	PacBio RS
502	PacBio RS II
503	Sequel
6	ION_TORRENT
601	Ion Torrent PGM
602	Ion Torrent Proton
603	Ion Torrent S5 XL
604	Ion Torrent S5
7	CAPILLARY
701	AB 3130 Genetic Analyzer
702	AB 310 Genetic Analyzer
703	AB 3130xL Genetic Analyzer
704	AB 3500 Genetic Analyzer
705	AB 3500xL Genetic Analyzer
706	AB 3730 Genetic Analyzer
707	AB 3730xL Genetic Analyzer
8	OXFORD_NANOPORE
801	GridION
802	MinION

表 测序平台代码表 (续)

803	PromethION
9	BGISEQ
901	BGISEQ50
902	BGISEQ100
903	BGISEQ500
904	BGISEQ1000
905	MGISEQ200
906	MGISEQ2000
907	DNBSEQ7
908	BGISEQ500RS
909	BGISEQ500CX
910	MGISEQ200RS
911	MGISEQ2000RS
912	MGISEQ200CX
913	MGISEQ200CX
914	DNBSEQ50
915	DNBSEQ400
916	DNBSEQ-
10	Berry Genomics
1001	NextSeq CN500
11	CapitalBio Company
1101	BioelectronSeq 4000
12	Bionano Genomics
1201	BioNano IRYS
1202	BioNano SAPHYR
13	DAAN GENE
1301	DA8600
14	Helicos BioSciences Corporation
1401	Helicos HeliScope
15	HYK Genetic
1501	HYKPSTARIA
16	PacBio SMRT
1601	PacBio RS
1602	PacBio RS II
1603	PacBio Sequel

组装方法代码

表B.3 组装方法代码表

1	ABySS
101	Newbler
102	SOAPdenovo
103	SPAdes
104	Velvet
105	phredPhrap
106	other
2	AllPaths
3	Arachne
4	CLC NGS Cell
5	Celera Assembler
6	GS De Novo Assembler
7	JAZZ
8	MIRA
9	MaSuRCA

分子类型代码

1

“1”

B.4

表B.4 分子类型代码表

1	genomic DNA
2	genomic RNA
3	viral cRNA

是否代码

1

“1”

B.5

表B.5 是否代码表

1	
2	

文件类型代码

1

“1”

B.6

表B.6 文件类型代码表

1	CRAM	
2	BAM	SAM
3	SFF	
4	FASTQ	
5	PacBio_HDF5	PacBio hdf5
6	CompleteGenomics_native	
7	OxfordNanopore_native	

文库设置代码

1

“1”

B.7

表B.7 文库设置代码表

1	FRAGMENT/SINGLE	read
2	PAIRED	

文库来源代码

1

“1”

B.8

表B.8 文库来源代码表

1	GENOMIC	DNA DNA PCR
2	TRANSCRIPTOMIC	DNA EST cDNA RT-PCR
3	METAGENOMIC	
4	METATRANSCRIPTOMIC	
5	SYNTHETIC	DNA
6	VIRAL RNA	RNA
7	OTHER	

文库构建策略代码

1

“1”

B.9

表B.9 文库构建策略代码表

1	WGA	pcr
101	AMPLICON	PCR RT-PCR
102	CLONEEND	5' 3'
103	FINISHING	
104	ChIP-Seq	
105	MNaseSeq	MNase
106	DNaseHypersensitivity	DNaseI
107	Bisulfite - Seq	DNA
108	EST	cDNA
109	FL-cDNA	cDNA
110	CTS	
2	WGS	
201	MRESeq	
202	MeDIPSeq	DNA
203	MBDSeq	
204	Synthetic - Long-Read	DNA
205	ssRNAseq	
206	ncRNAseq	RNA snRNA RNA snoRNA RNA piRNA/piwi/RNA piwi siRNA RNA RNA
207	Hi- C	DNA
208	ATACseq	(ATAC) Tn5 DNA DNA DNA DNaseSeq
209	RADSeq	DNA
210	VALIDATION	
3	WXS	
301	FAIREseq	
302	SELEX	
303	RIP_seq	RNA CLIP-Seq HITS-CLIP PARCLIP
304	ChIA-PET	
305	Targeted-Capture	
306	Tethered Chromatin Conformation Capture	
307	OTHER	
4	RNASeq	
5	miRNASeq	miRNA
6	Tn-Seq	
7	WCS	

表 文库构建策略代码表 (续)

8	CLONE	
9	POOLCND	BACs Fosmids

SNV

表 检测软件代码表（续）

	SNV
71	SNPSVM
72	SNPTools
73	SNVer
74	SNVMix
75	SNVMix
76	SNVPPILP
77	SOAPgaea
78	SOAPgaea
79	SOAPsnp
80	SOAPsnp
81	SolSNP
82	SomaticCall
83	SomaticSniper
84	SPLINTER
85	Strelka
86	Syzygy
87	TrioCaller
88	UnifiedGenotyper
89	VAAL
90	VariantMaster
91	VariantMaster
92	VARiD
93	VarScan
94	vipR
95	Virmid
99	Other

SV检测软件代码

SV

2

SV

“01”

B.11

表B.11 SV 检测软件代码表

	SV
01	AGE
02	APOLLOH
03	BreakDancer
04	Breakpointe r
05	BreakSeq
06	Breakway

表 检测软件代码表（续）

	SV
07	CLEVER Toolkit
08	clipcrop
09	Clippers
10	Cloudbreak
11	CREST
12	DELLY
13	deStruct
14	detecttd
15	forestSV
16	FusionMap
17	GASV
18	GASVPro
19	Hydra
20	inGAP
21	nFuse
22	PEMer package
23	Pindel
24	SLOPE
25	SOAPsv
26	Socrates
27	SoftSearch
28	SPLITREAD
29	SVDetect
30	SVMerge
31	SVMiner
32	SVseq
33	TEMP
34	VariationHunter
35	ANISE and BASIL
36	Breakmer
37	cortex
38	Gustaf
39	IMR DENOM
40	inGAP

表 检测软件代码表（续）

	SV
47	Platypus
48	PopIns
49	PRISM
50	RAPTOR
51	Reprever
52	ViVar
99	Other

变异注释软件代码

2

“01”

B.12

表B.12 变异注释软件代码表

01	ANNOVAR
02	AnnTools
03	ASOoViR
04	AVIA
05	BioR
06	CADD
07	CandiSNPer
08	CHAOs
09	CNVannotator
10	COX
11	dbNSFP
12	FamAnn
13	FunctSNP
14	GEMINI
15	GeneTalk
16	GERP
17	GESND
18	LS-SNP/PDB
19	NGSSNP
20	Oncotator
21	pfSNP
22	SCAN
23	SeqAnt
24	SiPhy
25	SNPAAMapper

表 检测软件代码表 (续)

	CNV
05	CNAseq
06	CnD
07	CNValidator
08	CNVer
09	cnvHiTSeq
10	CNVnator
11	CNVrd2
12	CNVseq
13	Control - FREEC
14	CopySeq
15	GENSENG
16	HMMcopy
17	JointSLM
