

Ribosome Profiling Analysis

Bilal Wajid*, Azhar Bilal, Ahmed Zahoor, Maaz Ahmed, M Tauseeq

Department of Electrical Engineering and Technology Lahore, KSK Campus

Roll Numbers:

- **Azhar Bilal** (2017-EE-275)
- **Ahmed Zahoor** (2017-EE-276)
- **Maaz Ahmed** (2017-EE-317)
- **M Tauseeq** (2017-EE-301)

Abstract We have used a specific criteria for the Selection of these softwares. We have followed this criteria to include softwares for our project:

- Publication Date
 - Freeware
 - Linux Based
 - Offline
-

Table 1

Sr.	Software	post 2014	Free	Offline	Linux	Installed
			Data Preprocessing Data Preprocessing			
1	systemPipeR [26]	✓	✓	✓	✓	
2	RiboGalaxy [323]	✓	✓		✓	
3	riboSeqR [86]	✓	✓	✓	✓	
4	RiboTools [492]	✓	✓		✓	
5	Ribosome profiling analysis framework [106]	✓	✓	✓	✓	
6	RSCU RS [360]	✓	✓	✓	✓	
7	RIVET [133]	✓	✓		✓	
8	riboviz [62]	✓	✓	✓	✓	
9	altORFev [239]	✓	✓	✓	✓	
10	Shoelaces [38]	✓	✓	✓	✓	
11	RiboProP [539]	✓	✓	✓	✓	
			Base calling			
12	BCL2FASTQ Conversion Software [378]	✓		✓	✓	
13	Multipass [383]	✓	✓		✓	
14	VariantTools [253]	✓	✓	✓	✓	
15	PECaller [218]	✓	✓	✓	✓	
16	GBS-SNP-CROP [320]	✓	✓	✓	✓	
17	Scrappie [211]	✓	✓	✓	✓	
18	Nanocall [103]	✓	✓	✓	✓	
19	DeepNano [45]	✓	✓	✓	✓	
20	PBHoover [381]	✓	✓	✓	✓	
21	3Dec [489]	✓		✓	✓	
22	Pathogen Host Analysis Tool [159]	✓	✓	✓	✓	
23	Chiron [464]	✓	✓		✓	
24	MutAid [356]	✓	✓	✓	✓	
25	Metrichor [468]	✓	✓	✓	✓	
26	basecRAWller [444]	✓	✓	✓	✓	
			Adapter trimming			
27	Picard [332]	✓	✓	✓	✓	
28	BCL2FASTQ [186]	✓			✓	
29	MICRA [56]	✓		✓	✓	
30	Flexbar[392]	✓	✓	✓	✓	
31	ScaleHD [261]	✓	✓	✓	✓	
32	TagDust[250]	✓	✓	✓	✓	
33	QuasR[150]	✓	✓	✓	✓	
34	SAMSA [509]	✓	✓	✓	✓	
35	SOAPnuke[77]	✓	✓	✓		
36	seqclean [407]	✓	✓	✓	✓	
37	AMPTk [354]	✓	✓	✓		
38	RNA-Rocket [504]	✓		✓		

Sr.	Software	post 2014	Free	Offline	Linux	Installed
			Data Preprocessing Adapter trimming			
39	aRNApipe [13]	✓	✓	✓		
40	cutPrimers [229]	✓	✓	✓		
41	DRAP [55]	✓	✓	✓		
42	ST Pipeline [336]	✓		✓		
43	SEPIa[100]	✓	✓	✓	✓	
44	Seqpurge[448]	✓	✓	✓	✓	
45	Fast-GBS[470]	✓	✓	✓	✓	
46	PEAT[270]	✓	✓	✓	✓	
47	Atropos[119]	✓	✓	✓	✓	
48	DNApi[473]	✓	✓	✓	✓	
49	RepeatSoaker[123]	✓	✓	✓	✓	
50	G-CNV[298]	✓		✓	✓	
51	ADEPT[141]	✓	✓	✓	✓	
52	Biopieces[445]	✓	✓	✓	✓	
53	fqtrim [526]	✓	✓	✓	✓	
54	TRAPLINE [512]	✓	✓		✓	
55	bcbio-nextgen[59]	✓	✓	✓	✓	
56	PHYLUCE[138]	✓	✓	✓	✓	
57	NxTrim[345]	✓	✓	✓	✓	
58	illumiprocessor[390]	✓	✓	✓	✓	
59	RADIS [95]	✓	✓	✓	✓	
60	Cookiecutter [440]	✓	✓	✓	✓	
61	SeekDeep [188]	✓	✓	✓	✓	
62	Horse Trans[301]	✓	✓	✓	✓	
63	TRUFA[242]	✓	✓		✓	
64	Porechop[140]	✓	✓	✓	✓	
65	EAGER[365]	✓		✓	✓	
66	fastp[73]	✓	✓	✓	✓	
67	TRAPR [272]	✓	✓	✓	✓	
68	UrQt [326]	✓	✓	✓	✓	
69	GBS-SNP-CROP[320]	✓	✓	✓	✓	
70	MAGERI[428]	✓	✓	✓	✓	
71	Nephele[158]	✓	✓	✓	✓	
72	TEtools [260]	✓	✓	✓	✓	
73	miARma-Seq[18]	✓	✓	✓	✓	
74	FAST-iCLIP [145]	✓	✓	✓	✓	
75	expHTS [446]	✓	✓	✓	✓	
76	SPARTA[216]	✓	✓	✓	✓	
77	DNAp[63]	✓	✓	✓	✓	
78	infoTrim[372]	✓	✓	✓	✓	
79	LSTrAP [376]	✓	✓	✓	✓	
80	JohnsonEtAl2018[217]	✓	✓	✓	✓	
81	BLR[63]	✓	✓	✓	✓	
82	SPLICIFY[240]	✓	✓	✓	✓	

Sr.	Software	post 2014	Free	Offline	Linux	Installed
			Data Preprocessing Adapter trimming			
83	NGseqBasic[463]	✓	✓	✓	✓	
			Read quality control			
84	CLC Genomics [332] Workbench	✓		✓	✓	
85	Picard [369]	✓	✓	✓	✓	
86	HTSeq [17]	✓	✓	✓	✓	
87	bengal-bay [441]	✓	✓	✓	✓	
88	Partek Flow[333]	✓			✓	
89	SOAPnuke[77]	✓			✓	
90	UEA sRNA toolkit [443]	✓	✓	✓	✓	
91	Biopieces[445]	✓	✓	✓	✓	
92	MOCAT[476]	✓	✓	✓	✓	
93	MOLGENIS Research [333]	✓	✓		✓	
94	phantompeakqual tools [341]	✓	✓	✓	✓	
95	V-pipew[102]	✓	✓	✓	✓	
96	MultiQC[137]	✓	✓	✓	✓	
97	Agalma[174]	✓	✓	✓	✓	
98	bcbio-nextgen[?]	✓	✓	✓	✓	
99	ChiLin[379]	✓		✓	✓	
100	mistagging[134]	✓	✓	✓	✓	
101	BAMStats[196]	✓	✓	✓	✓	
102	GotCloud[219]	✓	✓	✓	✓	
103	RUBioSeq[?]	✓	✓	✓	✓	
104	fastp[73]	✓	✓	✓	✓	
105	KneadData[232]	✓	✓	✓	✓	
106	QuickNGS[94]	✓	✓	✓	✓	
107	BBT[84]	✓		✓	✓	
108	NGS TOOLBOX [61]	✓	✓	✓	✓	
109	KAT[302]	✓	✓	✓	✓	
110	Nephele[158]	✓	✓	✓	✓	
111	AfterQC[72]	✓	✓	✓	✓	
112	NGS-QC Generator [322]	✓	✓	✓	✓	
113	MetaTrans[311]	✓	✓	✓	✓	
114	miARma-Seq[18]	✓	✓	✓	✓	
115	FAST-iCLIP[145]	✓	✓	✓	✓	
116	QuickRNASeq[191]	✓	✓	✓	✓	
117	NanoPlot[105]	✓	✓	✓	✓	

Sr.	Software	post 2014	Free	Offline	Linux	Installed
			Data Preprocessing Read quality control			
118	aRNApipe[13]	✓	✓	✓	✓	
119	DRAP[55]	✓	✓	✓	✓	
120	Fast-GBS[470]	✓	✓	✓	✓	
121	XWAS[153]	✓	✓	✓	✓	
122	SePIA[205]	✓	✓	✓	✓	
123	ST Pipeline[336]	✓	✓	✓	✓	
124	MuffinInfo[7]	✓	✓	✓	✓	
125	A-GAME[349]	✓	✓	✓	✓	
126	DNAp[63]	✓	✓	✓	✓	
127	ClinQC[355]	✓	✓	✓	✓	
128	ChronQC[461]	✓	✓	✓	✓	
129	Pheniqs[151]	✓	✓	✓	✓	
130	JohnsonEtAl2018[217]	✓	✓	✓	✓	
131	filterFillIn[453]	✓	✓	✓	✓	
132	Gimpute[70]	✓	✓	✓	✓	
133	Zika-RNAseq [501] Pipeline	✓	✓	✓	✓	
134	sigQC[461]	✓		✓	✓	
135	zUMIs[357]	✓	✓	✓	✓	
136	NGS-pipe[429]	✓	✓	✓	✓	
137	miRPursuit[68]	✓	✓	✓	✓	
138	metaWRAP[475]	✓	✓	✓	✓	
139	NGseqBasic[463]	✓	✓	✓	✓	
140	FastqPuri[461]	✓	✓		✓	
141	FQC[51]	✓	✓	✓	✓	
142	RNA workbench[172]	✓	✓	✓	✓	
143	AlmostSignificant[503]	✓	✓	✓	✓	
144	XYalign[506]	✓	✓	✓	✓	
145	Hercules tool[12]	✓	✓	✓	✓	
146	artMAP[213]	✓	✓	✓	✓	
147	OncoRep[319]	✓	✓	✓	✓	
148	QASDRA[245]	✓	✓	✓	✓	
149	RNA-seq portal[268]	✓	✓		✓	
150	Zseq[9]	✓	✓	✓	✓	
151	MVA-NGS[508]	✓	✓	✓	✓	
152	CoVaCS[80]	✓			✓	
153	RNA-QC-Chain[547]	✓	✓		✓	
154	Sunbeam[88]	✓	✓	✓	✓	
			Error Connection			
155	CoLoRMap [178]	✓			✓	
156	AutoCurE [412]	✓	✓	✓	✓	
157	PBcR [35]	✓	✓	✓	✓	
158	SOAPec [465]	✓	✓	✓	✓	

Sr.	Software	post 2014	Free	Offline	Linux	Installed
			Data Preprocessing Error Connection			
159	TRAPLINE[512]	✓	✓			
160	MiRCA [228]	✓	✓	✓	✓	
161	Jabba [324]	✓	✓	✓	✓	
162	sprai [332]	✓	✓	✓	✓	
163	Bridger [67]	✓	✓	✓	✓	
164	Rcorrector [435]	✓	✓	✓	✓	
165	BFC[263]	✓	✓	✓	✓	
166	iDES [338]	✓	✓	✓	✓	
167	MAGERI[428]	✓		✓	✓	
168	ECtools[126]	✓	✓	✓	✓	
169	UMI-tools[432]	✓	✓	✓	✓	
170	BLESS[194]	✓	✓	✓	✓	
171	DUDE-Seq[257]	✓	✓			
172	Pollux[308]	✓	✓	✓	✓	
173	PoreSeq[454]	✓	✓	✓	✓	
174	MultiRes[295]	✓	✓	✓	✓	
175	EC[397]	✓		✓	✓	
176	debarcer[439]	✓	✓	✓	✓	
177	Bloocoo[6]	✓	✓	✓	✓	
178	Karect[10]	✓	✓	✓	✓	
179	NGS-eval[315]	✓	✓			
180	LQS[287]	✓	✓	✓	✓	
181	ABruijn assembler [275]	✓	✓	✓	✓	
182	RTCR[156]	✓	✓	✓	✓	
183	LoRMA[404]	✓	✓	✓	✓	
184	ntHash[327]	✓	✓		✓	
185	G-CNV[298]	✓	✓		✓	
186	ADEPt[208]	✓	✓	✓	✓	
187	UMI-Reducer[299]	✓	✓		✓	
188	C3POA[482]	✓	✓		✓	
189	ACE[424]	✓	✓	✓	✓	
190	BarraCUDA[247]	✓	✓	✓	✓	
191	Frame-Pro[127]	✓	✓	✓	✓	
192	Canu[241]	✓	✓	✓	✓	
193	EP-metagenomic[411]	✓	✓	✓	✓	
194	ISEA[267]	✓	✓	✓	✓	
195	hivmmer[199]	✓	✓		✓	
196	C3S-LAA[146]	✓	✓	✓	✓	
197	RECKONER[120]	✓	✓	✓	✓	
198	BAND-DYN-PROG[130]	✓	✓	✓	✓	
199	Shannon[224]	✓	✓	✓	✓	
200	HMCan-diff[22]	✓	✓	✓	✓	
201	HECIL[83]	✓	✓	✓	✓	

Sr.	Software	post 2014	Free	Offline	Linux	Installed
			Data Preprocessing Error Connection			
202	GuEtAl2015[173]	✓	✓	✓	✓	
203	FMOE[201]	✓	✓	✓	✓	
204	SMARTcleaner[538]	✓	✓	✓	✓	
205	TARDIS[317]	✓	✓	✓	✓	
206	MECAT[515]	✓	✓	✓	✓	
207	NaS[293]	✓	✓	✓	✓	
208	ngsReports[502]	✓	✓		✓	
209	FreeBarcodes[189]	✓	✓	✓	✓	
210	TranscriptClean[514]	✓	✓	✓	✓	
211	MiniScrub[249]	✓	✓	✓	✓	
212	QuorUM[304]	✓	✓	✓	✓	
213	SGA-ICE[405]	✓	✓	✓	✓	
214	HALC[29]	✓	✓	✓	✓	
215	TNER[112]	✓	✓	✓	✓	
216	LRCstats[246]	✓	✓	✓	✓	
217	Pacasus[505]	✓	✓	✓	✓	
218	HG-CoLoR[331]	✓	✓	✓	✓	
219	FMLRC[493]	✓	✓	✓	✓	
220	ELECTOR[306]	✓	✓	✓	✓	
221	AGILE[204]	✓	✓	✓	✓	
222	Gencore[74]	✓	✓	✓	✓	
			Depth of coverage			
223	GATK [488]	✓		✓	✓	
224	NOISEq [458]	✓	✓	✓	✓	
225	Sambamba [457]	✓	✓	✓	✓	
226	RefCov [154]	✓	✓			
227	VarAFT[117]	✓	✓	✓	✓	
228	BasePlayer [226]	✓	✓		✓	
229	SUSHI [187]	✓	✓	✓	✓	
230	mosdepth [363]	✓	✓	✓	✓	
231	BadRegionFinder [406]	✓	✓	✓	✓	
232	cljam [456]	✓	✓	✓	✓	
233	ORIO[252]	✓	✓	✓		
234	Goleft [361]	✓	✓	✓	✓	
235	Octopus-toolkit[234]	✓	✓	✓		
236	hts-nim[364]	✓	✓	✓	✓	
			Read clustering			
237	USEARCH [129]	✓		✓	✓	
238	FASTX-Toolkit [220]	✓	✓	✓	✓	
239	bx-python [190]	✓	✓	✓	✓	
240	GBS-SNP-CROP [447]	✓	✓			
241	AMPTk[354]	✓	✓	✓	✓	
242	Starcode [552]	✓	✓	✓	✓	
243	SHARAKU [472]	✓	✓	✓	✓	

Sr.	Software	post 2014	Free	Offline	Linux	Installed
			Read clustering			
244	QCluster [91]	✓		✓	✓	
245	clusterdv [309]	✓	✓	✓	✓	
246	nAWC [32]	✓	✓	✓	✓	
247	MapReduce Inchworm[233]	✓	✓	✓	✓	
248	C3S-LAA [146]	✓	✓	✓	✓	
249	FASTCAR[212]	✓	✓	✓	✓	
250	SCLUST[34]	✓	✓	✓	✓	
251	LULU[148]	✓	✓	✓	✓	
252	DRAGON[423]	✓	✓	✓	✓	
253	Bartender [541]	✓	✓	✓	✓	
254	CARNAC-LR [305]	✓	✓	✓	✓	
255	Hetero-RP [288]	✓	✓	✓	✓	
256	XenofilteR [238]	✓	✓	✓	✓	
257	isONclust [400]	✓	✓	✓	✓	
258	necklace [104]	✓	✓	✓	✓	
			Data Processing			
			Duplicate read removal			
259	Horse Trans [301]	✓	✓	✓	✓	
260	ShortRead [107]	✓	✓	✓	✓	
261	BamTools [486]	✓	✓	✓	✓	
262	ConDeTri [141]	✓	✓	✓	✓	
263	cd-hit-454 [368]	✓	✓	✓	✓	
264	FastUniq [166]	✓	✓	✓	✓	
265	NextClip [192]	✓	✓	✓	✓	
266	SILVAngs [162]	✓			✓	
267	SAMBLASTER [386]	✓	✓	✓	✓	
268	biobambam [544]	✓	✓	✓	✓	
269	QUASR [271]	✓	✓	✓	✓	
270	FGP [447]	✓	✓		✓	
271	Pyicos [497]	✓	✓	✓	✓	
272	DeDup [497]	✓	✓	✓	✓	
273	UMI-tools [497]	✓	✓	✓	✓	
274	expHTS [446]	✓	✓	✓	✓	
275	FAST-iCLIP [145]	✓	✓	✓	✓	
276	RepeatSoaker [123]	✓	✓	✓	✓	
277	Mikado [479]	✓	✓	✓	✓	
278	Octopus-toolkit [234]	✓	✓	✓	✓	
279	Subread [271]	✓	✓	✓	✓	
280	Splign [182]	✓	✓		✓	
281	viGEN [36]	✓			✓	
282	QuickNGS [94]	✓	✓	✓	✓	
283	ContextMap [40]	✓	✓	✓	✓	
284	Portcullis [303]	✓	✓	✓	✓	

Sr.	Software	post 2014	Free	Offline	Linux	Installed
			Duplicate read removal			
285	iMapSplice [303]	✓	✓	✓	✓	
286	Scavenger [523]	✓	✓	✓	✓	
287	SPLICIFY [240]	✓			✓	
288	OncoRep [319]	✓	✓	✓	✓	
289	CUSHAW [165]	✓	✓	✓	✓	
290	GEM-Mapper [307]	✓	✓	✓	✓	
291	CLC Genomics Workbench [332]	✓		✓	✓	
292	QualiMap [351]	✓	✓	✓	✓	
293	PHYLUCE [138]	✓	✓	✓	✓	
294	NGS-Bits [417]	✓	✓	✓	✓	
295	dupRadar [409]	✓	✓	✓	✓	
296	SUSHI [187]	✓	✓	✓	✓	
297	USEARCH [129]	✓			✓	
298	FASTX-Toolkit [81]	✓	✓	✓	✓	
299	AMPTk [354]	✓	✓	✓	✓	
300	SHARAKU [472]	✓	✓	✓	✓	
301	LULU [148]	✓	✓	✓	✓	
			Read alignment			
302	CLC Assembly Cell [521]	✓	✓		✓	
303	Subread [271]	✓	✓	✓	✓	
304	NovoAlign [23]	✓		✓	✓	
305	ELAND [237]	✓		✓	✓	
306	BBTools [167]	✓	✓	✓	✓	
307	SMALT [27]	✓	✓	✓	✓	
308	LifeScope [410]	✓	✓	✓	✓	
309	NextGENe [181]	✓		✓	✓	
310	Kronos [455]	✓	✓	✓	✓	
311	ScaleHD [261]	✓	✓	✓	✓	
312	Segemehl [197]	✓	✓	✓	✓	
313	Nesoni [69]	✓		✓	✓	
314	Genomatix Software Suite [431]	✓		✓	✓	
315	Genomic Alignments [33]	✓	✓	✓	✓	
316	LAMSA [278]	✓		✓	✓	
317	SAMSA [509]	✓	✓	✓	✓	
318	Genedata Expressionist [193]	✓	✓		✓	
321	BioKanga [438]	✓	✓	✓	✓	
322	PHYLUCE [138]	✓	✓	✓	✓	
323	RSF [207]	✓	✓	✓	✓	
324	Minimap [264]	✓	✓	✓	✓	

Sr.	Software	post 2014	Free	Offline	Linux	Installed
			Read Alignment			
325	Crossbow [193]	✓	✓	✓	✓	
326	CUSHAW [165]	✓	✓	✓	✓	
327	GEM-Mapper [307]	✓	✓		✓	
328	NextFlow [118]	✓	✓	✓	✓	
329	BALAU [371]	✓	✓	✓	✓	
330	systemPipeR [26]	✓	✓	✓	✓	
331	NovoAlignCS [485]	✓	✓		✓	
332	GraphMap [436]	✓	✓	✓	✓	
333	DALIGNER [413]	✓	✓	✓	✓	
334	EAGER [365]	✓	✓		✓	
335	PECaller [218]	✓	✓	✓	✓	
336	pbdagcon [329]	✓	✓	✓	✓	
337	piPipes [183]	✓	✓	✓	✓	
338	GBS-SNP-CROP [320]	✓	✓	✓	✓	
339	Meta-aligner [335]	✓	✓	✓	✓	
340	SeqPipe [21]	✓	✓	✓	✓	
341	marginAlign [210]	✓	✓	✓	✓	
342	Tychus [108]	✓	✓	✓	✓	
343	TEtools [260]	✓	✓	✓	✓	
344	EQP [418]	✓	✓	✓	✓	
345	SPARTA [216]	✓	✓	✓	✓	
346	HISEA [231]	✓	✓	✓	✓	
347	BatAlign [273]	✓	✓	✓	✓	
348	CORA [527]	✓	✓	✓	✓	
349	DuffyNGS [43]	✓	✓	✓	✓	
350	Ngmlr [419]	✓	✓	✓	✓	
351	Kart [274]	✓	✓	✓	✓	
352	SparkBWA [2]	✓	✓	✓	✓	
353	FEM [532]	✓	✓	✓	✓	
354	ALFALFA [483]	✓	✓	✓	✓	
355	OSS [518]	✓		✓	✓	
356	QCumber [31]	✓		✓	✓	
357	CarrierSeq [328]	✓	✓	✓	✓	
358	BitMapper [78]	✓		✓	✓	
359	sRNAAnalyzer [513]	✓	✓	✓	✓	
360	Umap [225]	✓	✓	✓	✓	
361	Automation [416]	✓	✓	✓	✓	
362	Arioc [511]	✓	✓	✓	✓	
363	SSBT [434]	✓	✓	✓	✓	
364	SeqLib [486]	✓	✓	✓	✓	
365	MetaSpark [548]	✓	✓	✓	✓	
366	rHAT [279]	✓	✓	✓	✓	
367	deBWT [281]	✓	✓	✓	✓	
368	DREAM-Yara [97]	✓	✓	✓	✓	
369	GAIA [478]	✓	✓	✓	✓	

Sr.	Software	post 2014	Free	Offline	Linux	Installed
			Read Alignment			
370	Magic-BLAST [44]	✓	✓	✓	✓	
371	BGREAT [530]	✓	✓	✓	✓	
372	BigBWA [1]	✓	✓	✓	✓	
373	BETSY [75]	✓	✓	✓	✓	
374	ParaBWT [283]	✓	✓	✓	✓	
375	SCRAM [144]	✓	✓	✓	✓	
376	ConsHMM [437]	✓	✓	✓	✓	
377	SparkSW [540]	✓	✓	✓	✓	
378	LSTrAP [376]	✓	✓	✓	✓	
379	BCseq [71]	✓	✓	✓	✓	
380	Scavenger [523]	✓	✓	✓	✓	
381	BLR [385]	✓	✓	✓	✓	
382	EIM [401]	✓	✓	✓	✓	
383	NanoBLASTer [16]	✓	✓	✓	✓	
384	tarSVM [161]	✓	✓	✓	✓	
385	diffsplicing [469]	✓	✓	✓	✓	
386	FHAST [142]	✓	✓	✓	✓	
387	Canary [121]	✓	✓	✓	✓	
388	LSG [42]	✓	✓	✓	✓	
389	SPRINT [531]	✓	✓	✓	✓	
390	ChIPdig [135]	✓	✓	✓	✓	
391	sirFAST [258]	✓	✓	✓	✓	
392	ICRG [57]	✓	✓	✓	✓	
393	Genalice [467]	✓	✓	✓	✓	
394	parSRA [164]	✓	✓	✓	✓	
395	GUTSS [49]	✓	✓	✓	✓	
396	Tailor [82]	✓	✓	✓	✓	
397	Qtip [248]	✓	✓	✓	✓	
398	deBGA [280]	✓	✓	✓	✓	
399	FBB [391]	✓	✓	✓	✓	
400	HiLive [286]	✓	✓	✓	✓	
401	Epimetheus [403]	✓	✓	✓	✓	
402	lordFAST [179]	✓	✓	✓	✓	
403	BisPin [373]	✓	✓	✓	✓	
404	Whisper [114]	✓	✓	✓	✓	
405	libgaba [452]	✓	✓	✓	✓	
406	NEPAL [522]	✓	✓	✓	✓	
407	CS-BWAMEM [522]	✓	✓	✓	✓	
408	DSA [520]	✓	✓	✓	✓	
409	MutaNET [198]	✓	✓	✓	✓	
410	deGSM [176]	✓	✓	✓	✓	
411	Nucl2Vec [152]	✓	✓	✓	✓	
412	Minimap2 [265]	✓	✓	✓	✓	
413	MICA-aligner [64]	✓	✓	✓	✓	
414	GBSA [402]	✓	✓	✓	✓	

Sr.	Software	post 2014	Free	Offline	Linux	Installed
			Read Alignment			
415	DEAR-O [537]	✓	✓	✓	✓	
416	RNFtools [48]	✓	✓	✓	✓	
417	MashMap [209]	✓	✓	✓	✓	
418	COSINE [4]	✓	✓	✓	✓	
419	VarHMM [374]	✓	✓	✓	✓	
420	Nimbus [50]	✓	✓	✓	✓	
421	last-split-pe [427]	✓	✓	✓	✓	
422	PipeCraft [19]	✓	✓	✓	✓	
423	BELLA [175]	✓	✓	✓	✓	
424	HiGene [111]	✓	✓	✓	✓	
			Data analysis			
			RNase footprint detection			
425	Rfoot[215]	✓	✓	✓		
			Normalization			
426	RUST[346]	✓	✓	✓		
			Ribosome P-site localization			
427	RiboProfiling[370]	✓	✓	✓	✓	
428	riboWaltz[251]	✓	✓	✓	✓	
			Ribosome stalling prediction			
429	ROSE[534]	✓	✓	✓	✓	
			Isoform-level footprint estimation			
430	Ribomap[490]	✓	✓	✓	✓	
431	Toil[480]	✓	✓	✓	✓	
432	RiboAsiteDeblur[491]	✓	✓	✓	✓	
433	riboShape[282]	✓	✓	✓	✓	
434	RiboPip[492]	✓	✓	✓	✓	
435	RiboRL[200]	✓	✓	✓	✓	
			Translated ORF prediction			
436	RiboTaper[58]	✓	✓	✓	✓	
437	RiboProfiling[370]	✓	✓	✓	✓	
438	Rp-Bp[296]	✓	✓	✓	✓	
439	riboHMM[380]	✓	✓	✓	✓	
440	uORF-Tools[414]	✓	✓	✓	✓	
441	REPARATION[337]	✓	✓	✓	✓	
442	SPECTre[85]	✓	✓	✓	✓	
443	RiboCode[516]	✓	✓	✓	✓	

Sr.	Software	post 2014	Free	Offline	Linux	Installed
			Differential translation prediction			
444	Xtail[517]	✓	✓	✓	✓	
445	RiboDiff[546]	✓	✓	✓	✓	
446	anota2seq[347]	✓	✓	✓	✓	
447	RIVET[133]	✓		✓		
448	Riborex[269]	✓	✓	✓	✓	
449	diricore[285]	✓	✓	✓	✓	
			Integrative analysis data integration			
450	PROTEO FORMER [93]	✓	✓	✓	✓	
			File manipulation			
			File merging			
451	GATK [193]	✓		✓	✓	
452	BioD [430]	✓	✓	✓	✓	
453	Seqtk [420]	✓	✓	✓	✓	
454	FASconCAT [499]	✓	✓	✓	✓	
455	BamUtil [184]	✓	✓	✓	✓	
456	TBtools [394]	✓		✓	✓	
457	JVARKIT [277]	✓	✓	✓	✓	
458	Bio-samtools [136]	✓	✓	✓	✓	
459	Htsjdk [54]	✓	✓	✓		
460	FAST [254]	✓	✓	✓	✓	
461	HTDP [292]	✓	✓	✓	✓	
462	bedr [180]	✓	✓	✓	✓	
463	Illumina-utils [132]	✓	✓	✓	✓	
464	elPrep [195]	✓	✓	✓	✓	
			File parsing-extraction			
465	fastq-tools [193]	✓	✓	✓	✓	
466	Pysam [99]	✓	✓	✓	✓	
467	ChIP-Extract [131]	✓	✓			
468	affyio [451]	✓	✓	✓	✓	
469	pbh5tools [535]	✓	✓	✓	✓	
470	rbamtools [221]	✓	✓	✓	✓	
471	Japsa [339]	✓	✓	✓	✓	
472	bioawk [39]	✓	✓	✓	✓	
473	VCF-kit [92]	✓	✓	✓	✓	
474	Fsm-lite [28]	✓	✓	✓		
475	BAMQL [312]	✓	✓	✓	✓	
476	pbcore [46]	✓	✓	✓	✓	
477	Fasta-O-Matic [425]	✓	✓	✓	✓	
478	bioSyntax [25]	✓	✓	✓	✓	
479	SeqTailor [533]	✓	✓			
480	poRe [442]	✓	✓	✓	✓	

Sr.	Software	post 2014	Free	Offline	Linux	Installed
			File parsing-extraction			
481	HGVS [494]	✓	✓	✓	✓	
482	FASTdoop [143]	✓	✓	✓	✓	
483	Genome Build Predictor [223]	✓	✓	✓	✓	
484	VariantValidator [147]	✓	✓		✓	
485	GfaPy [163]	✓	✓	✓	✓	
486	GTFtools [266]	✓	✓		✓	
487	GiraFFe Browse [155]	✓	✓	✓	✓	
488	FasParser [450]	✓	✓	✓		
489	Bazam [395]	✓	✓	✓	✓	
490	Bqbttool [47]	✓	✓	✓	✓	
			File format conversion			
491	BCL2FASTQ Conversion Software [186]	✓		✓	✓	
492	DNAMAN [37]	✓		✓	✓	
493	liftover [206]	✓		✓	✓	
494	GenePattern [387]	✓	✓	✓	✓	
495	ea-utils [545]	✓	✓	✓	✓	
496	BaseSpace [222]	✓		✓	✓	
497	vcflib [53]	✓	✓	✓	✓	
498	bam2fastq [291]	✓	✓	✓	✓	
499	bam2wig [426]	✓	✓	✓	✓	
500	bigWigToWig [422]	✓	✓	✓	✓	
501	bigWigToBedGraph [177]	✓	✓	✓	✓	
502	NanoOK [259]	✓	✓	✓	✓	
503	catfasta2phym [484]	✓	✓	✓	✓	
504	BEASTmaster [314]	✓	✓	✓	✓	
505	BIRAP [462]	✓	✓	✓	✓	
506	genePredToGtf [214]	✓	✓	✓	✓	
507	gtfToGenePred [510]	✓		✓	✓	
508	sam2bam [348]	✓	✓	✓	✓	
509	GFF3toEMBL [353]	✓	✓	✓	✓	
510	Exdir [124]	✓	✓	✓	✓	
511	NGS-FC [528]	✓	✓	✓	✓	
512	glactools [388]	✓		✓	✓	
513	Beachmat [289]	✓	✓	✓	✓	
514	EasyQC [382]	✓	✓	✓	✓	
515	KCF-Convoy [408]	✓	✓	✓	✓	
516	AMF [139]	✓	✓	✓	✓	
517	bedToGenePred [76]	✓	✓	✓	✓	
518	TCGA2BED [96]	✓	✓	✓	✓	
519	VCF2RDF [366]	✓	✓			
520	Arteria [98]	✓	✓	✓	✓	

Sr.	Software	post 2014	Free	Offline	Linux	Installed
			File format conversion			
521	ChIP-Convert [15]	✓	✓			
522	EMBLmyGFF3 [342]	✓		✓	✓	
523	PathMe [122]	✓	✓	✓	✓	
524	bax2bam [20]	✓	✓	✓	✓	
525	MutAid [356]	✓	✓	✓	✓	
526	HPG pore [459]	✓	✓	✓	✓	
527	elPrep [195]	✓	✓	✓	✓	
528	EscherConverter [235]	✓	✓			
529	pslToBigPsl [474]	✓	✓	✓	✓	
530	convert-glycoct-inp [89]	✓	✓	✓	✓	
531	CSVToQuiXML [471]	✓	✓	✓		
532	EMBL2checklists [171]	✓	✓	✓	✓	
533	XSQ Tools [389]	✓	✓	✓	✓	
			File Idexation			
534	Picard [332]	✓	✓	✓	✓	
535	Rsamtools [33]	✓	✓	✓		
536	htslib [157]	✓	✓	✓	✓	
537	Sambamba [457]	✓	✓	✓	✓	
538	BamUtil [184]	✓	✓	✓	✓	
539	NovoSort [330]	✓		✓	✓	
540	RSF [207]	✓	✓	✓	✓	
541	NextFlow [118]	✓	✓	✓	✓	
542	JVARKIT [277]	✓	✓	✓	✓	
543	Cdbfasta [466]	✓	✓	✓	✓	
544	ropeBWT [262]	✓	✓	✓	✓	
545	Htsjdk [262]	✓	✓		✓	
546	BIRAP [462]	✓	✓	✓	✓	
547	bedr [180]	✓	✓	✓	✓	
548	FMtree [79]	✓	✓	✓	✓	
549	chopBAI [230]	✓	✓	✓	✓	
550	JfxNgs [276]	✓	✓	✓	✓	
551	pufferfish [11]	✓	✓	✓	✓	
552	Big-BWT [244]	✓	✓	✓	✓	
553	cljam [456]	✓	✓	✓	✓	
554	sBWT [66]	✓	✓	✓	✓	
555	GIGGLE [256]	✓	✓	✓	✓	
556	lordFAST [179]	✓	✓	✓	✓	
557	elPrep [195]	✓	✓	✓	✓	
558	Octopus-toolkit [234]	✓	✓	✓	✓	
559	HiGene [111]	✓	✓	✓	✓	
			File comparison			
560	vcflib [53]	✓	✓	✓	✓	

Sr.	Software	post 2014	Free	Offline	Linux	Installed
			File comparison			
561	BamUtil [184]	✓	✓	✓	✓	
562	NGS-Bits [417]	✓	✓	✓	✓	
563	UBU [433]	✓	✓	✓	✓	
564	VCFcomparator [227]	✓	✓	✓	✓	
565	FuzzyWuzzy [109]	✓	✓	✓	✓	
566	VCF-kit [92]	✓	✓	✓	✓	
567	VTC [128]	✓	✓	✓	✓	
568	BenchCT [24]	✓	✓	✓	✓	
569	VCF [?]	✓	✓		✓	
570	BAM-matcher [495]	✓	✓	✓	✓	
			File sorting			
571	Galaxy [3]	✓	✓		✓	
572	fastq-tools [33]	✓	✓	✓	✓	
573	vcflib [53]	✓	✓	✓	✓	
574	NovoSort [330]	✓		✓	✓	
575	Bio-samtools [136]	✓	✓	✓	✓	
576	NGS-Bits [417]	✓	✓	✓	✓	
577	Rsamtools [339]	✓	✓	✓	✓	
578	Japsa [33]	✓	✓	✓	✓	
579	BIRAP [462]	✓	✓	✓	✓	
580	UBU [462]	✓	✓	✓	✓	
581	FAST [254]	✓	✓	✓	✓	
582	HTDP [292]	✓	✓	✓	✓	
583	cljam [456]	✓	✓	✓	✓	
584	GFF3sort [549]	✓	✓	✓	✓	
585	CoVaCS [549]	✓	✓	✓		
586	FasParser [450]	✓	✓	✓	✓	
587	Bioflow [415]	✓	✓	✓		
			File intersection			
588	Pgltools [170]	✓	✓	✓	✓	
			File sampling			
589	GATK [193]	✓	✓	✓	✓	
590	SERES [498]	✓	✓	✓	✓	
591	Seqtk [420]	✓	✓	✓	✓	
592	fastq-tools [99]	✓	✓	✓	✓	
593	JVARKIT [277]	✓	✓	✓	✓	
594	vcfsubsample [294]	✓	✓	✓	✓	

Sr.	Software	post 2014	Free	Offline	Linux	Installed
			File splitting			
595	FASTX-Toolkit [220]	✓	✓	✓	✓	
			File editing			
596	Biostrings [185]	✓	✓	✓	✓	
597	htslib [157]	✓	✓	✓	✓	
598	Anduril [236]	✓	✓	✓	✓	
599	SeqTailor [157]	✓	✓		✓	
600	BUSTools [321]	✓	✓	✓	✓	
601	Genome Build Predictor [223]	✓	✓	✓	✓	
602	GfaPy [163]	✓	✓	✓	✓	
			Variant aggregation			
603	Snpsift [87]	✓	✓		✓	
604	org Hs eg db [396]	✓	✓	✓	✓	
605	Maftools [316]	✓	✓	✓	✓	
606	MuCor [243]	✓	✓	✓	✓	
607	HNMF [290]	✓	✓	✓	✓	
			File compression			
608	bzip2 [310]	✓	✓	✓	✓	
609	GNU zip [449]	✓	✓	✓	✓	
610	GDC [113]	✓	✓	✓	✓	
611	MAFCO [313]	✓	✓	✓	✓	
612	DNAzip [551]	✓	✓	✓	✓	
613	GQT [255]	✓	✓	✓	✓	
614	fqtools [125]	✓	✓	✓	✓	
615	SeqArray [543]	✓	✓	✓	✓	
616	qProfiler [8]	✓	✓	✓	✓	
617	GTRAC [460]	✓	✓	✓	✓	
618	ORCOM [168]	✓	✓	✓	✓	
619	LW-FQZip [202]	✓	✓	✓	✓	
620	SECRAM [203]	✓	✓	✓		
621	KIC [536]	✓	✓	✓	✓	
622	MINCE [359]	✓	✓	✓	✓	
623	CompMap [550]	✓	✓	✓	✓	
624	AFRESH [358]	✓	✓	✓	✓	
625	NRGC [399]	✓		✓	✓	
626	Boiler [375]	✓	✓	✓	✓	
627	iDoComp [344]	✓	✓	✓	✓	
628	MassComp [525]	✓	✓	✓	✓	
629	LFQC [340]	✓	✓	✓	✓	
630	ChIPWig [384]	✓	✓	✓	✓	
631	GeneCodeq [169]	✓	✓	✓	✓	

Sr.	Software	post 2014	Free	Offline	Linux	Installed
			File Compression			
632	GTZ [519]	✓	✓	✓	✓	
633	SCCG [343]	✓	✓	✓	✓	
634	SPRING [507]	✓	✓	✓	✓	
635	CSAM [60]	✓	✓	✓	✓	
636	QVZ [297]	✓	✓	✓	✓	
637	ERGC [398]	✓		✓	✓	
638	TwoPaCo [325]	✓	✓	✓	✓	
639	HARC [65]	✓	✓	✓	✓	
640	LCTD [149]	✓	✓	✓	✓	
641	CoMSA [116]	✓	✓	✓	✓	
642	BdBG [496]	✓	✓	✓	✓	
643	LEON [477]	✓	✓	✓	✓	
644	Quartz [529]	✓	✓	✓	✓	
645	GTC [101]	✓	✓	✓	✓	
646	Picopore [160]	✓	✓	✓	✓	
647	MSAC [115]	✓	✓	✓	✓	
648	FaStore [393]	✓	✓	✓	✓	
649	smallWig [500]	✓	✓	✓	✓	
650	Genomic Scores [377]	✓	✓	✓	✓	
651	HiRGC [284]	✓	✓	✓	✓	
652	Metannot [334]	✓	✓	✓	✓	
653	Crumble [41]	✓	✓	✓	✓	
654	CALQ [481]	✓	✓	✓	✓	
			File filtering			
655	GATK [193]	✓	✓	✓	✓	
656	FASTX Toolkit [220]	✓	✓		✓	
657	Galaxy [3]	✓	✓		✓	
658	Seqtk [420]	✓	✓	✓	✓	
659	fastq-tools [99]	✓	✓	✓	✓	
660	dplyr [52]	✓	✓	✓	✓	
661	fqtrim [526]	✓	✓	✓	✓	
662	JVARKIT [277]	✓	✓	✓	✓	
663	mapexr [300]	✓	✓	✓	✓	
664	NGS TOOLBOX [61]	✓	✓	✓	✓	
665	nsearch [542]	✓	✓	✓	✓	
666	VarAFT [117]	✓	✓	✓	✓	
667	NGS-Bits [30]	✓	✓	✓	✓	
668	Japsa [339]	✓	✓	✓	✓	
669	UBU [433]	✓	✓	✓	✓	
670	AMPTk [354]	✓	✓	✓	✓	
671	dbHT-Trans [110]	✓	✓	✓	✓	

Sr.	Software	post 2014	Free	Offline	Linux	Installed
			File Filtering			
672	VCF-kit [92]	✓	✓	✓	✓	
673	SWEEP [90]	✓	✓	✓	✓	
674	Fasta-O-Matic [425]	✓	✓	✓	✓	
675	SSBT [434]	✓	✓	✓	✓	
676	RepeatSoaker [123]	✓	✓	✓	✓	
677	HTDP [318]	✓	✓	✓	✓	
678	fastQ brew [350]	✓	✓	✓	✓	
679	VCF/Plotein [352]	✓	✓	✓	✓	
680	FMFilter [5]	✓	✓	✓	✓	
681	Gigwa [421]	✓	✓	✓	✓	
682	SAMSVN [524]	✓	✓	✓	✓	
683	cyvcf [362]	✓	✓	✓	✓	
684	FastqPuri [367]	✓	✓	✓	✓	
685	GateKeeper [14]	✓	✓	✓	✓	
686	MutAid [356]	✓	✓	✓	✓	
687	VariantBam [487]	✓	✓	✓	✓	
688	elPrep [195]	✓	✓	✓	✓	
689	hts-nim [364]	✓	✓	✓	✓	
690	PipeCraft [19]	✓	✓	✓	✓	

References

- [1] José M Abuín, Juan C Pichel, Tomás F Pena, and Jorge Amigo. Bigbwa: approaching the burrows-wheeler aligner to big data technologies. *Bioinformatics*, 31(24):4003–4005, 2015.
- [2] José M Abuín, Juan C Pichel, Tomás F Pena, and Jorge Amigo. Sparkbwa: speeding up the alignment of high-throughput dna sequencing data. *PloS one*, 11(5):e0155461, 2016.
- [3] Enis Afgan, Dannon Baker, Bérénice Batut, Marius Van Den Beek, Dave Bouvier, Martin Čech, John Chilton, Dave Clements, Nate Coraor, Björn A Grüning, et al. The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic acids research*, 46(W1):W537–W544, 2018.
- [4] Pegah Tootoonchi Afshar and Wing Hung Wong. Cosine: non-seeding method for mapping long noisy sequences. *Nucleic acids research*, 45(14):e132–e132, 2017.
- [5] Mete Akgün, Ö Faruk Gerdan, Zeliha Görmez, and Hüseyin Demirci. Fmfilter: A fast model based variant filtering tool. *Journal of biomedical informatics*, 60:319–327, 2016.
- [6] Isaac Akogwu, Nan Wang, Chaoyang Zhang, and Ping Gong. A comparative study

- of k-spectrum-based error correction methods for next-generation sequencing data analysis. *Human genomics*, 10(2):20, 2016.
- [7] Andy S Alic and Ignacio Blanquer. Muffininfo: Html5-based statistics extractor from next-generation sequencing data. *Journal of Computational Biology*, 23(9):750–755, 2016.
- [8] Tyler S Alioto, Ivo Buchhalter, Sophia Derdak, Barbara Hutter, Matthew D Eldridge, Eivind Hovig, Lawrence E Heisler, Timothy A Beck, Jared T Simpson, Laurie Tonon, et al. A comprehensive assessment of somatic mutation detection in cancer using whole-genome sequencing. *Nature communications*, 6:10001, 2015.
- [9] Abedalrhman Alkhateeb and Luis Rueda. Zseq: an approach for preprocessing next-generation sequencing data. *Journal of Computational Biology*, 24(8):746–755, 2017.
- [10] Amin Allam, Panos Kalnis, and Victor Solovyev. Karect: accurate correction of substitution, insertion and deletion errors for next-generation sequencing data. *Bioinformatics*, 31(21):3421–3428, 2015.
- [11] Fatemeh Almodaresi, Hirak Sarkar, and Rob Patro. A space and time-efficient index for the compacted colored de bruijn graph. biorxiv. 2017.
- [12] Jamie Alnasir and Hugh P Shanahan. A novel method to detect bias in short reads data. *Journal of integrative bioinformatics*, 14(3), 2017.
- [13] Arnald Alonso, Brittany N Lasseigne, Kelly Williams, Josh Nielsen, Ryne C Ramaker, Andrew A Hardigan, Bobbi Johnston, Brian S Roberts, Sara J Cooper, Sara Marsal, et al. arnapipe: a balanced, efficient and distributed pipeline for processing rna-seq data in high-performance computing environments. *Bioinformatics*, 33(11):1727–1729, 2017.
- [14] Mohammed Alser, Hasan Hassan, Hongyi Xin, Oğuz Ergin, Onur Mutlu, and Can Alkan. Gatekeeper: a new hardware architecture for accelerating pre-alignment in dna short read mapping. *Bioinformatics*, 33(21):3355–3363, 2017.
- [15] Giovanna Ambrosini, René Dreos, Sunil Kumar, and Philipp Bucher. The chip-seq tools and web server: a resource for analyzing chip-seq and other types of genomic data. *BMC genomics*, 17(1):938, 2016.
- [16] Mohammad Ruhul Amin, Steven Skiena, and Michael C Schatz. Nanoblaster: Fast alignment and characterization of oxford nanopore single molecule sequencing reads. In *2016 IEEE 6th International Conference on Computational Advances in Bio and Medical Sciences (ICCABS)*, pages 1–6. IEEE, 2016.
- [17] Simon Anders, Paul Theodor Pyl, and Wolfgang Huber. Htseq—a python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2):166–169, 2015.

-
- [18] Eduardo Andrés-León, Rocío Núñez-Torres, and Ana M Rojas. miarma-seq: a comprehensive tool for mirna, mrna and circrna analysis. *Scientific reports*, 6:25749, 2016.
- [19] Sten Anslan, Mohammad Bahram, Indrek Hiiesalu, and Leho Tedersoo. Pipecraft: Flexible open-source toolkit for bioinformatics analysis of custom high-throughput amplicon sequencing data. *Molecular ecology resources*, 17(6):e234–e240, 2017.
- [20] Sten Anslan, Mohammad Bahram, Indrek Hiiesalu, and Leho Tedersoo. User manual for pipecraft 1.0. 2017.
- [21] Dvir Aran, Monther Abu-Remaileh, Revital Levy, Nurit Meron, Gidon Toperoff, Yifat Edrei, Yehudit Bergman, and Asaf Hellman. Embryonic stem cell (es)-specific enhancers specify the expression potential of es genes in cancer. *PLoS genetics*, 12(2):e1005840, 2016.
- [22] Haitham Ashoor, Caroline Louis-Brennetot, Isabelle Janoueix-Lerosey, Vladimir B Bajic, and Valentina Boeva. Hmcan-diff: a method to detect changes in histone modifications in cells with different genetic characteristics. *Nucleic acids research*, 45(8):e58–e58, 2017.
- [23] Shazia Ashraf, Hiroki Kudo, Jia Rao, Atsuo Kikuchi, Eugen Widmeier, Jennifer A Lawson, Weizhen Tan, Tobias Hermle, Jillian K Warejko, Shirlee Shril, et al. Mutations in six nephrosis genes delineate a pathogenic pathway amenable to treatment. *Nature communications*, 9(1):1960, 2018.
- [24] Jérôme Audoux, Mikaël Salson, Christophe F Grosset, Sacha Beaumeunier, Jean-Marc Holder, Thérèse Commes, and Nicolas Philippe. Simba: A methodology and tools for evaluating the performance of rna-seq bioinformatic pipelines. *BMC bioinformatics*, 18(1):428, 2017.
- [25] Artem Babaian, Anicet Ebou, Alyssa Fegen, Ho Yin Kam, German E Novakovsky, Jasper Wong, Dylan Aïssi, and Li Yao. biosyntax: syntax highlighting for computational biology. *BMC bioinformatics*, 19(1):303, 2018.
- [26] Tyler WH Backman and Thomas Girke. systempiper: Ngs workflow and report generation environment. *BMC bioinformatics*, 17(1):388, 2016.
- [27] Kate S Baker, Timothy J Dallman, Philip M Ashton, Martin Day, Gwenda Hughes, Paul D Crook, Victoria L Gilbert, Sandra Zittermann, Vanessa G Allen, Benjamin P Howden, et al. Intercontinental dissemination of azithromycin-resistant shigellosis through sexual transmission: a cross-sectional study. *The Lancet infectious diseases*, 15(8):913–921, 2015.
- [28] Kate S Baker, Timothy J Dallman, Nigel Field, Tristan Childs, Holly Mitchell, Martin Day, François-Xavier Weill, Sophie Lefèvre, Mathieu Tourdjman, Gwenda Hughes, et al. Horizontal antimicrobial resistance transfer drives epidemics of multiple shigella species. *Nature communications*, 9(1):1462, 2018.

-
- [29] Ergude Bao and Lingxiao Lan. Halc: High throughput algorithm for long read error correction. *BMC bioinformatics*, 18(1):204, 2017.
- [30] Riyue Bao, Lei Huang, Jorge Andrade, Wei Tan, Warren A Kibbe, Hongmei Jiang, and Gang Feng. Review of current methods, applications, and data management for the bioinformatics analysis of whole exome sequencing. *Cancer informatics*, 13:CIN-S13779, 2014.
- [31] Marco Barbieri, Emanuele Roccia, Luca Mancino, Marco Sbroscia, Ilaria Gianani, and Fabio Sciarrino. What hong-ou-mandel interference says on two-photon frequency entanglement. *Scientific reports*, 7(1):7247, 2017.
- [32] Dillon OR Barker, Joao Andre Carrico, Peter Kruczkiewicz, Federica Palma, Mirko Rossi, and Eduardo N Taboada. Rapid identification of stable clusters in bacterial populations using the adjusted wallace coefficient. *Biorxiv*, page 299347, 2018.
- [33] Benjamin G Barwick, Christopher D Scharer, Ryan J Martinez, Madeline J Price, Alexander N Wein, Robert R Haines, Alexander PR Bally, Jacob E Kohlmeier, and Jeremy M Boss. B cell activation and plasma cell differentiation are inhibited by de novo dna methylation. *Nature communications*, 9(1):1900, 2018.
- [34] Alexandre Bazin, Didier Debroas, and Engelbert Mephu Nguifo. A de novo robust clustering approach for amplicon-based sequence data. *Journal of Computational Biology*, 2018.
- [35] Konstantin Berlin, Sergey Koren, Chen-Shan Chin, James P Drake, Jane M Landolin, and Adam M Phillippy. Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. *Nature biotechnology*, 33(6):623, 2015.
- [36] Krithika Bhuvaneshwar, Lei Song, Subha Madhavan, and Yuriy Gusev. Detection and quantification of viral rna in human tumors using open source pipeline: vigen. *bioRxiv*, page 099788, 2017.
- [37] Gonzalo N Bidart, Jesús Rodríguez-Díaz, Gaspar Pérez-Martínez, and María J Yebra. The lactose operon from lactobacillus casei is involved in the transport and metabolism of the human milk oligosaccharide core-2 n-acetyllactosamine. *Scientific reports*, 8(1):7152, 2018.
- [38] Åsmund Birkeland, Katarzyna Chyżyńska, and Eivind Valen. Shoelaces: an interactive tool for ribosome profiling processing and visualization. *BMC genomics*, 19(1):543, 2018.
- [39] John D Blischak, Ludovic Tailleux, Marsha Myrthil, Cécile Charlois, Emmanuel Bergot, Aurélien Dinh, Gloria Morizot, Olivia Chény, Cassandre Von Platen, Jean-Louis Herrmann, et al. Predicting susceptibility to tuberculosis based on gene expression profiling in dendritic cells. *Scientific reports*, 7(1):5702, 2017.

-
- [40] Thomas Bonfert, Evelyn Kirner, Gergely Csaba, Ralf Zimmer, and Caroline C Friedel. Contextmap 2: fast and accurate context-based rna-seq mapping. *BMC bioinformatics*, 16(1):122, 2015.
- [41] James K Bonfield, Shane A McCarthy, and Richard Durbin. Crumble: reference free lossy compression of sequence quality values. *Bioinformatics*, 35(2):337–339, 2018.
- [42] Paola Bonizzoni, Gianluca Della Vedova, Yuri Pirola, Marco Previtali, and Raffaella Rizzi. Lsg: an external-memory tool to compute string graphs for next-generation sequencing data assembly. *Journal of Computational Biology*, 23(3):137–149, 2016.
- [43] Maikel Boot, Vincent JC van Winden, Marion Sparrius, Robert van de Weerd, Alexander Speer, Roy Ummels, Tige Rustad, David R Sherman, and Wilbert Bitter. Cell envelope stress in mycobacteria is regulated by the novel signal transduction atpase inir in response to trehalose. *PLoS genetics*, 13(12):e1007131, 2017.
- [44] Grzegorz M Boratyn, Jean Thierry-Mieg, Danielle Thierry-Mieg, Ben Busby, and Thomas L Madden. Magic-blast, an accurate dna and rna-seq aligner for long and short reads. *BioRxiv*, page 390013, 2018.
- [45] Vladimír Boža, Broňa Brejová, and Tomáš Vinař. Deepnano: deep recurrent neural networks for base calling in minion nanopore reads. *PLoS one*, 12(6):e0178751, 2017.
- [46] Sinisa Bratulic, Macarena Toll-Riera, and Andreas Wagner. Mistranslation can enhance fitness through purging of deleterious mutations. *Nature communications*, 8:15410, 2017.
- [47] Martin Brehm and Martin Thomas. An efficient lossless compression algorithm for trajectories of atom positions and volumetric data. *Journal of chemical information and modeling*, 58(10):2092–2107, 2018.
- [48] Karel Břinda, Valentina Boeva, and Gregory Kucherov. Rnf: a general framework to evaluate ngs read mappers. *Bioinformatics*, 32(1):136–139, 2015.
- [49] Mitchell J Brittnacher, Sonya L Heltshe, Hillary S Hayden, Matthew C Radey, Eli J Weiss, Christopher J Damman, Timothy L Zisman, David L Suskind, and Samuel I Miller. Gutss: An alignment-free sequence comparison method for use in human intestinal microbiome and fecal microbiota transplantation analysis. *PLoS one*, 11(7):e0158897, 2016.
- [50] RWW Brouwer, MCGN van den Hout, CEM Kockx, E Brosens, B Eussen, A de Klein, F Sleutels, and WFJ van IJcken. Nimbus: a design-driven analyses suite for amplicon-based ngs data. *Bioinformatics*, 34(16):2732–2739, 2018.

-
- [51] Joseph Brown, Meg Pirrung, and Lee Ann McCue. Fqc dashboard: integrates fastqc results into a web-based, interactive, and extensible fastq quality control tool. *Bioinformatics*, 33(19):3137–3139, 2017.
- [52] Rosalind Brown, Alice D Lam, Alfredo Gonzalez-Sulser, Andrew Ying, Mary Jones, Robert Chang-Chih Chou, Makis Tzioras, Crispin Y Jordan, Izabela Jedrasiak-Cape, Anne-Laure Hemonnot, et al. Circadian and brain state modulation of network hyperexcitability in alzheimer’s disease. *eNeuro*, 5(2), 2018.
- [53] Adam B Burkholder, Scott A Lujan, Christopher A Lavender, Sara A Grimm, Thomas A Kunkel, and David C Fargo. Muver, a computational framework for accurately calling accumulated mutations. *BMC genomics*, 19(1):345, 2018.
- [54] Benjamin J Burwitz, Jochen M Wettengel, Martin A Mück-Häusl, Marc Ringelhan, Chunkyu Ko, Marvin M Festag, Katherine B Hammond, Mina Northrup, Benjamin N Bimber, Thomas Jacob, et al. Hepatocytic expression of human sodium-taurocholate cotransporting polypeptide enables hepatitis b virus infection of macaques. *Nature communications*, 8(1):2146, 2017.
- [55] Cédric Cabau, Frédéric Escudié, Anis Djari, Yann Guiguen, Julien Bobe, and Christophe Klopp. Compacting and correcting trinity and oases rna-seq de novo assemblies. *PeerJ*, 5:e2988, 2017.
- [56] Ségolène Caboche, Gaël Even, Alexandre Loywick, Christophe Audebert, and David Hot. Micra: an automatic pipeline for fast characterization of microbial genomes from high-throughput sequencing data. *Genome biology*, 18(1):233, 2017.
- [57] Maurizio Callari, Ankita Sati Batra, Rajbir Nath Batra, Stephen-John Sammut, Wendy Greenwood, Harry Clifford, Colin Hercus, Suet-Feung Chin, Alejandra Bruna, Oscar M Rueda, et al. Computational approach to discriminate human and mouse sequences in patient-derived tumour xenografts. *BMC genomics*, 19(1):19, 2018.
- [58] Lorenzo Calviello, Neelanjan Mukherjee, Emanuel Wyler, Henrik Zaubler, Antje Hirsekorn, Matthias Selbach, Markus Landthaler, Benedikt Obermayer, and Uwe Ohler. Detecting actively translated open reading frames in ribosome profiling data. *Nature methods*, 13(2):165, 2016.
- [59] Juliana B Candido, Jennifer P Morton, Peter Bailey, Andrew D Campbell, Saa-dia A Karim, Thomas Jamieson, Laura Lapienyte, Aarthi Gopinathan, William Clark, Ewan J McGhee, et al. Csf1r+ macrophages sustain pancreatic tumor growth through t cell suppression and maintenance of key gene programs that define the squamous subtype. *Cell reports*, 23(5):1448–1460, 2018.
- [60] Rodrigo Cánovas, Alistair Moffat, and Andrew Turpin. Csam: Compressed sam format. *Bioinformatics*, 32(24):3709–3716, 2016.

-
- [61] E Capra, F Turri, B Lazzari, P Cremonesi, TM Gliozzi, I Fojadelli, A Stella, and F Pizzi. Small rna sequencing of cryopreserved semen from single bull revealed altered mirnas and pirnas expression between high-and low-motile sperm populations. *BMC genomics*, 18(1):14, 2017.
- [62] Oana Carja, Tongji Xing, Edward WJ Wallace, Joshua B Plotkin, and Premal Shah. riboviz: analysis and visualization of ribosome profiling datasets. *BMC bioinformatics*, 18(1):461, 2017.
- [63] Jason L Causey, Cody Ashby, Karl Walker, Zhiping Paul Wang, Mary Yang, Yuanfang Guan, Jason H Moore, and Xiuzhen Huang. Dnap: A pipeline for dna-seq data analysis. *Scientific reports*, 8(1):6793, 2018.
- [64] Sze-Hang Chan, Jeanno Cheung, Edward Wu, Heng Wang, Chi-Man Liu, Xiaoqian Zhu, Shaoliang Peng, Ruibang Luo, and Tak-Wah Lam. Mica: A fast short-read aligner that takes full advantage of intel many integrated core architecture (mic). *arXiv preprint arXiv:1402.4876*, 2014.
- [65] Shubham Chandak, Kedar Tatwawadi, and Tsachy Weissman. Compression of genomic sequencing reads via hash-based reordering: algorithm and analysis. *Bioinformatics*, 34(4):558–567, 2017.
- [66] Chia-Hua Chang, Min-Te Chou, Yi-Chung Wu, Ting-Wei Hong, Yun-Lung Li, Chia-Hsiang Yang, and Jui-Hung Hung. sbwt: memory efficient implementation of the hardware-acceleration-friendly schindler transform for the fast biological sequence mapping. *Bioinformatics*, 32(22):3498–3500, 2016.
- [67] Zheng Chang, Guojun Li, Juntao Liu, Yu Zhang, Cody Ashby, Deli Liu, Carole L Cramer, and Xiuzhen Huang. Bridger: a new framework for de novo transcriptome assembly using rna-seq data. *Genome biology*, 16(1):30, 2015.
- [68] Inês Chaves, Bruno Vasques Costa, Andreia S Rodrigues, Andreas Bohn, and Célia M Miguel. mi rp ursuit—a pipeline for automated analyses of small rna s in model and nonmodel plants. *FEBS letters*, 591(15):2261–2268, 2017.
- [69] Bachar Cheaib, Malo Le Boulch, Pierre-Luc Mercier, and Nicolas Derome. Taxon-function decoupling as an adaptive signature of lake microbial metacomunities under a chronic polymetallic pollution gradient. *Frontiers in microbiology*, 9, 2018.
- [70] Junfang Chen, Dietmar Lippold, Josef Frank, William Rayner, Andreas Meyer-Lindenberg, and Emanuel Schwarz. Gimpute: an efficient genetic data imputation pipeline. *Bioinformatics*, 2018.
- [71] Liang Chen and Sika Zheng. Bcseq: accurate single cell rna-seq quantification with bias correction. *Nucleic acids research*, 46(14):e82–e82, 2018.
- [72] Shifu Chen, Tanxiao Huang, Yanqing Zhou, Yue Han, Mingyan Xu, and Jia Gu. Afterqc: automatic filtering, trimming, error removing and quality control for fastq data. *BMC bioinformatics*, 18(3):80, 2017.

-
- [73] Shifu Chen, Yanqing Zhou, Yaru Chen, and Jia Gu. fastp: an ultra-fast all-in-one fastq preprocessor. *Bioinformatics*, 34(17):i884–i890, 2018.
- [74] Shifu Chen, Yanqing Zhou, Yaru Chen, Tanxiao Huang, Wenting Liao, Yun Xu, Zhihua Liu, and Jia Gu. gencore: an efficient tool to generate consensus reads for error suppressing and duplicate removing of ngs data. *bioRxiv*, page 501502, 2018.
- [75] Xiaoling Chen and Jeffrey T Chang. Planning bioinformatics workflows using an expert system. *Bioinformatics*, 33(8):1210–1215, 2017.
- [76] Xiaowei Chen, Yajing Hao, Ya Cui, Zhen Fan, and Runsheng Chen. Lncvar: Deciphering genetic variations associated with long noncoding genes. In *Epitranscriptomics*, pages 189–198. Springer, 2019.
- [77] Yuxin Chen, Yongsheng Chen, Chunmei Shi, Zhibo Huang, Yong Zhang, Shengkang Li, Yan Li, Jia Ye, Chang Yu, Zhuo Li, et al. Soapnuke: a mapreduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience*, 7(1):gix120, 2017.
- [78] Haoyu Cheng, Huaipan Jiang, Jiaoyun Yang, Yun Xu, and Yi Shang. Bitmapper: an efficient all-mapper based on bit-vector computing. *BMC bioinformatics*, 16(1):192, 2015.
- [79] Haoyu Cheng, Ming Wu, and Yun Xu. Fmtree: a fast locating algorithm of fm-indexes for genomic data. *Bioinformatics*, 34(3):416–424, 2017.
- [80] Matteo Chiara, Silvia Gioiosa, Giovanni Chillemi, Mattia D’Antonio, Tiziano Flati, Ernesto Picardi, Federico Zambelli, David Stephen Horner, Graziano Pesole, and Tiziana Castrignanò. Covacs: a consensus variant calling system. *BMC genomics*, 19(1):120, 2018.
- [81] Chih-Hung Chou, Hsi-Yuan Huang, Wei-Chih Huang, Sheng-Da Hsu, Chung-Der Hsiao, Chia-Yu Liu, Yu-Hung Chen, Yu-Chen Liu, Wei-Yun Huang, Meng-Lin Lee, et al. The aquatic animals’ transcriptome resource for comparative functional analysis. *BMC genomics*, 19(2):103, 2018.
- [82] Min-Te Chou, Bo W Han, Chiung-Po Hsiao, Phillip D Zamore, Zhiping Weng, and Jui-Hung Hung. Tailor: a computational framework for detecting non-templated tailing of small silencing rnas. *Nucleic acids research*, 43(17):e109–e109, 2015.
- [83] Olivia Choudhury, Ankush Chakrabarty, and Scott J Emrich. Hecil: A hybrid error correction algorithm for long reads with iterative learning. *Scientific reports*, 8(1):9936, 2018.
- [84] Justin Chu, Hamid Mohamadi, Emre Erhan, Jeffery Tse, Readman Chiu, Sarah Yeo, and Inanc Birol. Improving on hash-based probabilistic sequence classification using multiple spaced seeds and multi-index bloom filters. *BioRxiv*, page 434795, 2018.

-
- [85] Sang Y Chun, Caitlin M Rodriguez, Peter K Todd, and Ryan E Mills. Spectre: a spectral coherence-based classifier of actively translated transcripts from ribosome profiling sequence data. *BMC bioinformatics*, 17(1):482, 2016.
- [86] Betty Y Chung, Thomas J Hardcastle, Joshua D Jones, Nerea Irigoyen, Andrew E Firth, David C Baulcombe, and Ian Brierley. The use of duplex-specific nuclease in ribosome profiling and a user-friendly software package for ribo-seq data analysis. *Rna*, 21(10):1731–1745, 2015.
- [87] Pablo Cingolani, Adrian Platts, Le Lily Wang, Melissa Coon, Tung Nguyen, Luan Wang, Susan J Land, Xiangyi Lu, and Douglas M Ruden. A program for annotating and predicting the effects of single nucleotide polymorphisms, snpeff: Snps in the genome of drosophila melanogaster strain w1118; iso-2; iso-3. *Fly*, 6(2):80–92, 2012.
- [88] Erik L Clarke, Louis J Taylor, Chunyu Zhao, Andrew Connell, Jung-Jin Lee, Bryton Fett, Frederic D Bushman, and Kyle Bittinger. Sunbeam: an extensible pipeline for analyzing metagenomic sequencing experiments. *Microbiome*, 7(1):46, 2019.
- [89] Olivier Clerc, Julien Mariethoz, Alain Rivet, Frédérique Lisacek, Serge Pérez, and Sylvie Ricard-Blum. A pipeline to translate glycosaminoglycan sequences into 3d models. application to the exploration of glycosaminoglycan conformational space. *Glycobiology*, 29(1):36–44, 2018.
- [90] Josh P Clevenger and Peggy Ozias-Akins. Sweep: A tool for filtering high-quality snps in polyploid crops. *G3: Genes, Genomes, Genetics*, 5(9):1797–1803, 2015.
- [91] Matteo Comin, Andrea Leoni, and Michele Schimd. Clustering of reads with alignment-free measures and quality values. *Algorithms for Molecular Biology*, 10(1):4, 2015.
- [92] Daniel E Cook and Erik C Andersen. Vcf-kit: assorted utilities for the variant call format. *Bioinformatics*, 33(10):1581–1582, 2017.
- [93] Jeroen Crappé, Elvis Ndah, Alexander Koch, Sandra Steyaert, Daria Gawron, Sarah De Keulenaer, Ellen De Meester, Tim De Meyer, Wim Van Criekeing, Petra Van Damme, et al. Proteoformer: deep proteome coverage through ribosome profiling and ms integration. *Nucleic acids research*, 43(5):e29–e29, 2014.
- [94] Giuliano Crispatzu, Pranav Kulkarni, Mohammad R Toliati, Peter Nürnberg, Marco Herling, Carmen D Herling, and Peter Frommolt. Semi-automated cancer genome analysis using high-performance computing. *Human mutation*, 38(10):1325–1335, 2017.
- [95] Astrid Cruaud, Mathieu Gautier, Jean-Pierre Rossi, Jean-Yves Rasplus, and Jérôme Gouzy. Radis: analysis of rad-seq data for interspecific phylogeny. *Bioinformatics*, 32(19):3027–3028, 2016.

-
- [96] Fabio Cumbo, Giulia Fiscon, Stefano Ceri, Marco Masseroli, and Emanuel Weitschek. Tcga2bed: extracting, extending, integrating, and querying the cancer genome atlas. *BMC bioinformatics*, 18(1):6, 2017.
- [97] Temesgen Hailemariam Dadi, Enrico Siragusa, Vitor C Piro, Andreas Andrusch, Enrico Seiler, Bernhard Y Renard, and Knut Reinert. Dream-yara: An exact read mapper for very large databases with short update time. *Bioinformatics*, 34(17):i766–i772, 2018.
- [98] Johan Dahlberg, Johan Hermansson, Steinar Sturlaugsson, and Pontus Larsson. Arteria: An automation system for a sequencing core facility. *BioRxiv*, page 214858, 2017.
- [99] Ziwei Dai, Samantha J Mentch, Xia Gao, Sailendra N Nichenametla, and Jason W Locasale. Methionine metabolism influences genomic architecture and gene expression through h3k4me3 peak width. *Nature communications*, 9(1):1955, 2018.
- [100] Georgios A Dalkas and Marianne Rooman. Sepia, a knowledge-driven algorithm for predicting conformational b-cell epitopes from the amino acid sequence. *BMC bioinformatics*, 18(1):95, 2017.
- [101] Agnieszka Danek and Sebastian Deorowicz. Gtc: a novel attempt to maintenance of huge genome collections compressed. *BioRxiv*, page 131649, 2017.
- [102] Debanu Das, Hsiu-Ju Chiu, Carol L Farr, Joanna C Grant, Lukasz Jaroszewski, Mark W Knuth, Mitchell D Miller, Henry J Tien, Marc-André Elslinger, Ashley M Deacon, et al. Crystal structure of a putative quorum sensing-regulated protein (pa3611) from the pseudomonas-specific duf4146 family. *Proteins: Structure, Function, and Bioinformatics*, 82(6):1086–1092, 2014.
- [103] Matei David, Lewis Jonathan Dursi, Delia Yao, Paul C Boutros, and Jared T Simpson. Nanocall: an open source basecaller for oxford nanopore sequencing data. *Bioinformatics*, 33(1):49–55, 2016.
- [104] Nadia M Davidson and Alicia Oshlack. Necklace: combining reference and assembled transcriptomes for more comprehensive rna-seq analysis. *GigaScience*, 7(5):giy045, 2018.
- [105] Wouter De Coster, Sven D’Hert, Darrin T Schultz, Marc Cruts, and Christine Van Broeckhoven. Nanopack: visualizing and processing long-read sequencing data. *Bioinformatics*, 34(15):2666–2669, 2018.
- [106] Eleonora de Klerk, Ivo FAC Fokkema, Klaske AMH Thiadens, Jelle J Goeman, Magnus Palmblad, Johan T den Dunnen, Marieke von Lindern, and Peter AC ‘t Hoen. Assessing the translational landscape of myogenic differentiation by ribosome profiling. *Nucleic acids research*, 43(9):4408–4428, 2015.

-
- [107] Wélliton de Souza, Benilton de Sá Carvalho, and Iscia Lopes-Cendes. Rqc: A bio-conductor package for quality control of high-throughput sequencing data. *Journal of Statistical Software*, 87(CN 2):1–14, 2018.
- [108] Christopher Dean, Noelle Noyes, Steven M Lakin, Pablo Rovira-Sanz, Xiang Yang, Keith Belk, Paul Morley, Richard Meinersmann, and Zaid Abdo. Tychus: a whole genome sequencing pipeline for assembly, annotation and phylogenetics of bacterial genomes. *BioRxiv*, page 283101, 2018.
- [109] Christopher DeBoever, Yosuke Tanigawa, Malene E Lindholm, Greg McInnes, Adam Lavertu, Erik Ingelsson, Chris Chang, Euan A Ashley, Carlos D Bustamante, Mark J Daly, et al. Medical relevance of protein-truncating variants across 337,205 individuals in the uk biobank study. *Nature communications*, 9, 2018.
- [110] Feilong Deng and Shi-Yi Chen. dbht-trans: an efficient tool for filtering the protein-encoding transcripts assembled by rna-seq according to search for homologous proteins. *Journal of Computational Biology*, 23(1):1–9, 2016.
- [111] Liqun Deng, Guowei Huang, Yuzheng Zhuang, Jiansheng Wei, and Youliang Yan. Higene: A high-performance platform for genomic data analysis. In *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 576–583. IEEE, 2016.
- [112] Shibing Deng, Maruja Lira, Donghui Huang, Kai Wang, Crystal Valdez, Jennifer Kinong, Paul A Rejto, Jadwiga Bienkowska, James Hardwick, and Tao Xie. Tner: a novel background error suppression method for mutation detection in circulating tumor dna. *BMC bioinformatics*, 19(1):387, 2018.
- [113] Sebastian Deorowicz, Agnieszka Danek, and Marcin Niemiec. Gdc 2: Compression of large collections of genomes. *Scientific reports*, 5:11565, 2015.
- [114] Sebastian Deorowicz, Agnieszka Debudaj-Grabysz, Adam Gudyś, and Szymon Grabowski. Whisper: read sorting allows robust mapping of dna sequencing data. *Bioinformatics*, 2018.
- [115] Sebastian Deorowicz, Joanna Walczyszyn, and Agnieszka Debudaj-Grabysz. Msac: Compression of multiple sequence alignment files. *bioRxiv*, page 240341, 2017.
- [116] Sebastian Deorowicz, Joanna Walczyszyn, and Agnieszka Debudaj-Grabysz. Comsa: compression of protein multiple sequence alignment files. *Bioinformatics*, 35(2):227–234, 2018.
- [117] Jean-Pierre Desvignes, Marc Bartoli, Valérie Delague, Martin Krahn, Morgane Miltgen, Christophe Bérout, and David Salgado. Varaft: a variant annotation and filtration system for human next generation sequencing data. *Nucleic acids research*, 46(W1):W545–W553, 2018.

-
- [118] Paolo Di Tommaso, Maria Chatzou, Evan W Floden, Pablo Prieto Barja, Emilio Palumbo, and Cedric Notredame. Nextflow enables reproducible computational workflows. *Nature biotechnology*, 35(4):316, 2017.
- [119] John P Didion, Marcel Martin, and Francis S Collins. Atropos: specific, sensitive, and speedy trimming of sequencing reads. *PeerJ*, 5:e3720, 2017.
- [120] Maciej Długosz and Sebastian Deorowicz. Reckoner: read error corrector based on kmc. *Bioinformatics*, 33(7):1086–1089, 2016.
- [121] Kenneth D Doig, Jason Ellul, Andrew Fellowes, Ella R Thompson, Georgina Ryland, Piers Blombery, Anthony T Papenfuss, and Stephen B Fox. Canary: an atomic pipeline for clinical amplicon assays. *BMC bioinformatics*, 18(1):555, 2017.
- [122] Daniel Domingo-Fernandez, Sarah Mubeen, Josep Marin-Llao, Charles Hoyt, and Martin Hofmann-Apitius. Pathme: Merging and exploring mechanistic pathway knowledge. *bioRxiv*, page 451625, 2019.
- [123] Mikhail G Dozmorov, Indra Adrianto, Cory B Giles, Edmund Glass, Stuart B Glenn, Courtney Montgomery, Kathy L Sivils, Lorin E Olson, Tomoaki Iwayama, Willard M Freeman, et al. Detrimental effects of duplicate reads and low complexity regions on rna-and chip-seq data. In *BMC bioinformatics*, volume 16, page S10. BioMed Central, 2015.
- [124] Sverre-Arne Dragly, Milad Hobbi Mobarhan, Mikkel E Lepperød, Simen Tennøe, Marianne Fyhn, Torkel Hafting, and Anders Malthe-Sørensen. Experimental directory structure (exdir): An alternative to hdf5 without introducing a new file format. *Frontiers in neuroinformatics*, 12, 2018.
- [125] Alastair P Droop. fqtools: an efficient software suite for modern fastq file manipulation. *Bioinformatics*, 32(12):1883–1884, 2016.
- [126] Huilong Du, Ying Yu, Yanfei Ma, Qiang Gao, Yinghao Cao, Zhuo Chen, Bin Ma, Ming Qi, Yan Li, Xianfeng Zhao, et al. Sequencing and de novo assembly of a near complete indica rice genome. *Nature communications*, 8:15324, 2017.
- [127] Nan Du and Yanni Sun. Improve homology search sensitivity of pacbio data by correcting frameshifts. *Bioinformatics*, 32(17):i529–i537, 2016.
- [128] Mark TW Ebbert, Mark E Wadsworth, Kevin L Boehme, Kaitlyn L Hoyt, Aaron R Sharp, Brendan D O’Fallon, John SK Kauwe, and Perry G Ridge. Variant tool chest: an improved tool to analyze and manipulate variant call format (vcf) files. *BMC bioinformatics*, 15(7):S12, 2014.
- [129] Robert C Edgar and Henrik Flyvbjerg. Error filtering, pair assembly and error correction for next-generation sequencing reads. *Bioinformatics*, 31(21):3476–3482, 2015.

-
- [130] Daniel PW Ellis. Beat tracking by dynamic programming. *Journal of New Music Research*, 36(1):51–60, 2007.
- [131] Dei M Elurbe, Sarita S Paranjpe, Georgios Georgiou, Ila Van Kruijsbergen, Ozren Bogdanovic, Romain Gibeaux, Rebecca Heald, Ryan Lister, Martijn A Huynen, Simon J Van Heeringen, et al. Regulatory remodeling in the allo-tetraploid frog *xenopus laevis*. *Genome biology*, 18(1):198, 2017.
- [132] A Murat Eren, Joseph H Vineis, Hilary G Morrison, and Mitchell L Sogin. A filtering method to generate high quality short reads using illumina paired-end technology. *PloS one*, 8(6):e66643, 2013.
- [133] Amanda W Ernlund, Robert J Schneider, and Kelly V Ruggles. Rivet: comprehensive graphic user interface for analysis and exploration of genome-wide translatomics data. *BMC genomics*, 19(1):809, 2018.
- [134] Philippe Esling, Franck Lejzerowicz, and Jan Pawlowski. Accurate multiplexing and filtering for high-throughput amplicon-sequencing. *Nucleic acids research*, 43(5):2513–2524, 2015.
- [135] Ruben Esse and Alla Grishok. Chipdig: a comprehensive user-friendly tool for mining multi-sample chip-seq data. *bioRxiv*, page 220079, 2017.
- [136] Graham J Etherington, Ricardo H Ramirez-Gonzalez, and Dan MacLean. biosamtools 2: a package for analysis and visualization of sequence and alignment data with samtools in ruby. *Bioinformatics*, 31(15):2565–2567, 2015.
- [137] Philip Ewels, Måns Magnusson, Sverker Lundin, and Max Käller. Multiqc: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19):3047–3048, 2016.
- [138] Brant C Faircloth. Phyluce is a software package for the analysis of conserved genomic loci. *Bioinformatics*, 32(5):786–788, 2015.
- [139] Wei Fan, Jie Zong, Zhijing Luo, Mingjiao Chen, Xiangxiang Zhao, Dabing Zhang, Yiping Qi, and Zheng Yuan. Development of a rad-seq based dna polymorphism identification software, agromarker finder, and its application in rice marker-assisted breeding. *PloS one*, 11(1):e0147187, 2016.
- [140] Tondini Federico, Jiranek Vladimir, Cristobal A Onetto, et al. Genome sequence of australian indigenous wine yeast *torulaspora delbrueckii* coft1 using nanopore sequencing. *Genome Announcements*, 2018.
- [141] Shihai Feng, Chien-Chi Lo, Po-E Li, and Patrick SG Chain. Adept, a dynamic next generation sequencing data error-detection program with trimming. *BMC bioinformatics*, 17(1):109, 2016.
- [142] Edward B Fernandez, Jason Villarreal, Stefano Lonardi, and Walid A Najjar. Fhast: Fpga-based acceleration of bowtie in hardware. *IEEE/ACM transactions on computational biology and bioinformatics*, 12(5):973–981, 2015.

-
- [143] Umberto Ferraro Petrillo, Gianluca Roscigno, Giuseppe Cattaneo, and Raffaele Giancarlo. Fastdoop: a versatile and efficient library for the input of fasta and fastq files for mapreduce hadoop bioinformatics applications. *Bioinformatics*, 33(10):1575–1577, 2017.
- [144] Stephen J Fletcher, Mikael Boden, Neena Mitter, and Bernard J Carroll. Scram: a pipeline for fast index-free small rna read alignment and visualization. *Bioinformatics*, 34(15):2670–2672, 2018.
- [145] Ryan A Flynn, Lance Martin, Robert C Spitale, Brian T Do, Selena M Sagan, Brian Zarnegar, Kun Qu, Paul A Khavari, Stephen R Quake, Peter Sarnow, et al. Dissecting noncoding and pathogen rna–protein interactomes. *Rna*, 21(1):135–143, 2015.
- [146] Felix Francis, Michael D Dumas, Scott B Davis, and Randall J Wisser. Clustering of circular consensus sequences: accurate error correction and assembly of single molecule real-time reads from multiplexed amplicon libraries. *BMC bioinformatics*, 19(1):302, 2018.
- [147] Peter J Freeman, Reece K Hart, Liam J Gretton, Anthony J Brookes, and Raymond Dalgleish. Variantvalidator: Accurate validation, mapping, and formatting of sequence variation descriptions. *Human mutation*, 39(1):61–68, 2018.
- [148] Tobias Guldberg Frøslev, Rasmus Kjølner, Hans Henrik Bruun, Rasmus Ejrnæs, Ane Kirstine Brunbjerg, Carlotta Pietroni, and Anders Johannes Hansen. Algorithm for post-clustering curation of dna amplicon data yields reliable biodiversity estimates. *Nature communications*, 8(1):1188, 2017.
- [149] Jiabing Fu, Yacong Ma, Bixin Ke, and Shoubin Dong. Lctd: A lossless compression tool of fastq file based on transformation of original file distribution. In *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 864–869. IEEE, 2016.
- [150] Dimos Gaidatzis, Anita Lerch, Florian Hahne, and Michael B Stadler. Quasr: quantification and annotation of short reads in r. *Bioinformatics*, 31(7):1130–1132, 2014.
- [151] Lior Galanti, Dennis Shasha, and Kristin Gunsalus. Pheniqs: Fast and flexible quality-aware sequence demultiplexing. *bioRxiv*, page 128512, 2017.
- [152] Prakhar Ganesh, Gaurav Gupta, Shubhi Saini, and Kolin Paul. Nucl2vec: Local alignment of dna sequences using distributed vector representation. *BioRxiv*, page 401851, 2018.
- [153] Feng Gao, Diana Chang, Arjun Biddanda, Li Ma, Yingjie Guo, Zilu Zhou, and Alon Keinan. Xwas: a software toolset for genetic data analysis and association studies of the x chromosome. *Journal of Heredity*, 106(5):666–671, 2015.

-
- [154] Xin Gao, Rahul Tyagi, Vincent Magrini, Amy Ly, Douglas P Jasmer, and Make-donka Mitreva. Compartmentalization of functions and predicted mirna regulation among contiguous regions of the nematode intestine. *RNA biology*, 14(10):1335–1352, 2017.
- [155] Owen Garland, Amanda Clare, and Wayne Aubrey. Giraffe browse: A lightweight web based tool for inspecting gff and fasta data. *BioRxiv*, page 273631, 2018.
- [156] Bram Gerritsen, Aridaman Pandit, Arno C Andeweg, and Rob J De Boer. Rctr: a pipeline for complete and accurate recovery of t cell repertoires from high throughput sequencing data. *Bioinformatics*, 32(20):3098–3106, 2016.
- [157] Felipe C Geyer, Anqi Li, Anastasios D Papanastasiou, Alison Smith, Pier Selenica, Kathleen A Burke, Marcia Edelweiss, Huei-Chi Wen, Salvatore Piscuoglio, Anne M Schultheis, et al. Recurrent hotspot mutations in hras q61 and pi3k-akt pathway genes as drivers of breast adenomyoepitheliomas. *Nature communications*, 9(1):1816, 2018.
- [158] Filippo Giannetti, Ruggero Reggiannini, Marco Moretti, Elisa Adirosi, Luca Baldini, Luca Facheris, Andrea Antonini, Samantha Melani, Giacomo Bacci, Antonio Petrolino, et al. Real-time rain rate evaluation via satellite downlink signal attenuation measurement. *Sensors*, 17(8):1864, 2017.
- [159] Christopher M Gibb, Robert Jackson, Sabah Mohammed, Jinan Fiaidhi, and Ingeborg Zehbe. Pathogen–host analysis tool (phat): an integrative platform to analyze next-generation sequencing data. *Bioinformatics*, 2018.
- [160] Scott Gigante. Picopore: a tool for reducing the storage size of oxford nanopore technologies datasets without loss of functionality. *F1000Research*, 6, 2017.
- [161] Christopher E Gillies, Edgar A Otto, Virginia Vega-Warner, Catherine C Robertson, Simone Sanna-Cherchi, Ali Gharavi, Brendan Crawford, Rajendra Bhimma, Cheryl Winkler, Hyun Min Kang, et al. tarsvm: Improving the accuracy of variant calls derived from microfluidic pcr-based targeted next generation sequencing using a support vector machine. *BMC bioinformatics*, 17(1):233, 2016.
- [162] Frank Oliver Glöckner, Pelin Yilmaz, Christian Quast, Jan Gerken, Alan Beccati, Andreea Ciuprina, Gerrit Bruns, Pablo Yarza, Jörg Peplies, Ralf Westram, et al. 25 years of serving the community with ribosomal rna gene reference databases and tools. *Journal of biotechnology*, 261:169–176, 2017.
- [163] Giorgio Gonnella and Stefan Kurtz. Gfapy: a flexible and extensible software library for handling sequence graphs in python. *Bioinformatics*, 33(19):3094–3095, 2017.
- [164] Jorge González-Domínguez, Christian Hundt, and Bertil Schmidt. parsra: a framework for the parallel execution of short read aligners on compute clusters. *Journal of Computational Science*, 25:134–139, 2018.

-
- [165] Jorge Gonzalez-Dominguez, Yongchao Liu, and Bertil Schmidt. Parallel and scalable short-read alignment on multi-core clusters using upc++. *PloS one*, 11(1):e0145490, 2016.
- [166] Jorge González-Domínguez and Bertil Schmidt. Pardre: faster parallel duplicated reads removal tool for sequencing studies. *Bioinformatics*, 32(10):1562–1564, 2016.
- [167] Cene Gostinčar, Jason E Stajich, Jerneja Zupančič, Polona Zalar, and Nina Gunde-Cimerman. Genomic evidence for intraspecific hybridization in a clonal and extremely halotolerant yeast. *BMC genomics*, 19(1):364, 2018.
- [168] Szymon Grabowski, Sebastian Deorowicz, and Łukasz Roguski. Disk-based compression of data from genome sequencing. *Bioinformatics*, 31(9):1389–1395, 2014.
- [169] Daniel L Greenfield, Oliver Stegle, and Alban Rrustemi. Genecodeq: quality score compression and improved genotyping using a bayesian framework. *Bioinformatics*, 32(20):3124–3132, 2016.
- [170] William W Greenwald, He Li, Erin N Smith, Paola Benaglio, Naoki Nariai, and Kelly A Frazer. Pgltools: a genomic arithmetic tool suite for manipulation of hi-c peak and other chromatin interaction data. *BMC bioinformatics*, 18(1):207, 2017.
- [171] Michael Gruenstaeudl and Yannick Hartmaring. Embl2checklists: A python package to facilitate the user-friendly submission of plant dna barcoding sequences to ena. *bioRxiv*, page 435644, 2018.
- [172] Björn A Grüning, Jörg Fallmann, Dilmurat Yusuf, Sebastian Will, Anika Erxleben, Florian Eggenhofer, Torsten Houwaart, Bérénice Batut, Pavankumar Videm, Andrea Bagnacani, et al. The rna workbench: best practices for rna and high-throughput sequencing bioinformatics in galaxy. *Nucleic acids research*, 45(W1):W560–W566, 2017.
- [173] Yarong Gu, Xianying Liu, Qiang Zhu, Youchao Dong, C Titus Brown, and Sakti Pramanik. A new method for dna sequencing error verification and correction via an on-disk index tree. In *Proceedings of the 6th ACM Conference on Bioinformatics, Computational Biology and Health Informatics*, pages 503–504. ACM, 2015.
- [174] August Guang, Mark Howison, Felipe Zapata, Charles E Lawrence, and Casey Dunn. Revising transcriptome assemblies with phylogenetic information in agalma1. 0. *bioRxiv*, page 202416, 2017.
- [175] Giulia Guidi, Marquita Ellis, Daniel Rokhsar, Katherine Yelick, and Aydın Buluç. Bella: Berkeley efficient long-read to long-read aligner and overlapper. *bioRxiv*, page 464420, 2018.
- [176] Hongzhe Guo, Yilei Fu, Yan Gao, Junyi Li, Yadong Wang, and Bo Liu. degsm: memory scalable construction of large scale de bruijn graph. *bioRxiv*, page 388454, 2018.

-
- [177] Jingtao Guo, Edward J Grow, Chongil Yi, Hana Mlcochova, Geoffrey J Maher, Cecilia Lindskog, Patrick J Murphy, Candice L Wike, Douglas T Carrell, Anne Goriely, et al. Chromatin and single-cell rna-seq profiling reveal dynamic signaling and metabolic transitions during human spermatogonial stem cell development. *Cell Stem Cell*, 21(4):533–546, 2017.
- [178] Ehsan Haghshenas, Faraz Hach, S Cenk Sahinalp, and Cedric Chauve. Colormap: Correcting long reads by mapping short reads. *Bioinformatics*, 32(17):i545–i551, 2016.
- [179] Ehsan Haghshenas, S Cenk Sahinalp, and Faraz Hach. lordfast: sensitive and fast alignment search tool for long noisy read sequencing data. *Bioinformatics*, 35(1):20–27, 2018.
- [180] Syed Haider, Daryl Waggott, Emilie Lalonde, Clement Fung, Fei-Fei Liu, and Paul C Boutros. A bedr way of genomic interval processing. *Source code for biology and medicine*, 11(1):14, 2016.
- [181] Nikola Hajkova, Jan Hojny, Kristyna Nemejcova, Pavel Dundr, Jan Ulrych, Katerina Jirsova, Johana Glezgova, and Ivana Ticha. Germline mutation in the tp53 gene in uveal melanoma. *Scientific reports*, 8(1):7618, 2018.
- [182] Bulut Hamali, Sandra Pichler, Elisabeth Wischnitzki, Klaus Schicker, Melanie Burger, Marion Holy, Kathrin Jaentsch, Martina Molin, Eva Maria Sehr, Oliver Kudlacek, et al. Identification and characterization of the fasciola hepatica sodium- and chloride-dependent taurine transporter. *PLoS neglected tropical diseases*, 12(4):e0006428, 2018.
- [183] Bo W Han, Wei Wang, Phillip D Zamore, and Zhiping Weng. pipipes: a set of pipelines for pirna and transposon analysis via small rna-seq, rna-seq, degradome- and cage-seq, chip-seq and genomic dna sequencing. *Bioinformatics*, 31(4):593–595, 2014.
- [184] Ulrika A Hänninen, Riku Katainen, Tomas Tanskanen, Roosa-Maria Plaketti, Riku Laine, Jiri Hamberg, Ari Ristimäki, Eero Pukkala, Minna Taipale, Jukka-Pekka Mecklin, et al. Exome-wide somatic mutation characterization of small bowel adenocarcinoma. *PLoS genetics*, 14(3):e1007200, 2018.
- [185] Andrew Hart, María Paz Cortés, Mauricio Latorre, and Servet Martinez. Codon usage bias reveals genomic adaptations to environmental conditions in an acidophilic consortium. *PloS one*, 13(5):e0195869, 2018.
- [186] Junko Hasegawa, Takuya Sakamoto, Satoru Fujimoto, Tomoe Yamashita, Takamasa Suzuki, and Sachihito Matsunaga. Auxin decreases chromatin accessibility through the tir1/afb1 auxin signaling pathway in proliferative cells. *Scientific reports*, 8(1):7773, 2018.

-
- [187] Masaomi Hatakeyama, Lennart Opitz, Giancarlo Russo, Weihong Qi, Ralph Schlapbach, and Hubert Rehrauer. Sushi: an exquisite recipe for fully documented, reproducible and reusable ngs data analysis. *BMC bioinformatics*, 17(1):228, 2016.
- [188] Nicholas J Hathaway, Christian M Parobek, Jonathan J Juliano, and Jeffrey A Bailey. Seekdeep: single-base resolution de novo clustering for amplicon deep sequencing. *Nucleic acids research*, 46(4):e21–e21, 2017.
- [189] John A Hawkins, Stephen K Jones, Ilya J Finkelstein, and William H Press. Error-correcting dna barcodes for high-throughput sequencing. *bioRxiv*, page 315002, 2018.
- [190] Tracy H Hazen, Sean C Daugherty, Amol C Shetty, James P Nataro, and David A Rasko. Transcriptional variation of diverse enteropathogenic escherichia coli isolates under virulence-inducing conditions. *MSystems*, 2(4):e00024–17, 2017.
- [191] Wen He, Shanrong Zhao, Chi Zhang, Michael S Vincent, and Baohong Zhang. Quickrna-seq: Guide for pipeline implementation and for interactive results visualization. In *Transcriptome Data Analysis*, pages 57–70. Springer, 2018.
- [192] Darren Heavens, Gonzalo Garcia Accinelli, Bernardo Clavijo, and Matthew Derek Clark. A method to simultaneously construct up to 12 differently sized illumina nextera long mate pair libraries with reduced dna input, time, and cost. *Biotechniques*, 59(1):42–45, 2015.
- [193] Jacob R Heldenbrand, Saurabh Baheti, Matthew A Bockol, Travis M Drucker, Steven N Hart, Matthew E Hudson, Ravishankar K Iyer, Michael T Kalmbach, Eric W Klee, Eric D Wieben, et al. Performance benchmarking of gatk3. 8 and gatk4. *BioRxiv*, page 348565, 2018.
- [194] Yun Heo, Anand Ramachandran, Wen-Mei Hwu, Jian Ma, and Deming Chen. Bless 2: accurate, memory-efficient and fast error correction method. *Bioinformatics*, 32(15):2369–2371, 2016.
- [195] Charlotte Herzeel, Pascal Costanza, Dries Decap, Jan Fostier, and Wilfried Vercantert. elprep 4: A multithreaded framework for sequence analysis. *PloS one*, 14(2):e0209523, 2019.
- [196] Mareike Herzog, Fabio Puddu, Julia Coates, Nicola Geisler, Josep V Forment, and Stephen P Jackson. Detection of functional protein domains by unbiased genome-wide forward genetic screening. *Scientific reports*, 8(1):6161, 2018.
- [197] Steve Hoffmann, Peter F Stadler, and Korbinian Strimmer. A simple data-adaptive probabilistic variant calling model. *Algorithms for Molecular Biology*, 10(1):10, 2015.
- [198] Markus Hollander, Mohamed Hamed, Volkhard Helms, and Kerstin Neininger. Mutanet: a tool for automated analysis of genomic mutations in gene regulatory networks. *Bioinformatics*, 34(5):864–866, 2017.

-
- [199] Mark Howison, Mia Coetzer, and Rami Kantor. Measurement error and variant-calling in deep illumina sequencing of hiv. *bioRxiv*, page 276576, 2018.
- [200] Hailin Hu, Xianggen Liu, An Xiao, Sen Song, and Jianyang Zeng. Rationalizing translation elongation by reinforcement learning. *bioRxiv*, page 463976, 2018.
- [201] Yao-Ting Huang and Yu-Wen Huang. An efficient error correction algorithm using fm-index. *BMC bioinformatics*, 18(1):524, 2017.
- [202] Zhi-An Huang, Zhenkun Wen, Qingjin Deng, Ying Chu, Yiwen Sun, and Zexuan Zhu. Lw-fqzip 2: a parallelized reference-based compression of fastq files. *BMC bioinformatics*, 18(1):179, 2017.
- [203] Zhicong Huang, Erman Ayday, Huang Lin, Raeka S Aiyar, Adam Molyneaux, Zhenyu Xu, Jacques Fellay, Lars M Steinmetz, and Jean-Pierre Hubaux. A privacy-preserving solution for compressed storage and selective retrieval of genomic data. *Genome research*, 26(12):1687–1696, 2016.
- [204] Graham M Hughes and Emma C Teeling. Agile: an assembled genome mining pipeline. *Bioinformatics*, 2018.
- [205] Katherine Icaý, Ping Chen, Alejandra Cervera, Ville Rantanen, Rainer Lehtonen, and Sampsa Hautaniemi. Sepia: Rna and small rna sequence processing, integration, and analysis. *BioData mining*, 9(1):20, 2016.
- [206] Adriana I Iglesias, Aniket Mishra, Veronique Vitart, Yelena Bykhovskaya, René Höhn, Henriët Springelkamp, Gabriel Cuellar-Partida, Puya Gharahkhani, Jessica N Cooke Bailey, Colin E Willoughby, et al. Cross-ancestry genome-wide association analysis of corneal thickness strengthens link between complex and mendelian eye diseases. *Nature communications*, 9(1):1864, 2018.
- [207] Danny Incarnato, Edoardo Morandi, Lisa Marie Simon, and Salvatore Oliviero. Rna framework: an all-in-one toolkit for the analysis of rna structures and post-transcriptional modifications. *Nucleic acids research*, 46(16):e97–e97, 2018.
- [208] Ehtesham Iqbal, Robbie Mallah, Daniel Rhodes, Honghan Wu, Alvin Romero, Nynn Chang, Olubanke Dzahini, Chandra Pandey, Matthew Broadbent, Robert Stewart, et al. Adept, a semantically-enriched pipeline for extracting adverse drug events from free-text electronic health records. *PloS one*, 12(11):e0187121, 2017.
- [209] Chirag Jain, Alexander Dilthey, Sergey Koren, Srinivas Aluru, and Adam M Phillippy. A fast approximate algorithm for mapping long reads to large reference databases. In *International Conference on Research in Computational Molecular Biology*, pages 66–81. Springer, 2017.
- [210] Miten Jain, Ian T Fiddes, Karen H Miga, Hugh E Olsen, Benedict Paten, and Mark Akeson. Improved data analysis for the minion nanopore sequencer. *Nature methods*, 12(4):351, 2015.

-
- [211] Miten Jain, Sergey Koren, Karen H Miga, Josh Quick, Arthur C Rand, Thomas A Sasani, John R Tyson, Andrew D Beggs, Alexander T Dilthey, Ian T Fiddes, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature biotechnology*, 36(4):338, 2018.
- [212] Benjamin T James, Brian B Luczak, and Hani Z Girgis. Fastcar: Rapid alignment-free prediction of sequence alignment identity scores. *bioRxiv*, page 380824, 2018.
- [213] Peter Javorka, Vivek Kumar Raxwal, Jan Najvarek, and Karel Riha. artmap: a user-friendly tool for mapping ems-induced mutations in arabidopsis. *bioRxiv*, page 414433, 2018.
- [214] Zhe Ji. Rfoot: Transcriptome-scale identification of rna-protein complexes from ribosome profiling data. *Current protocols in molecular biology*, 124(1):e66, 2018.
- [215] Zhe Ji, Ruisheng Song, Hailiang Huang, Aviv Regev, and Kevin Struhl. Transcriptome-scale rnase-footprinting of rna-protein complexes. *Nature biotechnology*, 34(4):410, 2016.
- [216] Benjamin K Johnson, Matthew B Scholz, Tracy K Teal, and Robert B Abramovitch. Sparta: Simple program for automated reference-based bacterial rna-seq transcriptome analysis. *BMC bioinformatics*, 17(1):66, 2016.
- [217] Lisa Kristine Johnson, Harriet Alexander, and C Titus Brown. Re-assembly, quality evaluation, and annotation of 678 microbial eukaryotic reference transcriptomes. *BioRxiv*, page 323576, 2018.
- [218] H Richard Johnston, Pankaj Chopra, Thomas S Wingo, Viren Patel, Michael P Epstein, Jennifer G Mulle, Stephen T Warren, Michael E Zwick, David J Cutler, et al. Pemapper and pecaller provide a simplified approach to whole-genome sequencing. *Proceedings of the National Academy of Sciences*, 114(10):E1923–E1932, 2017.
- [219] Goo Jun, Mary Kate Wing, Gonçalo R Abecasis, and Hyun Min Kang. An efficient and scalable analysis framework for variant extraction and refinement from population-scale dna sequence data. *Genome Research*, 25(6):918–925, 2015.
- [220] Yuki Kagoya, Munehide Nakatsugawa, Kayoko Saso, Tingxi Guo, Mark Anczuroska, Chung-Hsi Wang, Marcus O Butler, Cheryl H Arrowsmith, and Naoto Hirano. Dot1l inhibition attenuates graft-versus-host disease by allogeneic t cells in adoptive immunotherapy models. *Nature communications*, 9(1):1915, 2018.
- [221] Wolfgang Kaisers, Heiner Schaal, and Holger Schwender. rbamtools: an r interface to samtools enabling fast accumulative tabulation of splicing events over multiple rna-seq samples. *Bioinformatics*, 31(10):1663–1664, 2015.
- [222] Sergey Kalinin, Marta González-Prieto, Hannah Scheiblich, Lucia Lisi, Handojo Kusumo, Michael T Heneka, Jose LM Madrigal, Subhash C Pandey, and Douglas L

- Feinstein. Transcriptome analysis of alcohol-treated microglia reveals downregulation of beta amyloid phagocytosis. *Journal of neuroinflammation*, 15(1):141, 2018.
- [223] Chakravarthi Kanduri, Diana Domanska, Eivind Hovig, and Geir Kjetil Sandve. Genome build information is an essential part of genomic track files. *Genome biology*, 18(1):175, 2017.
- [224] Sreeram Kannan, Joseph Hui, Kayvon Mazooji, Lior Pachter, and David Tse. Shannon: An information-optimal de novo rna-seq assembler. *bioRxiv*, page 039230, 2016.
- [225] Mehran Karimzadeh, Carl Ernst, Anshul Kundaje, and Michael M Hoffman. Umap and bismap: quantifying genome and methylome mappability. *Nucleic acids research*, 46(20):e120–e120, 2018.
- [226] Riku Katainen, Iikki Donner, Tatiana Cajuso, Eevi Kaasinen, Kimmo Palin, Veli Mäkinen, Lauri A Aaltonen, and Esa Pitkänen. Discovery of potential causative mutations in human coding and noncoding genome with the interactive software baseplayer. *Nature protocols*, 13(11):2580, 2018.
- [227] Akhilesh Kaushal, Hongmei Zhang, Wilfried JJ Karmaus, and Julie SL Wang. Which methods to choose to correct cell types in genome-scale blood-derived dna methylation data? *BMC bioinformatics*, 16(15):P7, 2015.
- [228] Mehdi Kchouk and Mourad Elloumi. Hybrid error correction approach and de novo assembly for minion sequencing long reads. In *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 122–125. IEEE, 2016.
- [229] Andrey Kechin, Uljana Boyarskikh, Alexander Kel, and Maxim Filipenko. cut-primers: A new tool for accurate cutting of primers from reads of targeted next generation sequencing. *Journal of Computational Biology*, 24(11):1138–1143, 2017.
- [230] Birte Kehr and Páll Melsted. chopbai: Bam index reduction solves i/o bottlenecks in the joint analysis of large sequencing cohorts. *Bioinformatics*, 32(14):2202–2204, 2016.
- [231] Nilesh Khiste and Lucian Ilie. Hisea: Hierarchical seed aligner for pacbio data. *BMC bioinformatics*, 18(1):564, 2017.
- [232] Silas Kieser, Shafiqul A Sarker, Bernard Berger, Shamima Sultana, Mohammad J Chisti, Shoeb B Islam, Francis Foata, Nadine Porta, Bertrand Betrisey, Coralie Fournier, et al. Antibiotic treatment leads to fecal escherichia coli and coliphage expansion in severely malnourished diarrhea patients. *Cellular and molecular gastroenterology and hepatology*, 5(3):458–460, 2018.

- [233] Chang Sik Kim, Martyn D Winn, Vipin Sachdeva, and Kirk E Jordan. K-mer clustering algorithm using a mapreduce framework: application to the parallelization of the inchworm module of trinity. *BMC bioinformatics*, 18(1):467, 2017.
- [234] Taemook Kim, Hogyu David Seo, Lothar Hennighausen, Daeyoup Lee, and Keunsoo Kang. Octopus-toolkit: a workflow to automate mining of public epigenomic and transcriptomic next-generation sequencing data. *Nucleic acids research*, 46(9):e53–e53, 2018.
- [235] Zachary A King, Andreas Dräger, Ali Ebrahim, Nikolaus Sonnenschein, Nathan E Lewis, and Bernhard O Palsson. Escher: a web application for building, sharing, and embedding data-rich visualizations of biological pathways. *PLoS computational biology*, 11(8):e1004321, 2015.
- [236] Una Kjällquist, Rikard Erlandsson, Nicholas P Tobin, Amjad Alkodsí, Ikram Ullah, Gustav Stålhammar, Eva Karlsson, Thomas Hatschek, Johan Hartman, Sten Linnarsson, et al. Exome sequencing of primary breast cancers with paired metastatic lesions reveals metastasis-enriched mutations in the a-kinase anchoring protein family (akaps). *BMC cancer*, 18(1):174, 2018.
- [237] Wanlada Klangnarak, Taketo Fukuyo, MD Rezanujjaman, Masahide Seki, Sumio Sugano, Yutaka Suzuki, and Toshinobu Tokumoto. Candidate gene identification of ovulation-inducing genes by rna sequencing with an in vivo assay in zebrafish. *PloS one*, 13(5):e0196544, 2018.
- [238] Roelof JC Kluin, Kristel Kemper, Thomas Kuilman, Julian R de Ruiter, Vivek Iyer, Josep V Forment, Paulien Cornelissen-Steijger, Iris de Rink, Petra ter Brugge, Ji-Ying Song, et al. Xenofilter: computational deconvolution of mouse and human reads in tumor xenograft sequence data. *BMC bioinformatics*, 19(1):366, 2018.
- [239] Alex V Kochetov, Jens Allmer, Alexandra I Klimenko, Bulat S Zuraev, Yury G Matushkin, and Sergey A Lashin. Altorfex facilitates the prediction of alternative open reading frames in eukaryotic mrnas. *Bioinformatics*, 33(6):923–925, 2016.
- [240] Malgorzata A Komor, Thang V Pham, Annemieke C Hiemstra, Sander R Piersma, Anne S Bolijn, Tim Schelfhorst, Pien M Delis-van Diemen, Marianne Tijssen, Robert P Sebra, Meredith Ashby, et al. Identification of differentially expressed splice variants by the proteogenomic pipeline splicify. *Molecular & Cellular Proteomics*, 16(10):1850–1863, 2017.
- [241] Sergey Koren, Arang Rhie, Brian P Walenz, Alexander T Dilthey, Derek M Bickhart, Sarah B Kingan, Stefan Hiendleder, John L Williams, Timothy PL Smith, and Adam Phillippy. Complete assembly of parental haplotypes with trio binning. *BioRxiv*, page 271486, 2018.
- [242] Etienne Kornobis, Luis Cabellos, Fernando Aguilar, Cristina Frías-López, Julio Rozas, Jesús Marco, and Rafael Zardoya. Trufa: a user-friendly web server for

- p>de novo rna-seq analysis using cluster computing.
- Evolutionary Bioinformatics*
- , 11:EBO-S23873, 2015.
- [243] Karl W Kroll, Ann-Katherin Eisfeld, Gerard Lozanski, Clara D Bloomfield, John C Byrd, and James S Blachly. Mucor: mutation aggregation and correlation. *Bioinformatics*, 32(10):1557–1558, 2016.
- [244] Alan Kuhnle, Taher Mun, Christina Boucher, Travis Gagie, Ben Langmead, and Giovanni Manzini. Efficient construction of a complete index for pan-genomics read alignment. *arXiv preprint arXiv:1811.06933*, 2018.
- [245] M Oğuzhan Külekci, Ali Fotouhi, and Mina Majidi. Quality assessment of high-throughput dna sequencing data via range analysis. *bioRxiv*, page 101469, 2017.
- [246] Sean La, Ehsan Haghshenas, and Cedric Chauve. Lrcstats, a tool for evaluating long reads correction methods. *Bioinformatics*, 33(22):3652–3654, 2017.
- [247] William B Langdon and Brian Yee Hong Lam. Genetically improved barracuda. *BioData mining*, 10(1):28, 2017.
- [248] Ben Langmead. A tandem simulation framework for predicting mapping quality. *Genome biology*, 18(1):152, 2017.
- [249] Nathan LaPierre, Rob Egan, Wei Wang, and Zhong Wang. Miniscrub: de novo long read scrubbing using approximate alignment and deep learning. *BioRxiv*, page 433573, 2018.
- [250] Timo Lassmann. Tagdust2: a generic method to extract reads from sequencing data. *BMC bioinformatics*, 16(1):24, 2015.
- [251] Fabio Lauria, Toma Tebaldi, Paola Bernabò, Ewout JN Groen, Thomas H Gillingwater, and Gabriella Viero. ribowaltz: optimization of ribosome p-site positioning in ribosome profiling data. *PLoS computational biology*, 14(8):e1006169, 2018.
- [252] Christopher A Lavender, Andrew J Shapiro, Adam B Burkholder, Brian D Bennett, Karen Adelman, and David C Fargo. Orio (online resource for integrative omics): a web-based platform for rapid integration of next generation sequencing data. *Nucleic acids research*, 45(10):5678–5690, 2017.
- [253] Michael Lawrence and Robert Gentleman. Varianttools: an extensible framework for developing and testing variant callers. *Bioinformatics*, 33(20):3311–3313, 2017.
- [254] Travis J Lawrence, Kyle T Kauffman, Katherine CH Amrine, Dana L Carper, Raymond S Lee, Peter J Becich, Claudia J Canales, and David H Ardell. Fast: Fast analysis of sequences toolbox. *Frontiers in genetics*, 6:172, 2015.
- [255] Ryan M Layer, Neil Kindlon, Konrad J Karczewski, Aaron R Quinlan, Exome Aggregation Consortium, et al. Efficient genotype compression and analysis of large genetic-variation data sets. *Nature methods*, 13(1):63, 2016.

-
- [256] Ryan M Layer, Brent S Pedersen, Tonya DiSera, Gabor T Marth, Jason Gertz, and Aaron R Quinlan. Giggle: a search engine for large-scale integrated genome analysis. *Nature methods*, 15(2):123, 2018.
- [257] Byunghan Lee, Taesup Moon, Sungroh Yoon, and Tsachy Weissman. Dude-seq: Fast, flexible, and robust denoising for targeted amplicon sequencing. *PloS one*, 12(7):e0181463, 2017.
- [258] Donghyuk Lee, Farhad Hormozdiari, Hongyi Xin, Faraz Hach, Onur Mutlu, and Can Alkan. Fast and accurate mapping of complete genomics reads. *Methods*, 79:3–10, 2015.
- [259] Richard M Leggett, Darren Heavens, Mario Caccamo, Matthew D Clark, and Robert P Davey. Nanook: multi-reference alignment analysis of nanopore sequencing data, quality and error profiles. *Bioinformatics*, 32(1):142–144, 2015.
- [260] Emmanuelle Lerat, Marie Fablet, Laurent Modolo, H  lene Lopez-Maestre, and Cristina Vieira. Tertools facilitates big data expression analysis of transposable elements and reveals an antagonism between their activity and that of pirna genes. *Nucleic acids research*, 45(4):e17–e17, 2016.
- [261] Matthew S Leslie and Phillip A Morin. Structure and phylogeography of two tropical predators, spinner (*stenella longirostris*) and pantropical spotted (s. *attenuata*) dolphins, from snp data. *Royal Society open science*, 5(4):171615, 2018.
- [262] Heng Li. Fast construction of fm-index for long sequence reads. *Bioinformatics*, 30(22):3274–3275, 2014.
- [263] Heng Li. Bfc: correcting illumina sequencing errors. *Bioinformatics*, 31(17):2885–2887, 2015.
- [264] Heng Li. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics*, 32(14):2103–2110, 2016.
- [265] Heng Li. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18):3094–3100, 2018.
- [266] Hongdong Li. Gtftools: a python package for analyzing various modes of gene models. *bioRxiv*, page 263517, 2018.
- [267] Min Li, Zhongxiang Liao, Yiming He, Jianxin Wang, Junwei Luo, and Yi Pan. Isea: iterative seed-extension algorithm for de novo assembly using paired-end information and insert size distribution. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, 14(4):916–925, 2017.
- [268] Weizhong Li, R Alexander Richter, Yunsup Jung, Qiyun Zhu, and Robert W Li. Web-based bioinformatics workflows for end-to-end rna-seq data computation and analysis in agricultural animal species. *BMC genomics*, 17(1):761, 2016.

-
- [269] Wenzheng Li, Weili Wang, Philip J Uren, Luiz OF Penalva, and Andrew D Smith. Riborex: fast and flexible identification of differential translation from ribo-seq data. *Bioinformatics*, 33(11):1735–1737, 2017.
- [270] Yun-Lung Li, Jui-Cheng Weng, Chiung-Chih Hsiao, Min-Te Chou, Chin-Wen Tseng, and Jui-Hung Hung. Peat: an intelligent and efficient paired-end sequencing adapter trimming algorithm. In *BMC bioinformatics*, volume 16, page S2. BioMed Central, 2015.
- [271] Yang Liao, Gordon K Smyth, and Wei Shi. The r package rsubread is easier, faster, cheaper and better for alignment and quantification of rna sequencing reads. *bioRxiv*, page 377762, 2018.
- [272] Jae Hyun Lim, Soo Youn Lee, and Ju Han Kim. Trapr: R package for statistical analysis and visualization of rna-seq data. *Genomics & informatics*, 15(1):51, 2017.
- [273] Jing-Quan Lim, Chandana Tennakoon, Peiyong Guan, and Wing-Kin Sung. Batalign: an incremental method for accurate alignment of sequencing reads. *Nucleic acids research*, 43(16):e107–e107, 2015.
- [274] Hsin-Nan Lin and Wen-Lian Hsu. Kart: a divide-and-conquer algorithm for ngs read alignment. *Bioinformatics*, 33(15):2281–2287, 2017.
- [275] Yu Lin, Jeffrey Yuan, Mikhail Kolmogorov, Max W Shen, Mark Chaisson, and Pavel A Pevzner. Assembly of long error-prone reads using de bruijn graphs. *Proceedings of the National Academy of Sciences*, 113(52):E8396–E8405, 2016.
- [276] Pierre Lindenbaum, Matilde Karakachoff, and Richard Redon. Jfxngs: A bam/vcf viewer with javascript-based filtering/reformatting functionalities. *BioRxiv*, page 120196, 2017.
- [277] Pierre Lindenbaum and Richard Redon. bioalcaide, samjs and vcfilterjs: object-oriented formatters and filters for bioinformatics files. *Bioinformatics*, 34(7):1224–1225, 2017.
- [278] Bo Liu, Yan Gao, and Yadong Wang. Lamsa: fast split read alignment with long approximate matches. *Bioinformatics*, 33(2):192–201, 2017.
- [279] Bo Liu, Dengfeng Guan, Mingxiang Teng, and Yadong Wang. rhat: fast alignment of noisy long reads with regional hashing. *Bioinformatics*, 32(11):1625–1631, 2015.
- [280] Bo Liu, Hongzhe Guo, Michael Brudno, and Yadong Wang. debga: read alignment with de bruijn graph-based seed and extension. *Bioinformatics*, 32(21):3224–3232, 2016.
- [281] Bo Liu, Dixian Zhu, and Yadong Wang. debwt: parallel construction of burrows–wheeler transform for large collection of genomes with de bruijn-branch encoding. *Bioinformatics*, 32(12):i174–i182, 2016.

- [282] Tzu-Yu Liu and Yun S Song. Prediction of ribosome footprint profile shapes from transcript sequences. *Bioinformatics*, 32(12):i183–i191, 2016.
- [283] Yongchao Liu, Thomas Hankeln, and Bertil Schmidt. Parallel and space-efficient construction of burrows-wheeler transform and suffix array for big genome data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, 13(3):592–598, 2016.
- [284] Yuansheng Liu, Hui Peng, Limsoon Wong, and Jinyan Li. High-speed and high-ratio referential genome compression. *Bioinformatics*, 33(21):3364–3372, 2017.
- [285] Fabricio Loayza-Puch, Koos Rooijers, Jelle Zijlstra, Behzad Mousmbeini, Esther A Zaal, Joachim F Oude Vrielink, Rui Lopes, Alejandro P Ugalde, Celia R Berkers, and Reuven Agami. Tgf β 1-induced leucine limitation uncovered by differential ribosome codon reading. *EMBO reports*, 18(4):549–557, 2017.
- [286] Tobias P Loka, Simon H Tausch, and Bernhard Y Renard. Reliable variant calling during runtime of illumina sequencing. *BioRxiv*, page 387662, 2018.
- [287] Nicholas J Loman, Joshua Quick, and Jared T Simpson. A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nature methods*, 12(8):733, 2015.
- [288] Yang Young Lu, Jinchi Lv, Jed A Fuhrman, and Fengzhu Sun. Towards enhanced and interpretable clustering/classification in integrative genomics. *Nucleic acids research*, 45(20):e169–e169, 2017.
- [289] Aaron TL Lun, Hervé Pagès, and Mike L Smith. beachmat: A bioconductor c++ api for accessing high-throughput biological data from a variety of r matrix types. *PLoS computational biology*, 14(5):e1006135, 2018.
- [290] Yuan Luo, Chengsheng Mao, Yiben Yang, Fei Wang, Faraz S Ahmad, Donna Arnett, Marguerite R Irvin, and Sanjiv J Shah. Integrating hypertension phenotype and genotype with hybrid non-negative matrix factorization. *arXiv preprint arXiv:1805.05008*, 2018.
- [291] Joseph D Lutgring, Wenming Zhu, Tom JB de Man, Johannetsy J Avillan, Karen F Anderson, David R Lonsway, Lori A Rowe, Dhvani Batra, J Kamile Rasheed, and Brandi M Limbago. Phenotypic and genotypic characterization of enterobacteriaceae producing oxacillinase-48–like carbapenemases, united states. *Emerging infectious diseases*, 24(4):700, 2018.
- [292] Piotr Madanecki, Magdalena Bałut, Patrick G Buckley, J Renata Ochocka, Rafał Bartoszewski, David K Crossman, Ludwine M Messiaen, and Arkadiusz Piotrowski. High-throughput tabular data processor–platform independent graphical tool for processing large data sets. *PloS one*, 13(2):e0192858, 2018.

- [293] Mohammed-Amin Madoui, Stefan Engelen, Corinne Cruaud, Caroline Belser, Laurie Bertrand, Adriana Alberti, Arnaud Lemainque, Patrick Wincker, and Jean-Marc Aury. Genome assembly using nanopore-guided long and error-free dna reads. *BMC genomics*, 16(1):327, 2015.
- [294] Niklas Mähler, Jing Wang, Barbara K Terebienieć, Pär K Ingvarsson, Nathaniel R Street, and Torgeir R Hvidsten. Gene co-expression network connectivity is an important determinant of selective constraint. *PLoS genetics*, 13(4):e1006402, 2017.
- [295] Raunaq Malhotra, Manjari Jha, Mary Poss, and Raj Acharya. A random forest classifier for detecting rare variants in ngs data from viral populations. *Computational and structural biotechnology journal*, 15:388–395, 2017.
- [296] Brandon Malone, Ilian Atanassov, Florian Aeschimann, Xiping Li, Helge Großhans, and Christoph Dieterich. Bayesian prediction of rna translation from ribosome profiling. *Nucleic acids research*, 45(6):2960–2972, 2017.
- [297] Greg Malysa, Mikel Hernaez, Idoia Ochoa, Milind Rao, Karthik Ganesan, and Tsachy Weissman. Qvz: lossy compression of quality values. *Bioinformatics*, 31(19):3122–3129, 2015.
- [298] Andrea Manconi, Emanuele Manca, Marco Moscatelli, Matteo Gnocchi, Alessandro Orro, Giuliano Armano, and Luciano Milanesi. G-cnv: a gpu-based tool for preparing data to detect cnvs with read-depth methods. *Frontiers in bioengineering and biotechnology*, 3:28, 2015.
- [299] Serghei Mangul, Sarah Van Driesche, Lana S Martin, Kelsey C Martin, and Eleazar Eskin. Umi-reducer: Collapsing duplicate sequencing reads via unique molecular identifiers. *bioRxiv*, page 103267, 2017.
- [300] Brian K Mannakee, Uthra Balaji, Agnieszka K Witkiewicz, Ryan N Gutenkunst, and Erik S Knudsen. Sensitive and specific post-call filtering of genetic variants in xenograft and primary tumors. *Bioinformatics*, 34(10):1713–1718, 2018.
- [301] Tamer A Mansour, Erica Y Scott, Carrie J Finno, Rebecca R Bellone, Michael J Mienaltowski, M Cecilia Penedo, Pablo J Ross, Stephanie J Valberg, James D Murray, and C Titus Brown. Tissue resolved, gene structure refined equine transcriptome. *BMC genomics*, 18(1):103, 2017.
- [302] Daniel Mapleson, Gonzalo Garcia Accinelli, George Kettleborough, Jonathan Wright, and Bernardo J Clavijo. Kat: a k-mer analysis toolkit to quality control ngs datasets and genome assemblies. *Bioinformatics*, 33(4):574–576, 2016.
- [303] Daniel Mapleson, Luca Venturini, Gemy Kaithakottil, and David Swarbreck. Efficient and accurate detection of splice junctions from rna-seq with portcullis. *Giga-Science*, 7(12):giy131, 2018.
- [304] Guillaume Marçais, James A Yorke, and Aleksey Zimin. Quorum: an error corrector for illumina reads. *PLoS One*, 10(6):e0130821, 2015.

-
- [305] Camille Marchet, Lolita Lecompte, Corinne Da Silva, Corinne Cruaud, Jean-Marc Aury, Jacques Nicolas, and Pierre Peterlongo. De novo clustering of long reads by gene from transcriptomics data. *Nucleic acids research*, 47(1):e2–e2, 2018.
- [306] Camille Marchet, Pierre Morisse, Lolita Lecompte, Antoine Limasset, Arnaud Lefebvre, Thierry Lecroq, and Pierre Peterlongo. Elector: Evaluator for long reads correction methods. *BioRxiv*, page 512889, 2019.
- [307] Santiago Marco-Sola and Paolo Ribeca. Efficient alignment of illumina-like high-throughput sequencing reads with the genomic multi-tool (gem) mapper. *Current protocols in bioinformatics*, 50(1):11–13, 2015.
- [308] Eric Marinier, Daniel G Brown, and Brendan J McConkey. Pollux: platform independent error correction of single and mixed genomes. *BMC bioinformatics*, 16(1):10, 2015.
- [309] Joao C Marques and Michael B Orger. Clusterdv, a simple density-based clustering method that is robust, general and automatic. *bioRxiv*, page 224840, 2018.
- [310] Ruth C Martin, Kelly Vining, and James E Dombrowski. Genome-wide (chip-seq) identification of target genes regulated by bdbzip10 during paraquat-induced oxidative stress. *BMC plant biology*, 18(1):58, 2018.
- [311] Xavier Martinez, Marta Pozuelo, Victoria Pascal, David Campos, Ivo Gut, Marta Gut, Fernando Azpiroz, Francisco Guarner, and Chaysavanh Manichanh. Meta-trans: an open-source pipeline for metatranscriptomics. *Scientific reports*, 6:26447, 2016.
- [312] Andre P Masella, Christopher M Lalansingh, Pragash Sivasundaram, Michael Fraser, Robert G Bristow, and Paul C Boutros. Bamql: a query language for extracting reads from bam files. *BMC bioinformatics*, 17(1):305, 2016.
- [313] Luís MO Matos, António JR Neves, Diogo Pratas, and Armando J Pinho. Mafco: A compression tool for maf files. *PloS one*, 10(3):e0116082, 2015.
- [314] Nicholas J Matzke and Randall B Irmis. Including autapomorphies is important for paleontological tip-dating with clocklike data, but not with non-clock data. *PeerJ*, 6:e4553, 2018.
- [315] Ali May, Sanne Abeln, Mark J Buijs, Jaap Heringa, Wim Crielaard, and Bernd W Brandt. Ngs-eval: Ngs error analysis and novel sequence variant detection tool. *Nucleic acids research*, 43(W1):W301–W305, 2015.
- [316] Anand Mayakonda, De-Chen Lin, Yassen Assenov, Christoph Plass, and H Phillip Koeffler. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome research*, 28(11):1747–1756, 2018.
- [317] Bradon R McDonald, Tania Contente-Cuomo, Stephen-John Sammut, Ahuva Odenheimer-Bergman, Brenda Ernst, Nieves Perdigones, Suet-Feung Chin, Maria

- Farooq, Patricia A Cronin, Karen S Anderson, et al. Detection of residual disease after neoadjuvant therapy in breast cancer using personalized circulating tumor dna analysis. *BioRxiv*, page 425470, 2018.
- [318] Aaron McKenna, Matthew Hanna, Eric Banks, Andrey Sivachenko, Kristian Cibulskis, Andrew Kernytzsky, Kiran Garimella, David Altshuler, Stacey Gabriel, Mark Daly, et al. The genome analysis toolkit: a mapreduce framework for analyzing next-generation dna sequencing data. *Genome research*, 20(9):1297–1303, 2010.
- [319] Tobias Meißner, Kathleen M Fisch, Louis Gioia, and Andrew I Su. Oncorep: an n-of-1 reporting tool to support genome-guided treatment for breast cancer patients using rna-sequencing. *BMC medical genomics*, 8(1):24, 2015.
- [320] Arthur TO Melo and Iago Hale. Expanded functionality, increased accuracy, and enhanced speed in the de novo genotyping-by-sequencing pipeline gbs-snp-crop. *Bioinformatics*, 2018.
- [321] Páll Melsted, Vasilis Ntranos, and Lior Pachter. The barcode, umi, set format and bustools. *bioRxiv*, page 472571, 2018.
- [322] Marco-Antonio Mendoza-Parra, Vincent Saravaki, Pierre-Etienne Cholley, Matthias Blum, Benjamin Billoré, and Hinrich Gronemeyer. Antibody performance in chip-sequencing assays: From quality scores of public data sets to quantitative certification. *F1000Research*, 5, 2016.
- [323] Audrey M Michel, James PA Mullan, Vimalkumar Velayudhan, Patrick BF O’Connor, Claire A Donohue, and Pavel V Baranov. Ribogalaxy: a browser based platform for the alignment, analysis and visualization of ribosome profiling data. *RNA biology*, 13(3):316–319, 2016.
- [324] Giles Miclotte, Mahdi Heydari, Piet Demeester, Stephane Rombauts, Yves Van de Peer, Pieter Audenaert, and Jan Fostier. Jabba: hybrid error correction for long sequencing reads. *Algorithms for Molecular Biology*, 11(1):10, 2016.
- [325] Ilia Minkin, Son Pham, and Paul Medvedev. Twopaco: An efficient algorithm to build the compacted de bruijn graph from many complete genomes. *Bioinformatics*, 33(24):4024–4032, 2016.
- [326] Laurent Modolo and Emmanuelle Lerat. Urqt: an efficient software for the unsupervised quality trimming of ngs data. *BMC bioinformatics*, 16(1):137, 2015.
- [327] Hamid Mohamadi, Justin Chu, Benjamin P Vandervalk, and Inanc Birol. nthash: recursive nucleotide hashing. *Bioinformatics*, 32(22):3492–3494, 2016.
- [328] Angel Mojarro, Julie Hachey, Gary Ruvkun, Maria T Zuber, and Christopher E Carr. Carrierseq: a sequence analysis workflow for low-input nanopore sequencing. *BMC bioinformatics*, 19(1):108, 2018.

- [329] Paula Moolhuijzen, Pao Theen See, James K Hane, Gongjun Shi, Zhaohui Liu, Richard P Oliver, and Caroline S Moffat. Comparative genomics of the wheat fungal pathogen *pyrenophora tritici-repentis* reveals chromosomal variations and genome plasticity. *BMC genomics*, 19(1):279, 2018.
- [330] Yoshiro Morimoto, Mihoko Shimada-Sugimoto, Takeshi Otowa, Shintaro Yoshida, Akira Kinoshita, Hiroyuki Mishima, Naohiro Yamaguchi, Takatoshi Mori, Akira Imamura, Hiroki Ozawa, et al. Whole-exome sequencing and gene-based rare variant association tests suggest that *pla2g4e* might be a risk gene for panic disorder. *Translational psychiatry*, 8(1):41, 2018.
- [331] Pierre Morisse, Thierry Lecroq, and Arnaud Lefebvre. Hybrid correction of highly noisy long reads using a variable-order de bruijn graph. *Bioinformatics*, 34(24):4213–4222, 2018.
- [332] Nobuhiko Muramoto, Arisa Oda, Hidenori Tanaka, Takahiro Nakamura, Kazuto Kugou, Kazuki Suda, Aki Kobayashi, Shiori Yoneda, Akinori Ikeuchi, Hiroki Sugimoto, et al. Phenotypic diversification by enhanced genome restructuring after induction of multiple dna double-strand breaks. *Nature communications*, 9(1):1995, 2018.
- [333] Ruth J Muschel, Hong Bing Zhang, George Iliakis, and W Gillies McKenna. Cyclin b expression in hela cells during the g2 block induced by ionizing radiation. *Cancer Research*, 51(19):5113–5117, 1991.
- [334] Harun Mustafa, Andre Kahles, Mikhail Karasikov, and Gunnar Raetsch. Metannot: A succinct data structure for compression of colors in dynamic de bruijn graphs. *bioRxiv*, page 236711, 2017.
- [335] Damoon Nashta-ali, Ali Aliyari, Ahmad Ahmadian Moghadam, Mohammad Amin Edrisi, Seyed Abolfazl Motahari, and Babak Hossein Khalaj. Meta-aligner: long-read alignment based on genome statistics. *BMC bioinformatics*, 18(1):126, 2017.
- [336] José Fernández Navarro, Joel Sjöstrand, Fredrik Salmén, Joakim Lundeberg, and Patrik L Ståhl. St pipeline: an automated pipeline for spatial mapping of unique transcripts. *Bioinformatics*, 33(16):2591–2593, 2017.
- [337] Elvis Ndah, Veronique Jonckheere, Adam Giess, Eivind Valen, Gerben Menschaert, and Petra Van Damme. Reparation: ribosome profiling assisted (re-) annotation of bacterial genomes. *Nucleic acids research*, 45(20):e168–e168, 2017.
- [338] Aaron M Newman, Alexander F Lovejoy, Daniel M Klass, David M Kurtz, Jacob J Chabon, Florian Scherer, Henning Stehr, Chih Long Liu, Scott V Bratman, Carmen Say, et al. Integrated digital error suppression for improved detection of circulating tumor dna. *Nature biotechnology*, 34(5):547, 2016.
- [339] Son Hoang Nguyen, Tania PS Duarte, Lachlan JM Coin, and Minh Duc Cao. Real-time demultiplexing nanopore barcoded sequencing data with npbarcode. *Bioinformatics*, 33(24):3988–3990, 2017.

-
- [340] Marius Nicolae, Sudipta Pathak, and Sanguthevar Rajasekaran. Lfqc: a lossless compression algorithm for fastq files. *Bioinformatics*, 31(20):3276–3281, 2015.
- [341] Daniil Nikitin, Dmitry Penzar, Andrew Garazha, Maxim Sorokin, Victor Tkachev, Nicolas Borisov, Alexander Poltorak, Vladimir Prassolov, and Anton A Buzdin. Profiling of human molecular pathways affected by retrotransposons at the level of regulation by transcription factor proteins. *Frontiers in immunology*, 9:30, 2018.
- [342] Martin Norling, Niclas Jareborg, and Jacques Dainat. Emblmygff3: a converter facilitating genome annotation submission to european nucleotide archive. *BMC research notes*, 11(1):584, 2018.
- [343] Ibrahim Numanagić, James K Bonfield, Faraz Hach, Jan Voges, Jörn Ostermann, Claudio Alberti, Marco Mattavelli, and S Cenk Sahinalp. Comparison of high-throughput sequencing data compression tools. *Nature methods*, 13(12):1005, 2016.
- [344] Idoia Ochoa, Mikel Hernaez, and Tsachy Weissman. idocomp: a compression scheme for assembled genomes. *Bioinformatics*, 31(5):626–633, 2014.
- [345] Jared O’Connell, Ole Schulz-Trieglaff, Emma Carlson, Matthew M Hims, Niall A Gormley, and Anthony J Cox. Nxtrim: optimized trimming of illumina mate pair reads. *Bioinformatics*, 31(12):2035–2037, 2015.
- [346] Patrick BF O’Connor, Dmitry E Andreev, and Pavel V Baranov. Comparative survey of the relative impact of mrna features on local ribosome profiling read density. *Nature communications*, 7:12915, 2016.
- [347] Christian Oertlin, Julie Lorent, Valentina Gandin, Carl Murie, Laia Masvidal, Marie Cargnello, Luc Furic, Ivan Topisirovic, and Ola Larsson. Genome-wide analysis of differential translation and differential translational buffering using anota2seq. *bioRxiv*, page 106922, 2017.
- [348] Takeshi Ogasawara, Yinhe Cheng, and Tzy-Hwa Kathy Tzeng. Sam2bam: High-performance framework for ngs data preprocessing tools. *PloS one*, 11(11):e0167100, 2016.
- [349] Julia Oh, Allyson L Byrd, Clay Deming, Sean Conlan, Betty Barnabas, Robert Blakesley, Gerry Bouffard, Shelise Brooks, Holly Coleman, Mila Dekhtyar, et al. Biogeography and individuality shape function in the human skin metagenome. *Nature*, 514(7520):59, 2014.
- [350] Damien M O’Halloran. fastq_brew: module for analysis, preprocessing, and reformatting of fastq sequence data. *BMC research notes*, 10(1):275, 2017.
- [351] Konstantin Okonechnikov, Ana Conesa, and Fernando García-Alcalde. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*, 32(2):292–294, 2015.

-
- [352] Raul Ossio, Diego Said Anaya-Mancilla, O Isaac Garcia-Salinas, Jair S Garcia-Sotelo, Luis A Aguilar, David J Adams, and Carla Robles-Espinoza. Vcf/plotein: A web application to facilitate the clinical interpretation of genetic and genomic variants from exome sequencing projects. *bioRxiv*, page 466490, 2018.
- [353] Andrew J Page, Sascha Steinbiss, Ben Taylor, Torsten Seemann, and Jacqueline A Keane. {GFF} 3toembl: Preparing annotated assemblies for submission to {EMBL}. *The Journal of Open Source Software*, 1(6), 2016.
- [354] Jonathan M Palmer, Michelle A Jusino, Mark T Banik, and Daniel L Lindner. Non-biological synthetic spike-in controls and the amptk software pipeline improve mycobionome data. *PeerJ*, 6:e4925, 2018.
- [355] Ram Vinay Pandey, Stephan Pabinger, Albert Kriegner, and Andreas Weinhäusel. Clinqc: a tool for quality control and cleaning of sanger and ngs data in clinical research. *BMC bioinformatics*, 17(1):56, 2016.
- [356] Ram Vinay Pandey, Stephan Pabinger, Albert Kriegner, and Andreas Weinhäusel. Mutaaid: Sanger and ngs based integrated pipeline for mutation identification, validation and annotation in human molecular genetics. *PloS one*, 11(2):e0147697, 2016.
- [357] Swati Parekh, Christoph Ziegenhain, Beate Vieth, Wolfgang Enard, and Ines Hellmann. zumis-a fast and flexible pipeline to process rna sequencing data with umis. *Gigascience*, 7(6):giy059, 2018.
- [358] Tom Paridaens, Glenn Van Wallendael, Wesley De Neve, and Peter Lambert. Afresh: an adaptive framework for compression of reads and assembled sequences with random access functionality. *Bioinformatics*, 33(10):1464–1472, 2017.
- [359] Rob Patro and Carl Kingsford. Data-dependent bucketing improves reference-free compression of sequencing reads. *Bioinformatics*, 31(17):2770–2777, 2015.
- [360] Damien Paulet, Alexandre David, and Eric Rivals. Ribo-seq enlightens codon usage bias. *DNA Research*, 24(3):303–210, 2017.
- [361] Brent S Pedersen, Ryan L Collins, Michael E Talkowski, and Aaron R Quinlan. Indexcov: fast coverage quality control for whole-genome sequencing. *GigaScience*, 6(11):gix090, 2017.
- [362] Brent S Pedersen and Aaron R Quinlan. cyvcf2: fast, flexible variant analysis with python. *Bioinformatics*, 2017.
- [363] Brent S Pedersen and Aaron R Quinlan. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics*, 34(5):867–868, 2017.
- [364] Brent S Pedersen and Aaron R Quinlan. hts-nim: scripting high-performance genomic analyses. *bioRxiv*, page 261735, 2018.

-
- [365] Alexander Peltzer, Günter Jäger, Alexander Herbig, Alexander Seitz, Christian Kniep, Johannes Krause, and Kay Nieselt. Eager: efficient ancient genome reconstruction. *Genome biology*, 17(1):60, 2016.
- [366] Emanuel Diego S Penha, Egiebade Iriabho, Alex Dussaq, Diana Magalhães de Oliveira, and Jonas S Almeida. Isomorphic semantic mapping of variant call format (vcf2rdf). *Bioinformatics*, 33(4):547–548, 2016.
- [367] Paula Perez-Rubio, Claudio Lottaz, and Julia Catherine Engelmann. Fastqpur: high-performance preprocessing of rna-seq data. *bioRxiv*, page 480707, 2018.
- [368] Sari Peura, Lucas Sinclair, Stefan Bertilsson, and Alexander Eiler. Metagenomic insights into strategies of aerobic and anaerobic carbon and nitrogen transformation in boreal lakes. *Scientific reports*, 5:12102, 2015.
- [369] Marguerite Picard-Bennoun. Genetic evidence for ribosomal antisuppressors in *inopospora anserina*. *Molecular and General Genetics MGG*, 147(3):299–306, 1976.
- [370] Alexandra Popa, Kevin Lebrigand, Agnes Paquet, Nicolas Nottet, Karine Robbe-Sermesant, Rainer Waldmann, and Pascal Barbry. Riboprofiling: a bioconductor package for standard ribo-seq pipeline processing. *F1000Research*, 5, 2016.
- [371] Victoria Popic and Serafim Batzoglou. A hybrid cloud read aligner based on min-hash and kmer voting that preserves privacy. *Nature communications*, 8:15311, 2017.
- [372] Jacob Porter and Liqing Zhang. Infotrim: A dna read quality trimmer using entropy. In *2017 IEEE 7th International Conference on Computational Advances in Bio and Medical Sciences (ICCABS)*, pages 1–2. IEEE, 2017.
- [373] Jacob Porter and Liqing Zhang. Bispin and bfast-gap: Mapping bisulfite-treated reads. *BioRxiv*, page 284596, 2018.
- [374] Thomas M Poulsen and Martin Frith. Variable-order sequence modeling improves bacterial strain discrimination for ion torrent dna reads. *BMC bioinformatics*, 18(1):299, 2017.
- [375] Jacob Pritt and Ben Langmead. Boiler: lossy compression of rna-seq alignments using coverage vectors. *Nucleic acids research*, 44(16):e133–e133, 2016.
- [376] Sebastian Proost, Agnieszka Krawczyk, and Marek Mutwil. Lstrap: efficiently combining rna sequencing data into co-expression networks. *BMC bioinformatics*, 18(1):444, 2017.
- [377] Pau Puigdevall and Robert Castelo. Genomicscores: seamless access to genomewide position-specific scores from r and bioconductor. *Bioinformatics*, 34(18):3208–3210, 2018.

- [378] Emma Puighermanal, Anne Biever, Vincent Pascoli, Su Melser, Marine Pratlong, Laura Cutando, Stephanie Rialle, Dany Severac, Jihane Boubaker-Vitre, Oded Meyuhas, et al. Ribosomal protein s6 phosphorylation is involved in novelty-induced locomotion, synaptic plasticity and mrna translation. *Frontiers in molecular neuroscience*, 10:419, 2017.
- [379] Qian Qin, Shenglin Mei, Qiu Wu, Hanfei Sun, Lewyn Li, Len Taing, Sujun Chen, Fugen Li, Tao Liu, Chongzhi Zang, et al. Chilin: a comprehensive chip-seq and dnase-seq quality control and analysis pipeline. *BMC bioinformatics*, 17(1):404, 2016.
- [380] Anil Raj, Sidney H Wang, Heejung Shim, Arbel Harpak, Yang I Li, Brett Engelmann, Matthew Stephens, Yoav Gilad, and Jonathan K Pritchard. Thousands of novel translated open reading frames in humans inferred by ribosome footprint profiling. *Elife*, 5:e13328, 2016.
- [381] Sarah M Ramirez-Busby, Afif Elghraoui, Yeon Bin Kim, Kellie Kim, and Faramarz Valafar. Pbhoover and cigarroller: a method for confident haploid variant calling on pacific biosciences data and its application to heterogeneous population analysis. *bioRxiv*, page 360370, 2018.
- [382] Vijaya Raghavan Rangamaran, Bharathram Uppili, Dharani Gopal, and Kirubakaran Ramalingam. Easyqc: Tool with interactive user interface for efficient next-generation sequencing data quality control. *Journal of Computational Biology*, 25(12):1301–1311, 2018.
- [383] Thomas S Rask, Bent Petersen, Donald S Chen, Karen P Day, and Anders Gorm Pedersen. Using expected sequence features to improve basecalling accuracy of amplicon pyrosequencing data. *BMC bioinformatics*, 17(1):176, 2016.
- [384] Vida Ravanmehr, Minji Kim, Zhiying Wang, and Olgica Milenković. Chipwig: a random access-enabling lossless and lossy compression method for chip-seq data. *Bioinformatics*, 34(6):911–919, 2017.
- [385] David Redin, Tobias Frick, Hooman Aghelpasand, Jennifer Theland, Max Kaller, Erik Borgstrom, Remi-Andre Olsen, and Afshin Ahmadian. Efficient whole genome haplotyping and high-throughput single molecule phasing with barcode-linked reads. *bioRxiv*, page 356121, 2018.
- [386] Allison A Regier, Yossi Farjoun, David E Larson, Olga Krasheninina, Hyun Min Kang, Daniel P Howrigan, Bo-Juen Chen, Manisha Kher, Eric Banks, Darren C Ames, et al. Functional equivalence of genome sequencing analysis pipelines enables harmonized variant calling across human genetics projects. *Nature communications*, 9(1):4038, 2018.
- [387] Michael Reich, Thorin Tabor, Ted Liefeld, Helga Thorvaldsdóttir, Barbara Hill, Pablo Tamayo, and Jill P Mesirov. The genepattern notebook environment. *Cell systems*, 5(2):149–151, 2017.

- [388] Gabriel Renaud. glactools: a command-line toolset for the management of genotype likelihoods and allele counts. *Bioinformatics*, 34(8):1398–1400, 2017.
- [389] Martial Rey, Menglin Yang, Linda Lee, Ye Zhang, Joey G Sheff, Christoph W Sensen, Hynek Mrazek, Petr Halada, Petr Man, Justin L McCarville, et al. Addressing proteolytic efficiency in enzymatic degradation therapy for celiac disease. *Scientific reports*, 6:30980, 2016.
- [390] Zachary B Rodriguez, Susan L Perkins, and Christopher C Austin. Multiple origins of green blood in new guinea lizards. *Science advances*, 4(5):eaao5017, 2018.
- [391] Irene Rodriguez-Lujan, Jeff Hasty, and Ramón Huerta. Fbb: a fast bayesian-bound tool to calibrate rna-seq aligners. *Bioinformatics*, 33(2):210–218, 2017.
- [392] Johannes T Roehr, Christoph Dieterich, and Knut Reinert. Flexbar 3.0—simd and multicore parallelization. *Bioinformatics*, 33(18):2941–2942, 2017.
- [393] Łukasz Roguski, Idoia Ochoa, Mikel Hernaez, and Sebastian Deorowicz. Fastore: a space-saving solution for raw sequencing data. *Bioinformatics*, 34(16):2748–2756, 2018.
- [394] Cynara CT Romero, Jasper P Vermeulen, Anton Vels, Axel Himmelbach, Martin Mascher, and Rients E Niks. Mapping resistance to powdery mildew in barley reveals a large-effect nonhost resistance qtl. *Theoretical and applied genetics*, pages 1–15, 2018.
- [395] Simon P Sadedin and Alicia Oshlack. Bazam: A rapid method for read extraction and realignment of high throughput sequencing data. *bioRxiv*, page 433003, 2018.
- [396] Jessica E Sagers, Adam S Brown, Sasa Vasilijic, Rebecca Lewis, Mehmet I Sahin, Lukas D Landegger, Roy H Perlis, Isaac S Kohane, D Bradley Welling, Chirag J Patel, et al. Computational repositioning and preclinical validation of mifepristone for human vestibular schwannoma. *Scientific reports*, 8(1):5437, 2018.
- [397] Subrata Saha and Sanguthevar Rajasekaran. Ec: an efficient error correction algorithm for short reads. *BMC bioinformatics*, 16(17):S2, 2015.
- [398] Subrata Saha and Sanguthevar Rajasekaran. Ergc: an efficient referential genome compression algorithm. *Bioinformatics*, 31(21):3468–3475, 2015.
- [399] Subrata Saha and Sanguthevar Rajasekaran. Nrgc: a novel referential genome compression algorithm. *Bioinformatics*, 32(22):3405–3412, 2016.
- [400] Kristoffer Sahlin and Paul Medvedev. De novo clustering of long-read transcriptome data using a greedy, quality-value based algorithm. *BioRxiv*, page 463463, 2018.
- [401] Farzaneh Salari, Fatemeh Zare-Mirakabad, Mehdi Sadeghi, and Hassan Rokni-Zadeh. Assessing the impact of exact reads on reducing the error rate of read mapping. *BMC bioinformatics*, 19(1):406, 2018.

-
- [402] Jose Salavert, Andres Tomas, Joaquin Tarraga, Ignacio Medina, Joaquin Dopazo, and Ignacio Blanquer. Fast inexact mapping using advanced tree exploration on backward search methods. *BMC bioinformatics*, 16(1):18, 2015.
- [403] Mohamed-Ashick M Saleem, Marco-Antonio Mendoza-Parra, Pierre-Etienne Cholley, Matthias Blum, and Hinrich Gronemeyer. Epimetheus-a multi-profile normalizer for epigenomic sequencing data. *BMC bioinformatics*, 18(1):259, 2017.
- [404] Leena Salmela, Riku Walve, Eric Rivals, and Esko Ukkonen. Accurate self-correction of errors in long reads using de bruijn graphs. *Bioinformatics*, 33(6):799–806, 2016.
- [405] Katrin Sameith, Juliana G Roscito, and Michael Hiller. Iterative error correction of long sequencing reads maximizes accuracy and improves contig assembly. *Briefings in bioinformatics*, 18(1):1–8, 2016.
- [406] Sarah Sandmann. Badregionfinder—an r/bioconductor package for identifying regions with bad coverage. 2017.
- [407] Camilla A Santos, Sónia Andrade, Ana K Teixeira, Flávio Farias, Karin Kurkjian, Ana C Guerrelhas, João L Rocha, Pedro M Galetti Jr, and Patrícia D Freitas. *Litopenaeus vannamei* transcriptome profile of populations evaluated for growth performance and exposed to white spot syndrome virus (wssv). *Frontiers in genetics*, 9:120, 2018.
- [408] Masayuki Sato, Hirotaka Suetake, and Masaaki Kotera. Kcf-convoy: efficient python package to convert kegg chemical function and substructure fingerprints. *BioRxiv*, page 452383, 2018.
- [409] Sergi Sayols, Denise Scherzinger, and Holger Klein. dupradar: a bioconductor package for the assessment of pcr artifacts in rna-seq data. *BMC bioinformatics*, 17(1):428, 2016.
- [410] Ute Scheller, Kathrin Pfisterer, Steffen Uebe, Arif B Ekici, André Reis, Rami Jamra, and Fulvia Ferrazzi. Integrative bioinformatics analysis characterizing the role of edc3 in mrna decay and its association to intellectual disability. *BMC medical genomics*, 11(1):41, 2018.
- [411] Melanie Schirmer, Rosalinda D’Amore, Umer Z Ijaz, Neil Hall, and Christopher Quince. Illumina error profiles: resolving fine-scale variation in metagenomic sequencing data. *BMC bioinformatics*, 17(1):125, 2016.
- [412] Sarah E Schmedes, Jonathan L King, and Bruce Budowle. Correcting inconsistencies and errors in bacterial genome metadata using an automated curation tool in excel (autocure). *Frontiers in bioengineering and biotechnology*, 3:138, 2015.
- [413] Emanuel Schmid-Siebert, Sophie Richard, Amanda Luraschi, Konrad Mühlethaler, Marco Pagni, and Philippe M Hauser. Mechanisms of surface antigenic variation

- in the human pathogenic fungus *pneumocystis jirovecii*. *MBio*, 8(6):e01470–17, 2017.
- [414] Anica Scholz, Florian Eggenhofer, Rick Gelhausen, Björn Grüning, Kathi Zarnack, Bernhard Brüne, Rolf Backofen, and Tobias Schmid. uorf-tools-workflow for the determination of translation-regulatory upstream open reading frames. *bioRxiv*, page 415018, 2018.
- [415] Patrick Schorderet. Neat: a framework for building fully automated ngs pipelines and analyses. *BMC bioinformatics*, 17(1):53, 2016.
- [416] Jacob Schreiber and Kevin Karplus. Analysis of nanopore data using hidden markov models. *Bioinformatics*, 31(12):1897–1903, 2015.
- [417] Christopher M Schroeder, Franz J Hilke, Markus W Löffler, Michael Bitzer, Florian Lenz, and Marc Sturm. A comprehensive quality control workflow for paired tumor-normal ngs experiments. *Bioinformatics*, 33(11):1721–1722, 2017.
- [418] Sven Schuierer and Guglielmo Roma. The exon quantification pipeline (eqp): a comprehensive approach to the quantification of gene, exon and junction expression from rna-seq data. *Nucleic acids research*, 44(16):e132–e132, 2016.
- [419] Fritz J Sedlazeck, Philipp Rescheneder, Moritz Smolka, Han Fang, Maria Nattestad, Arndt von Haeseler, and Michael C Schatz. Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods*, 15(6):461–468, 2018.
- [420] Shobana Sekar, Lori Cuyugan, Jonathan Adkins, Philipp Geiger, and Winnie S Liang. Circular rna expression and regulatory network prediction in posterior cingulate astrocytes in elderly subjects. *BMC genomics*, 19(1):340, 2018.
- [421] Guilhem Sempéré, Florian Philippe, Alexis Dereeper, Manuel Ruiz, Gautier Sarah, and Pierre Larmande. Gigwa—genotype investigator for genome-wide analyses. *GigaScience*, 5(1):25, 2016.
- [422] Ali Sharifi-Zarchi, Daniela Gerovska, Kenjiro Adachi, Mehdi Totonchi, Hamid Pezeshk, Ryan J Taft, Hans R Schöler, Hamidreza Chitsaz, Mehdi Sadeghi, Hossein Baharvand, et al. Dna methylation regulates discrimination of enhancers from promoters through a h3k4me1-h3k4me3 seesaw mechanism. *BMC genomics*, 18(1):964, 2017.
- [423] Alok Sharma, Yosvany López, and Tatsuhiko Tsunoda. Divisive hierarchical maximum likelihood clustering. *BMC bioinformatics*, 18(16):546, 2017.
- [424] Siavash Sheikhezadeh and Dick de Ridder. Ace: accurate correction of errors using k-mer tries. *Bioinformatics*, 31(19):3216–3218, 2015.
- [425] Jennifer Marie Shelton and Susan J Brown. Fasta-o-matic: a tool to sanity check and if needed reformat fasta files. *BioRxiv*, page 024448, 2015.

- [426] Li Shen, Ningyi Shao, Xiaochuan Liu, and Eric Nestler. ngs. plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC genomics*, 15(1):284, 2014.
- [427] Anish MS Shrestha, Naruki Yoshikawa, and Kiyoshi Asai. Combining probabilistic alignments with read pair information improves accuracy of split-alignments. *Bioinformatics*, 34(21):3631–3637, 2018.
- [428] Mikhail Shugay, Andrew R Zaretsky, Dmitriy A Shagin, Irina A Shagina, Ivan A Volchenkov, Andrew A Shelenkov, Mikhail Y Lebedin, Dmitriy V Bagaev, Sergey Lukyanov, and Dmitriy M Chudakov. Mageri: Computational pipeline for molecular-barcoded targeted resequencing. *PLoS computational biology*, 13(5):e1005480, 2017.
- [429] Jochen Singer, Hans-Joachim Ruscheweyh, Ariane L Hofmann, Thomas Thurnherr, Franziska Singer, Nora C Toussaint, Charlotte KY Ng, Salvatore Pisuoglio, Christian Beisel, Gerhard Christofori, et al. Ngs-pipe: a flexible, easily extendable and highly configurable framework for ngs analysis. *Bioinformatics*, 34(1):107–108, 2017.
- [430] Swati Singh, Garima Khare, Ritika Kar Bahal, Prahlad C Ghosh, and Anil K Tyagi. Identification of mycobacterium tuberculosis bioa inhibitors by using structure-based virtual screening. *Drug design, development and therapy*, 12:1065, 2018.
- [431] Saloni Sinha, Venkata Anudeep Bheemsetty, and Maneesha S Inamdar. A double helical motif in ociad2 is essential for its localization, interactions and stat3 activation. *Scientific reports*, 8(1):7362, 2018.
- [432] Tom Smith, Andreas Heger, and Ian Sudbery. Umi-tools: modeling sequencing errors in unique molecular identifiers to improve quantification accuracy. *Genome research*, 27(3):491–499, 2017.
- [433] Anastasiya Vladimirovna Snezhkina, George Sergeevich Krasnov, Andrew Rostislavovich Zaretsky, Alex Zhavoronkov, Kirill Mikhailovich Nyushko, Alexey Alexandrovich Moskaev, Irina Yurievna Karpova, Anastasiya Isaevna Afremova, Anastasiya Valerievna Lipatova, Dmitriy Vladimirovich Kochetkov, et al. Differential expression of alternatively spliced transcripts related to energy metabolism in colorectal cancer. *BMC genomics*, 17(14):1011, 2016.
- [434] Brad Solomon and Carl Kingsford. Improved search of large transcriptomic sequencing databases using split sequence bloom trees. In *International Conference on Research in Computational Molecular Biology*, pages 257–271. Springer, 2017.
- [435] Li Song and Liliana Florea. Rcorrector: efficient and accurate error correction for illumina rna-seq reads. *GigaScience*, 4(1):48, 2015.
- [436] Ivan Sović, Mile Šikić, Andreas Wilm, Shannon Nicole Fenlon, Swaine Chen, and Niranjan Nagarajan. Fast and sensitive mapping of nanopore sequencing reads with graphmap. *Nature communications*, 7:11307, 2016.

- [437] Adriana C Sperlea and Jason Ernst. Systematic discovery of conservation states for single-nucleotide annotation of the human genome. *bioRxiv*, page 262097, 2018.
- [438] Andrew Spriggs, Steven T Henderson, Melanie L Hand, Susan D Johnson, Jennifer M Taylor, and Anna Koltunow. Assembled genomic and tissue-specific transcriptomic data resources for two genetically distinct lines of cowpea (*vigna unguiculata* (l.) walp). *Gates open research*, 2, 2018.
- [439] Anders Ståhlberg, Paul M Krzyzanowski, Matthew Egyud, Stefan Filges, Lincoln Stein, and Tony E Godfrey. Simple multiplexed pcr-based barcoding of dna for ultrasensitive mutation detection by next-generation sequencing. *nature protocols*, 12(4):664, 2017.
- [440] Ekaterina Starostina, Gaik Tamazian, Pavel Dobrynin, Stephen O’Brien, and Aleksey Komissarov. Cookiecutter: a tool for kmer-based read filtering and extraction. *BioRxiv*, page 024679, 2015.
- [441] Eike J Steinig, Sebastian Duchene, D Ashley Robinson, Stefan Monecke, Maho Yokoyama, Maisem Laabei, Peter Slickers, Patiyan Andersson, Deborah Williamson, Angela Kearns, et al. Global phylogenomics of multidrug-resistant staphylococcus aureus sequence type 772: the bengal bay clone. *bioRxiv*, 2017.
- [442] Robert D Stewart and Mick Watson. pore guis for parallel and real-time processing of minion sequence data. *Bioinformatics*, 33(14):2207–2208, 2017.
- [443] Matthew B Stocks, Irina Mohorianu, Matthew Beckers, Claudia Paicu, Simon Moxon, Joshua Thody, Tamas Dalmay, and Vincent Moulton. The uea srna workbench (version 4.4): a comprehensive suite of tools for analyzing mirnas and srnas. *Bioinformatics*, 34(19):3382–3384, 2018.
- [444] Marcus Stoiber and James Brown. Basecrawler: streaming nanopore basecalling directly from raw signal. *bioRxiv*, page 133058, 2017.
- [445] Jakob Stokholm, Martin J Blaser, Jonathan Thorsen, Morten A Rasmussen, Johannes Waage, Rebecca K Vinding, Ann-Marie M Schoos, Asja Kunøe, Nadia R Fink, Bo L Chawes, et al. Maturation of the gut microbiome and risk of asthma in childhood. *Nature communications*, 9(1):141, 2018.
- [446] David A Streett, Kristen R Petersen, Alida T Gerritsen, Samuel S Hunter, and Matthew L Settles. exphts: analysis of high throughput sequence data in an experimental framework. In *Proceedings of the 6th ACM Conference on Bioinformatics, Computational Biology and Health Informatics*, pages 523–524. ACM, 2015.
- [447] Michal Strejcek, Qiong Wang, Jakub Ridl, and Ondrej Uhlik. Hunting down frame shifts: Ecological analysis of diverse functional gene sequences. *Frontiers in microbiology*, 6:1267, 2015.
- [448] Marc Sturm, Christopher Schroeder, and Peter Bauer. Seqpurge: highly-sensitive adapter trimming for paired-end ngs data. *BMC bioinformatics*, 17(1):208, 2016.

-
- [449] Manuel Suárez-Albela, Tiago Fernández-Caramés, Paula Fraga-Lamas, and Luis Castedo. A practical evaluation of a high-security energy-efficient gateway for iot fog computing applications. *Sensors*, 17(9):1978, 2017.
- [450] Yan-Bo Sun. Fasparser: a package for manipulating sequence data. *Zoological research*, 38(2):110, 2017.
- [451] Ewa Surdziel, Ieuan Clay, Florian Nigsch, Anke Thiemeyer, Cyril Allard, Gregory Hoffman, John S Reece-Hoyes, Tanushree Phadke, Romain Gambert, Caroline Gubser Keller, et al. Multidimensional pooled shrna screens in human thp-1 cells identify candidate modulators of macrophage polarization. *PloS one*, 12(8):e0183679, 2017.
- [452] Hajime Suzuki and Masahiro Kasahara. Introducing difference recurrence relations for faster semi-global alignment of long sequences. *BMC bioinformatics*, 19(1):45, 2018.
- [453] Masako Suzuki, Will Liao, Frank Wos, Andrew D Johnston, Justin DeGrazia, Jennifer Ishii, Toby Bloom, Michael C Zody, Soren Germer, and John M Greally. Whole-genome bisulfite sequencing with improved accuracy and cost. *Genome research*, 28(9):1364–1371, 2018.
- [454] Tamas Szalay and Jene A Golovchenko. De novo sequencing and variant calling with nanopores using poreseq. *Nature biotechnology*, 33(10):1087, 2015.
- [455] M Jafar Taghiyar, Jamie Rosner, Diljot Grewal, Bruno M Grande, Radhouane Aniba, Jasleen Grewal, Paul C Boutros, Ryan D Morin, Ali Bashashati, and Sohrab P Shah. Kronos: a workflow assembler for genome analytics and informatics. *GigaScience*, 6(7):gix042, 2017.
- [456] Toshiaki Takeuchi, Atsuo Yamada, Takashi Aoki, and Kunihiro Nishimura. cljam: a library for handling dna sequence alignment/map (sam) with parallel processing. *Source code for biology and medicine*, 11(1):12, 2016.
- [457] Artem Tarasov, Albert J Vilella, Edwin Cuppen, Isaac J Nijman, and Pjotr Prins. Sambamba: fast processing of ngs alignment formats. *Bioinformatics*, 31(12):2032–2034, 2015.
- [458] Sonia Tarazona, Pedro Furió-Tarí, David Turra, Antonio Di Pietro, María José Nueda, Alberto Ferrer, and Ana Conesa. Data quality aware analysis of differential expression in rna-seq with noiseq r/bioc package. *Nucleic acids research*, 43(21):e140–e140, 2015.
- [459] Joaquin Tarraga, Asunción Gallego, Vicente Arnau, Ignacio Medina, and Joaquin Dopazo. Hpg pore: an efficient and scalable framework for nanopore sequencing data. *BMC bioinformatics*, 17(1):107, 2016.

-
- [460] Kedar Tatwawadi, Mikel Hernaez, Idoia Ochoa, and Tsachy Weissman. Gtrac: fast retrieval from compressed collections of genomic variants. *Bioinformatics*, 32(17):i479–i486, 2016.
- [461] Nilesch R Tawari, Justine Jia Wen Seow, Dharuman Perumal, Jack L Ow, Shimin Ang, Arun George Devasia, and Pauline C Ng. Chronqc: a quality control monitoring system for clinical next generation sequencing. *Bioinformatics*, 34(10):1799–1800, 2017.
- [462] Hasan C Tekedar, Attila Karsi, Joseph S Reddy, Seong W Nho, Safak Kalindamar, and Mark L Lawrence. Comparative genomics and transcriptional analysis of flavobacterium columnare strain atcc 49512. *Frontiers in microbiology*, 8:588, 2017.
- [463] Jelena M Telenius, Jim R Hughes, WIGWAM Consortium, et al. Ngseqbasic-a single-command unix tool for atac-seq, dnasei-seq, cut-and-run, and chip-seq data mapping, high-resolution visualisation, and quality control. *bioRxiv*, page 393413, 2018.
- [464] Haotian Teng, Minh Duc Cao, Michael B Hall, Tania Duarte, Sheng Wang, and Lachlan JM Coin. Chiron: translating nanopore raw signal directly into nucleotide sequence using deep learning. *GigaScience*, 7(5):giy037, 2018.
- [465] GP Tiley, RT Kimball, EL Braun, and JG Burleigh. Comparison of the chinese bamboo partridge and red junglefowl genome sequences highlights the importance of demography in genome evolution. *BMC genomics*, 19(1):336, 2018.
- [466] Renata Velozo Timbó, Roberto Coiti Togawa, Marcos MC Costa, David A Andow, and Débora P Paula. Mitogenome sequence accuracy using different elucidation methods. *PloS one*, 12(6):e0179971, 2017.
- [467] Bas Tolhuis and Hans Karten. Validation of an ultra-fast cnv calling tool for next generation sequencing data using mlpa-verified copy number alterations. *BioRxiv*, page 340505, 2018.
- [468] Kim NT Ton, Simone L Cree, Sabine J Gronert-Sum, Tony R Merriman, Lisa K Stamp, and Martin A Kennedy. Multiplexed nanopore sequencing of hla-b locus in māori and pacific island samples. *Frontiers in genetics*, 9, 2018.
- [469] Hande Topa and Antti Honkela. Analysis of differential splicing suggests different modes of short-term splicing regulation. *Bioinformatics*, 32(12):i147–i155, 2016.
- [470] Davoud Torkamaneh, Jérôme Laroche, Maxime Bastien, Amina Abed, and François Belzile. Fast-gbs: a new pipeline for the efficient and highly accurate calling of snps from genotyping-by-sequencing data. *BMC bioinformatics*, 18(1):5, 2017.

-
- [471] Marco Trevisan-Herraz, Inmaculada Jorge, Elena Bonzon-Kulichenko, Pedro Navarro, and Jesús Vázquez. Quixot: quantification and statistics of high-throughput proteomics by stable isotope labelling. *BioRxiv*, page 193607, 2017.
- [472] Mariko Tsuchiya, Kojiro Amano, Masaya Abe, Misato Seki, Sumitaka Hase, Kengo Sato, and Yasubumi Sakakibara. Sharaku: an algorithm for aligning and clustering read mapping profiles of deep sequencing in non-coding rna processing. *Bioinformatics*, 32(12):i369–i377, 2016.
- [473] Junko Tsuji and Zhiping Weng. Dnapi: a de novo adapter prediction algorithm for small rna sequencing data. *PloS one*, 11(10):e0164228, 2016.
- [474] Cath Tyner, Galt P Barber, Jonathan Casper, Hiram Clawson, Mark Diekhans, Christopher Eisenhart, Clayton M Fischer, David Gibson, Jairo Navarro Gonzalez, Luvina Guruvadoo, et al. The ucsc genome browser database: 2017 update. *Nucleic acids research*, 45(D1):D626–D634, 2016.
- [475] Gherman V Uritskiy, Jocelyne DiRuggiero, and James Taylor. Metawrap—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome*, 6(1):158, 2018.
- [476] K Joeri van der Velde, Floris Imhann, Bart Charbon, Chao Pang, David van Enckevort, Mariska Slofstra, Ruggero Barbieri, Rudi Alberts, Dennis Hendriksen, Fleur Kelpin, et al. Molgenis research: advanced bioinformatics data software for non-bioinformaticians. *Bioinformatics*, 35(6):1076–1078, 2018.
- [477] Renaud Vanhoutreuve, Arnaud Kress, Baptiste Legrand, Hélène Gass, Olivier Poch, and Julie D Thompson. Leon-bis: multiple alignment evaluation of sequence neighbours using a bayesian inference system. *BMC bioinformatics*, 17(1):271, 2016.
- [478] Giovanna MM Ventola, Teresa MR Noviello, Salvatore D’Aniello, Antonietta Spagnuolo, Michele Ceccarelli, and Luigi Cerulo. Identification of long non-coding transcripts with feature selection: a comparative study. *BMC bioinformatics*, 18(1):187, 2017.
- [479] Luca Venturini, Shabhonam Caim, Gemy George Kaithakottil, Daniel Lee Mapleson, and David Swarbreck. Leveraging multiple transcriptome assembly methods for improved gene structure annotation. *GigaScience*, 7(8):giy093, 2018.
- [480] John Vivian, Arjun Arkal Rao, Frank Austin Nothhaft, Christopher Ketchum, Joel Armstrong, Adam Novak, Jacob Pfeil, Jake Narkizian, Alden D Deran, Audrey Musselman-Brown, et al. Toil enables reproducible, open source, big biomedical data analyses. *Nature biotechnology*, 35(4):314, 2017.
- [481] Jan Voges, Jörn Ostermann, and Mikel Hernaez. Calq: compression of quality values of aligned sequencing data. *Bioinformatics*, 34(10):1650–1658, 2017.

-
- [482] Roger Volden, Theron Palmer, Ashley Byrne, Charles Cole, Robert J Schmitz, Richard Edward Green, and Christopher Vollmers. R2c2: Improving nanopore read accuracy enables the sequencing of highly-multiplexed full-length single-cell cDNA. *BioRxiv*, page 338020, 2018.
- [483] Michaël Vyverman, Bernard De Baets, Veerle Fack, and Peter Dawyndt. A long fragment aligner called alfalfa. *BMC bioinformatics*, 16(1):159, 2015.
- [484] Sam Wagner, Nadejda Lupolova, David L Gally, and Sally A Argyle. Convergence of plasmid architectures drives emergence of multi-drug resistance in a clonally diverse *Escherichia coli* population from a veterinary clinical care setting. *Veterinary microbiology*, 211:6–14, 2017.
- [485] Ching Man Wai, Jisen Zhang, Tyler C Jones, Chifumi Nagai, and Ray Ming. Cell wall metabolism and hexose allocation contribute to biomass accumulation in high yielding extreme segregants of a *Saccharum* interspecific f2 population. *BMC genomics*, 18(1):773, 2017.
- [486] Jeremiah Wala and Rameen Beroukhim. SeqLib: a c++ api for rapid bam manipulation, sequence alignment and sequence assembly. *Bioinformatics*, 33(5):751–753, 2016.
- [487] Jeremiah Wala, Cheng-Zhong Zhang, Matthew Meyerson, and Rameen Beroukhim. Variantbam: filtering and profiling of next-generation sequencing data using region-specific rules. *Bioinformatics*, 32(13):2029–2031, 2016.
- [488] Mark A Walker, Chandra Sekhar Pedomallu, Akinyemi I Ojesina, Susan Bullman, Ted Sharpe, Christopher W Whelan, and Matthew Meyerson. Gatk pathseq: a customizable computational tool for the discovery and identification of microbial sequences in libraries from eukaryotic hosts. *Bioinformatics*, 34(24):4287–4289, 2018.
- [489] Bo Wang, Lin Wan, Anqi Wang, and Lei M Li. An adaptive decorrelation method removes illumina dna base-calling errors caused by crosstalk between adjacent clusters. *Scientific reports*, 7:41348, 2017.
- [490] Hao Wang, Joel McManus, and Carl Kingsford. Isoform-level ribosome occupancy estimation guided by transcript abundance with ribomap. *Bioinformatics*, 32(12):1880–1882, 2016.
- [491] Hao Wang, Joel McManus, and Carl Kingsford. Accurate recovery of ribosome positions reveals slow translation of wobble-pairing codons in yeast. *Journal of Computational Biology*, 24(6):486–500, 2017.
- [492] Hongwei Wang, Yan Wang, and Zhi Xie. Computational resources for ribosome profiling: from database to web server and software. *Briefings in bioinformatics*, 20(1):144–155, 2017.

-
- [493] Jeremy R Wang, James Holt, Leonard McMillan, and Corbin D Jones. Fmlrc: Hybrid long read error correction using an fm-index. *BMC bioinformatics*, 19(1):50, 2018.
- [494] Meng Wang, Keith M Callenberg, Raymond Dalglish, Alexandre Fedtsov, Naomi K Fox, Peter J Freeman, Kevin B Jacobs, Piotr Kaleta, Andrew J McMurry, Andreas Prlić, et al. hgvs: A python package for manipulating sequence variants using hgvs nomenclature: 2018 update. *Human mutation*, 39(12):1803–1813, 2018.
- [495] Paul PS Wang, Wendy T Parker, Susan Branford, and Andreas W Schreiber. Bam-matcher: a tool for rapid ngs sample matching. *Bioinformatics*, 32(17):2699–2701, 2016.
- [496] Rongjie Wang, Junyi Li, Yang Bai, Tianyi Zang, and Yadong Wang. Bdbg: a bucket-based method for compressing genome sequencing data with dynamic de bruijn graphs. *PeerJ*, 6:e5611, 2018.
- [497] Tao Wang, Guanghua Xiao, Yongjun Chu, Michael Q Zhang, David R Corey, and Yang Xie. Design and bioinformatics analysis of genome-wide clip experiments. *Nucleic acids research*, 43(11):5263–5274, 2015.
- [498] Wei Wang, Jack Smith, Hussein A Hejase, and Kevin J Liu. Non-parametric and semi-parametric support estimation using sequential resampling random walks on biomolecular sequences. In *RECOMB International conference on Comparative Genomics*, pages 294–308. Springer, 2018.
- [499] Yan Wang, Matt Stata, Wei Wang, Jason E Stajich, Merlin M White, and Jean-Marc Moncalvo. Comparative genomics reveals the core gene toolbox for the fungus-insect symbiosis. *mBio*, 9(3):e00636–18, 2018.
- [500] Zhiying Wang, Tsachy Weissman, and Olgica Milenkovic. smallwig: parallel compression of rna-seq wig files. *Bioinformatics*, 32(2):173–180, 2015.
- [501] Zichen Wang and Avi Ma’ayan. An open rna-seq data analysis pipeline tutorial with an example of reprocessing data from a recent zika virus study. *F1000Research*, 5, 2016.
- [502] Christopher M Ward, Hien To, and Stephen M Pederson. ngsreports: An r package for managing fastqc reports and other ngs related log files. *BioRxiv*, page 313148, 2018.
- [503] Joseph Ward, Christian Cole, Melanie Febrer, and Geoffrey J Barton. Almostsignificant: simplifying quality control of high-throughput sequencing data. *Bioinformatics*, 32(24):3850–3851, 2016.
- [504] Andrew S Warren, Cristina Aurrecochea, Brian Brunk, Prerak Desai, Scott Emrich, Gloria I Giraldo-Calderón, Omar Harb, Deborah Hix, Daniel Lawson, Dustin

- Machi, et al. Rna-rocket: an rna-seq analysis resource for infectious disease research. *Bioinformatics*, 31(9):1496–1498, 2015.
- [505] Sven Warris, Elio Schijlen, Henri Van De Geest, Rahulsimham Vegesna, Thamara Hesselink, Bas te Lintel Hekkert, Gabino Sanchez Perez, Paul Medvedev, Kateryna D Makova, and Dick De Ridder. Correcting palindromes in long reads after whole-genome amplification. *BMC genomics*, 19(1):798, 2018.
- [506] Timothy H Webster, Madeline Couse, Bruno M Grande, Eric Karlins, Tanya N Phung, Phillip A Richmond, Whitney Whitford, and Melissa A Wilson Sayres. Identifying, understanding, and correcting technical biases on the sex chromosomes in next-generation sequencing data. *BioRxiv*, page 346940, 2018.
- [507] Caleb Weinreb, Samuel Wolock, and Allon M Klein. Spring: a kinetic interface for visualizing high dimensional single-cell expression data. *Bioinformatics*, 34(7):1246–1248, 2017.
- [508] Matthijs RA Welkers, Marcel Jonges, Rienk E Jeeninga, Marion PG Koopmans, and Menno D de Jong. Improved detection of artifactual viral minority variants in high-throughput sequencing data. *Frontiers in microbiology*, 5:804, 2015.
- [509] Samuel T Westreich, Michelle L Treiber, David A Mills, Ian Korf, and Danielle G Lemay. Samsa2: a standalone metatranscriptome analysis pipeline. *BMC bioinformatics*, 19(1):175, 2018.
- [510] Richard J White, John E Collins, Ian M Sealy, Neha Wali, Christopher M Dooley, Zsofia Digby, Derek L Stemple, Daniel N Murphy, Konstantinos Billis, Thibaut Hourlier, et al. A high-resolution mrna expression time course of embryonic development in zebrafish. *Elife*, 6:e30860, 2017.
- [511] Richard Wilton, Xin Li, Andrew P Feinberg, and Alexander S Szalay. Ar-ioc: Gpu-accelerated alignment of short bisulfite-treated reads. *Bioinformatics*, 34(15):2673–2675, 2018.
- [512] Markus Wolfien, Christian Rimmbach, Ulf Schmitz, Julia Jeannine Jung, Stefan Krebs, Gustav Steinhoff, Robert David, and Olaf Wolkenhauer. Trapline: a standardized and automated pipeline for rna sequencing data analysis, evaluation and annotation. *BMC bioinformatics*, 17(1):21, 2016.
- [513] Xiaogang Wu, Taek-Kyun Kim, David Baxter, Kelsey Scherler, Aaron Gordon, Olivia Fong, Alton Etheridge, David J Galas, and Kai Wang. srnanalyzer—a flexible and customizable small rna sequencing data analysis pipeline. *Nucleic acids research*, 45(21):12140–12151, 2017.
- [514] Dana Wyman and Ali Mortazavi. Transcriptclean: variant-aware correction of indels, mismatches and splice junctions in long-read transcripts. *Bioinformatics*, 35(2):340, 2019.

- [515] Chuan-Le Xiao, Ying Chen, Shang-Qian Xie, Kai-Ning Chen, Yan Wang, Yue Han, Feng Luo, and Zhi Xie. Mecat: fast mapping, error correction, and de novo assembly for single-molecule sequencing reads. *nature methods*, 14(11):1072, 2017.
- [516] Zhengtao Xiao, Rongyao Huang, Xudong Xing, Yuling Chen, Haiteng Deng, and Xuerui Yang. De novo annotation and characterization of the translome with ribosome profiling data. *Nucleic acids research*, 46(10):e61–e61, 2018.
- [517] Zhengtao Xiao, Qin Zou, Yu Liu, and Xuerui Yang. Genome-wide assessment of differential translations with ribosome profiling data. *Nature communications*, 7:11194, 2016.
- [518] Hongyi Xin, Sunny Nahar, Richard Zhu, John Emmons, Gennady Pekhimenko, Carl Kingsford, Can Alkan, and Onur Mutlu. Optimal seed solver: optimizing seed selection in read mapping. *Bioinformatics*, 32(11):1632–1642, 2015.
- [519] Yuting Xing, Gen Li, Zhenguo Wang, Bolun Feng, Zhuo Song, and Chengkun Wu. Gtz: a fast compression and cloud transmission tool optimized for fastq files. *BMC bioinformatics*, 18(16):549, 2017.
- [520] Bo Xu, Changlong Li, Hang Zhuang, Jiali Wang, Qingfeng Wang, Jinhong Zhou, and Xuehai Zhou. Dsa: scalable distributed sequence alignment system using simd instructions. In *Proceedings of the 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*, pages 758–761. IEEE Press, 2017.
- [521] Connie L Xu, Paul G Cantalupo, Maria Teresa Sáenz-Robles, Amy Baldwin, Dan Fitzpatrick, Douglas E Norris, Ethan Jackson, and James M Pipas. Draft genome sequence of a novel rhabdovirus isolated from deinocerites mosquitoes. *Genome Announc.*, 6(19):e01438–17, 2018.
- [522] Kazunori D Yamada. Derivative-free neural network for optimizing the scoring functions associated with dynamic programming of pairwise-profile alignment. *Algorithms for Molecular Biology*, 13(1):5, 2018.
- [523] Andrian Yang, Joshua YS Tang, Michael Troup, and Joshua WK Ho. Scavenger: A pipeline for recovery of unaligned reads utilising similarity with aligned reads. *bioRxiv*, page 345876, 2018.
- [524] Jianfeng Yang, Xiaofan Ding, Xing Sun, Shui-Ying Tsang, and Hong Xue. Samsvm: A tool for misalignment filtration of sam-format sequences with support vector machine. *Journal of bioinformatics and computational biology*, 13(06):1550025, 2015.
- [525] Ruochen Yang, Xi Chen, and Idoia Ochoa. Masscomp, a lossless compressor for mass spectrometry data. *bioRxiv*, page 542894, 2019.
- [526] Lynn Yi, Harold Pimentel, Nicolas L Bray, and Lior Pachter. Gene-level differential analysis at transcript-level resolution. *Genome biology*, 19(1):53, 2018.

-
- [527] Deniz Yorukoglu, Yun William Yu, Jian Peng, and Bonnie Berger. Compressive mapping for next-generation sequencing. *Nature biotechnology*, 34(4):374, 2016.
- [528] Chunjiang Yu, Wentao Wu, Jing Wang, Yuxin Lin, Yang Yang, Jiajia Chen, Fei Zhu, and Bairong Shen. Ngs-fc: A next-generation sequencing data format converter. *IEEE/ACM transactions on computational biology and bioinformatics*, 15(5):1683–1691, 2018.
- [529] Y William Yu, Deniz Yorukoglu, Jian Peng, and Bonnie Berger. Quality score compression improves genotyping accuracy. *Nature biotechnology*, 33(3):240, 2015.
- [530] Daniel R Zerbino and Ewan Birney. Velvet: algorithms for de novo short read assembly using de bruijn graphs. *Genome research*, 18(5):821–829, 2008.
- [531] Feng Zhang, Yulan Lu, Sijia Yan, Qinghe Xing, and Weidong Tian. Sprint: an snp-free toolkit for identifying rna editing sites. *Bioinformatics*, 33(22):3538–3548, 2017.
- [532] Haowen Zhang, Yuandong Chan, Kaichao Fan, Bertil Schmidt, and Weiguo Liu. Fast and efficient short read mapping based on a succinct hash index. *BMC bioinformatics*, 19(1):92, 2018.
- [533] Peng Zhang, Bertrand Boisson, Jean-Laurent Casanova, Laurent Abel, and Yuval Itan. Seqtailor: a user-friendly webserver for the extraction of dna or protein sequences from next-generation sequencing data. *bioRxiv*, page 408625, 2018.
- [534] Sai Zhang, Hailin Hu, Jingtian Zhou, Xuan He, Tao Jiang, and Jianyang Zeng. Rose: a deep learning based framework for predicting ribosome stalling. *Available at SSRN 3155721*, 2018.
- [535] Shi-Jian Zhang, Chenqu Wang, Shouyu Yan, Aisi Fu, Xuke Luan, Yumei Li, Qing Sunny Shen, Xiaoming Zhong, Jia-Yu Chen, Xiangfeng Wang, et al. Isoform evolution in primates through independent combination of alternative rna processing events. *Molecular biology and evolution*, 34(10):2453–2468, 2017.
- [536] Yeting Zhang, Khyati Patel, Tony Endrawis, Autumn Bowers, and Yazhou Sun. A fastq compressor based on integer-mapped k-mer indexing for biologist. *Gene*, 579(1):75–81, 2016.
- [537] Zong-Hong Zhang, Naomi R Wray, and Qiongyi Zhao. Dear-o: Differential expression analysis based on rna-seq data-online. *bioRxiv*, page 069807, 2016.
- [538] Dejian Zhao and Deyou Zheng. Smartcleaner: identify and clean off-target signals in smart chip-seq analysis. *BMC bioinformatics*, 19(1):544, 2018.
- [539] Dengke Zhao, William D Baez, Kurt Fredrick, and Ralf Bundschuh. Riboprop: a probabilistic ribosome positioning algorithm for ribosome profiling. *Bioinformatics*, 2018.

-
- [540] Guoguang Zhao, Cheng Ling, and Donghong Sun. Sparksw: scalable distributed computing system for large-scale biological sequence alignment. In *2015 15th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*, pages 845–852. IEEE, 2015.
- [541] Lu Zhao, Zhimin Liu, Sasha F Levy, and Song Wu. Bartender: a fast and accurate clustering algorithm to count barcode reads. *Bioinformatics*, 34(5):739–747, 2017.
- [542] Yongan Zhao, Haixu Tang, and Yuzhen Ye. Rapsearch2: a fast and memory-efficient protein similarity search tool for next-generation sequencing data. *Bioinformatics*, 28(1):125–126, 2011.
- [543] Xiuwen Zheng, Stephanie M Gogarten, Michael Lawrence, Adrienne Stilp, Matthew P Conomos, Bruce S Weir, Cathy Laurie, and David Levine. Seqarray—a storage-efficient high-performance data format for wgs variant calls. *Bioinformatics*, 33(15):2251–2257, 2017.
- [544] Xiangqun Zheng-Bradley, Ian Streeter, Susan Fairley, David Richardson, Laura Clarke, Paul Flicek, and 1000 Genomes Project Consortium. Alignment of 1000 genomes project reads to reference assembly grch38. *GigaScience*, 6(7):gix038, 2017.
- [545] Daibin Zhong, Eugenia Lo, Xiaoming Wang, Delenasaw Yewhalaw, Guofa Zhou, Harrysone E Atieli, Andrew Githeko, Elizabeth Hemming-Schroeder, Ming-Chieh Lee, Yaw Afrane, et al. Multiplicity and molecular epidemiology of plasmodium vivax and plasmodium falciparum infections in east africa. *Malaria journal*, 17(1):185, 2018.
- [546] Yi Zhong, Theofanis Karaletsos, Philipp Drewe, Vipin T Sreedharan, David Kuo, Kamini Singh, Hans-Guido Wendel, and Gunnar Rätsch. Ribodiff: detecting changes of mrna translation efficiency from ribosome footprints. *Bioinformatics*, 33(1):139–141, 2016.
- [547] Qian Zhou, Xiaoquan Su, Gongchao Jing, Songlin Chen, and Kang Ning. Rna-qc-chain: comprehensive and fast quality control for rna-seq data. *BMC genomics*, 19(1):144, 2018.
- [548] Wei Zhou, Ruilin Li, Shuo Yuan, ChangChun Liu, Shaowen Yao, Jing Luo, and Beifang Niu. Metaspark: a spark-based distributed processing tool to recruit metagenomic reads to reference genomes. *Bioinformatics*, 33(7):1090–1092, 2017.
- [549] Tao Zhu, Chengzhen Liang, Zhigang Meng, Sandui Guo, and Rui Zhang. Gff3sort: a novel tool to sort gff3 files for tabix indexing. *BMC bioinformatics*, 18(1):482, 2017.
- [550] Zexuan Zhu, Linsen Li, Yongpeng Zhang, Yanli Yang, and Xiao Yang. Compmap: a reference-based compression program to speed up read mapping to related reference sequences. *Bioinformatics*, 31(3):426–428, 2014.

-
- [551] Zexuan Zhu, Yongpeng Zhang, Zhen Ji, Shan He, and Xiao Yang. High-throughput dna sequence data compression. *Briefings in bioinformatics*, 16(1):1–15, 2013.
- [552] Eduard Zorita, Pol Cusco, and Guillaume J Filion. Starcode: sequence clustering based on all-pairs search. *Bioinformatics*, 31(12):1913–1919, 2015.