#### The Bioconductor Project: Current Status

Martin Morgan

Roswell Park Cancer Institute Buffalo, NY, USA martin.morgan@roswellpark.org

5 December, 2017

過 ト イヨ ト イヨト

## Bioconductor





Analysis and comprehension of high-throughput genomic data.

- Started 2002
- 1473 *R* packages developed by 'us' and user-contributed.

Well-used and respected.

- 53k unique IP downloads / month.
- 21,700 PubMedCentral citations.

More than 1000 maintainers!



- Packages
- Users
- Web & support sites
- Training, workflows, & meetings
- New package submission
- Release & devel builders
- Funding
- Governance: (annual) Scientific Advisory Board; (monthly) Technical Advisory Board



- Packages
- Users
- Web & support sites
- Training, workflows, & meetings
- New package submission
- Release & devel builders
- Funding
- Governance: (annual) Scientific Advisory Board; (monthly) Technical Advisory Board

- < A > < B > < B >

https://bioconductor.org https://support.bioconductor.org

- Packages
- Users
- Web & support sites
- Training, workflows, & meetings
- New package submission
- Release & devel builders
- Funding
- Governance: (annual) Scientific Advisory Board; (monthly) Technical Advisory Board

・ 同 ト ・ ヨ ト ・ ヨ ト

00	Θ	Iss	ues - Bloconductor/Contributions - Chromium			
			- Bioconduct ×			Martin
← -	2	с	GitHub, Inc. [US] https://github.com/Bioconductor/Contributions/Issues	U.	0 0	
0	1	TN	s repository Search Pull requests Issues Marketplace Explore	*	+-	8-
Q B	io	cor	ductor / Contributions	35	¥r.	n 8
C	0	ode	O Issues 33     In Pull requests ● III Projects ● III Wile ju Insights O Settings			
FI	ter	8 -	Q iscissue iscopen Labels Milestones		New	issue
e		•	3 Open 🗸 517 Closed Author + Labels + Projects + Milestones + As	signer		Sort -
0		0	mCSEAdata 2. review is progress WWRWINGS #002 openied is day ago by perimitational II (), ib of 8	E	1	<b>P</b> •
8		0	Bits         Environment         Environment           4551 opened is day ago by Malgorzabiches         IV, 8 of 8         IV.	g	2	Ри
6		0	CHARGE 2. review in progress RROR 4000 opened 6 days app by Derjamin-Addade IU, 5 of 8	3	2	Ģ a
e		0	BASINET 5. www.king.moderation #547 opened 8 days app by Dickles III, 8 of 8			Ç a
6		0	Numero 2. rostov in programa C2 #548 opened 9 days app by smnd2 ID, 8 of 8	1		<b>1</b> 9
0		•	Key2Enrich: An all-In-one R/Bioconductor package for gene list enrichment analysis and pathway visualization <u>Enrovementory</u> (2000)	1		12

Packages

Users

- Web & support sites
- Training, workflows, & meetings
- New package submission
- Release & devel builders
- Funding
- Governance: (annual) Scientific Advisory Board; (monthly) Technical Advisory Board



National Human Genome Research Institute



# SOUND

- Packages
- Users
- Web & support sites
- Training, workflows, & meetings
- New package submission
- Release & devel builders
- Funding
- Governance: (annual) Scientific Advisory Board; (monthly) Technical Advisory Board

・ 同 ト ・ ヨ ト ・ ヨ ト

- Recognition.
- Access & permanance.
- Interoperability.
- Documentation.
- Support.
- Tested.



- Recognition.
- Access & permanance.
- Interoperability.
- Documentation.
- Support.
- Tested.

#### in Bioc > 12.5 years

(B)

- Recognition.
- Access & permanance.
- Interoperability.
- Documentation.
- Support.
- Tested.

```
git$ grep -1 SummarizedExperiment \
 */DESCRIPTION | wc -1
165
```

A B F A B F

- Recognition.
- Access & permanance.
- Interoperability.
- Documentation.
- Support.
- Tested.

#### Documentation

HTML	R Script	Analyzing RNA-seq data with DESeq2
PDF		Reference Manual
Text		NEWS

3

A B M A B M

- Recognition.
- Access & permanance.
- Interoperability.
- Documentation.
- Support.
- Tested.

	Lates	t Posts		× Bioc-c	level ini	fo Pag 🛛 🗙	y biocon	ductor - T	wit ×			
←	→ C	🔒 Secu	re   ht	tps://support.l	piocond	uctor.org						
	-	sages 🙎	<ul> <li>votes</li> </ul>	• posts • tags • t	ollowing	<ul> <li>bookmarks</li> </ul>	5					
Ĵ	Bio OPEN SOUR	cond	UCT	Or ASI	K QUEST	ΓΙΟΝ	LATES	T <sup>1</sup>	NEWS <sup>1</sup>	JOB		
	Limit <del>•</del>			Sort 👻		Search						
	0	0	1	News: Org	)bs in /	Annotatio	nHub					
	votes	answers	views	orgdb annot	ationhub	news				writt		
	0	1 answer	25 views	Error in file.choose() : file choice cancelled								
	votes			bioconductor	rstudio							
						×	vritten 14 hour	s ago by cla	risbaby • 0 •	updated '		
	0	1	23	error displa	ying b	am files						
	votes	answer	views	genomicalign	nents	writte	n 5 hours ago	by kamal.fa	rtiyal84 • 0 •	updated '		
	1	1	51	Normalizati	on of A	lignment	sTrack / co	verage pl	ot by libra	ry size		
	vote	answer	views	normalization	gviz		written 1 d	ay ago by w	iedemak • 0	updated		
Traffic:	310 users	visited in the	last hour	Appropriate		a? deciar	for multi-f	actor con	anarieon y	within au		

E + 4 E +

- Recognition.
- Access & permanance.
- Interoperability.
- Documentation.
- Support.
- Tested.

🛃 Mu											
> (	C () bioconductor.org/chee	kResults	/3.6/bioc-LATEST/					\$		0 0	
it log napshol	t Date: 2017-11-14 17:00:31 -	This p	age was generated on 2013				.6				
ostsawe				R version			Installed pkg				
	Linux (Ubuntu 16.04.1 LTS)				85) "Short Sum		195				
	Windows Server 2012 R2 Standar		mingw32 / x86_64-w64-mingw32								
	OS X 10.11.6 El Capitan				85) "Short Sum		193	1			
	hostname to see more info about the syste										
TIMEOL ERRC VARNING	Idetus is indicated by one of the fu UniVESTALL_BUALD_CHECK or BU NESTALL_BUALD_OFBUALD BIN SCHECK of package produced w MINSTALL_BUALD_CHECK or BU NESTALL_OF package was not na	ILD BIN of of packag amings ILD BIN of eded (ctcl	package took more than 40 mins a tailed, or CHECK produced erro I package was OK I on glyph to see why)	25 S	Lise the check boxes to show only packages with the selected status brass	Package propaga LEDs YES: Package v version was but NO: Package w (impossible dep	ras propagate sped as not propaga	d because ried becar	it didr	rt previos a problem	usły exte n
TIME OL ERRO VARNING SCINES OL Skippe	UT INSTALL BUILD, CHECK or BU DR INSTALL BUILD, or BUILD BIN S CHECK of package produced w INSTALL BUILD, CHECK or BU	ILD BIN of of packag emings ILD BIN of eded (ctic) je was skij isuit is not	package look more than 40 min f falled, or CHECK produced error package was OK i on glyph to see why) sped because the BUILD step fall available because of an anomaly	ans ver ver led	boxes to show only packages with the selected status types.	LEDs YES: Package w version was but NO: Package w (mpossible dep propagated) UNNEEDED: Pa the repository w to propagate it	ras propagate sped as not propage andencies, or ickage was no th this version	d because red becau version lo t propaga . A versio	i II didr use of i wer the bad bein h burns	rt previou a problem in what is bause it is b is requi	n s airead s airead red in o
TIMEOL ERRC VARNING Skippe Skippe	UT WASTALL BUND, CHECK OF BU MISTALL BUND, OF BUND BIN CHECK of package produced w MINSTALL BUND, CHECK OF BU MINSTALL BUND, CHECK OF BU MINSTALL BUND, BUND BIN OF package CHECK OF BUND BUND OF package NA BUND, CHECK OF BUND BIN OF package NA BUND, CHECK OF BUND BIN OF package	ILD BIN of of packag emings ILD BIN of eded (ctic) je was skij isuit is not	Tpackage bolk more than 40 mins a failed, or CHECK produced error I package was OK ca right to see why) oped because the BUKD stop fail available because of an anomaly bed results.	ans ver ver led	boxes to show only packages with the selected status types. kem	LEDs VES: Package v version was but NO: Package w (impossible dep propagated) UNNEEDED: P the repossibly w to propagate it A crossed-out pack	vas propagate sped as not propaga endencies, or ickage was no th this version sage name inc	d because resion lo d propaga A version loates the	i II didr use of i wer the bad bein h burns	It previou a problem in what is bause it is b is required ge is <u>clea</u>	usty exis n s airead s airead ned in o precates
TIME OL ERRO VARNING SCINES OL Skippe	UT INSTALL INAD CHECK OF BU WINSTALL BUND OF BUILD BU WINSTALL BUND OF BUILD BU WINSTALL CHECK OF BUILD OF BOYCE OF WINSTALL CHECK OF BUILD OF BUILD WINSTALL CHECK OF BUILD BUILT IN OF CHECK OF BUILT BUILT IN OF CHECK OF BUILT IN OF CHECK OF BUILT BUILT IN OF CHECK OF CHECK OF BUILT IN OF CHECK OF CHECK OF BUILT IN OF CHECK OF CHECK OF CHECK OF BUILT IN OF CHECK OF CHECK OF CHECK OF CHECK OF BUILT IN OF CHECK OF C	ILD BIN of of packag emings ILD BIN of eded (ctic) je was skij isuit is not	Tpackage took more than 40 mins taked, or C/HECK produced enror package was OK Long pyth to site withy) oped because the BULD shap that watabble because of an anomaly leaf results.	es grant and a syst	boxes to show only packages with the selected status types. Jam	LEDs YES: Package w version was but NO: Package w (mpossible dep propagated) UNNEEDED: Pa the repository w to propagate it	vas propagate sped andencies, or ickage was no th this version kage name inc	d because red becau version lo t propaga . A versio	itt didr use of i wer the bed been bump packs	rt previou a problem in what is bause it is b is requi	usty exis n s airead s airead ned in o precates
INFOL ERRC ARNING Street Skippi Skon an MMARY Nabec Skay1	UT (INSTALL, BIALD, CHECKY er BL WINSTALL, BUALD, CREUCE BU WINSTALL, BUALD, OF BUALD BUT CHECKY of parkage produced wo GAINSTALL, SUBJ, OF parkage were into in a set of the CREUK BUT BUT DE STATE INSTALL, SUBJ, CHECKY OF BUALD BUT INSTALL, SUBJ, CHECKY OF BUALD BUALD BUALD BUALD INSTALL, SUBJ, CHECKY OF BUALD BUALD BUALD BUALD BUALD INSTALL, SUBJ, SUBJ, CHECKY OF BUALD BUALD BUALD INSTALL, SUBJ,	ILD BIN of of packag emings ILD BIN of eded (ctic) je was skij isuit is not	Jpackage book move than 40 mins Baked, or C/HCK produced end package was OK to regist to see why more than 100 minsteries package was OK to regist to see why in the DULD stop fail washold to because of an internet work of the DULD stop fail washold to because of an internet work of the DULD stop fail work of the DULD stop fails of the DULD stop fails work of th	ed /in the Build Syst /x56_64 tandard / x64	bores to show only packages with the sakeched status types. Nem NVSTALL 0 4 513 0 3 498	HEDS     YEIS: Package v     version was bun     No: Package v     version was bun     No: Package v     version was bun     Impossible dep propagated)     Intereposition v     to propagate i     A crossed-out pack     BURD     BURD     BURD     See 2 331     BURD     See 2 333	ras propagate sped andencies, or ackage was no th this version cage name inc 438 2 16 407 5 18	d because resion lo 4 propaga A versio loates the	i it didr use of i wer the bed be n burns packs	rt previou a problem in what is cause it is a is requir ge is <u>dec</u> BUALD 6	usly exi n s airead s airead ned in o precate sw
Interou Erano Conversion Skippe Interonan	UT (INSTALL, BIALD, CHECKY er BL WINSTALL, BUALD, CREUCE BU WINSTALL, BUALD, OF BUALD BUT CHECKY of parkage produced wo GAINSTALL, SUBJ, OF parkage were into in a set of the CREUK BUT BUT DE STATE INSTALL, SUBJ, CHECKY OF BUALD BUT INSTALL, SUBJ, CHECKY OF BUALD BUALD BUALD BUALD INSTALL, SUBJ, CHECKY OF BUALD BUALD BUALD BUALD BUALD INSTALL, SUBJ, SUBJ, CHECKY OF BUALD BUALD BUALD INSTALL, SUBJ,	ILD BIN of of packag emings ILD BIN of eded (ctic) je was skij isuit is not	Tpackage book more than 400 mine tabled, or C/HECK than 400 mine tabled, or C/HECK than 400 mine or glyph to save withy) available because of an anomaly her result. OS / Arch Linux (Ubunhu 10.04.1 LT5)	ed /in the Build Syst /x56_64 tandard / x64	boxes to show only packages actin the selected status types. Nem INSTALL 0 4 513	HEDS     YEIS: Package v     version was bun     No: Package v     version was bun     No: Package v     version was bun     ImoveBible dep propagated)     Intereposition v     to propagate i     A crossed-out pack     BURD     BURD     BURD     BURD     BURD     BURD     See 2 331     BURD     See 2 332     A 29	vas propagate sped andencies, or ackage was no th this version tage name inc cage name inc cage 12 16	1 because rection to 4 propaga - A versio Icates the ECK 149 120	e it didr use of i wer the bed be n burns packs	rt previou a problem in what is cause it is a is requir ge is <u>dec</u> BUALD 6	usty ext n s airead s atread ned in o precate 3W
ERRO AUENING Skipp R Skipp Skipp R Skipp R Skipp R Skipp Skip	UT INTERLA LIAND, CHECK EN ORIGO EN INTERLA LIAND, ORIGO EN INTERLA CHECK OF provide produces or INTERLA LIAND, CHECK EN INTERLA LIAND, CHECK EN INTERLA EN CHECK OF BUILD BUILD BUILT INTERLA LIAND BUILD BUILT BUILT INTERLA LIAND EN INTERLA INTERLA LIAND EN INTERLA LIAND EN INTERLA LIAND INTERLA LIAND EN INTERLA LIAND EN INTERLA LIAND EN INTERLA INTERLA LIAND EN INTERLA LIAND EN INTERLA LIAND EN INTERLA LIAND EN INTERLA INTERLA LIAND EN INTERLA LIAND EN INTERLA INTERLA LIAND EN INTERLA LIAN	ILD BIN 6 of packag amings (LD BIN o eded (clock ge was skip isuit is not is the deta	jackage bok mon than 40 mms     jackage bok mon than 50 mms     jackage was 0K     us offprif to see with     pord because the BIALD line of that     monthline because the BIALD line of the BIALD line BIALD line of the BIALD line BI	ed /in the Build Syst /x56_64 tandard / x64	Bores fo show           only packages           add black           add black	VES: Peckape w Vices: Peckape w NO: Packape w NO: Packape w NO: Packape w Propagated University of the second propagated the reposition of pack the reposition of pack the reposition of packape the reposition of the second packape of the secon	ras propagate as not propaga endendies, of r ickage was no to the twision cage name inc 436 2 16 407 5 18 426 1 16 P Q B	d because ned because ension to 4 propaga icates the 140 122 140 122 140 122 140 122 140 122 140 122	e it didr use of i wer the bed been n bump packs	rt previou a problem in what is cause it is o is require ge is <u>dec</u> BLATLO E 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	asty ext n s airead s aread red in o arecates sw 1407 1426 <u>Y</u> Z
ERRO AJENING SUNAI AN Skippi B Skippi Skip Skip	UT INVESTUL, BUAD CONCOLOR DE UN INVESTUL, BUAD DE UNID BIN CONCOLOR (BUAD BIN CONCOLOR DE UNID BIN BUAD CONCOLOR DE UNID BIN BUAD CONCOLOR DU DI BIN BUAD CONCOLOR DU DI BIN PUED CONCOLOR DU DI BIN CONCOLOR DE UNID BIN	ILD BIN 6 of packag mings ILD BIN o eded (clock ge was skip isuit is not is the deta	I plackage bok more than 40 mm y I plackage bok more than 5 mm or stype: I bese why post because the BULO2 are but box / Amin Evra (Route 10,0411155) Evra (Route 10,0411555) Evra (Route 10,0411555) Evra (Route 10,0411555) Evra (Route 10,04115555) Evra (Route 10,0411555555555555555555555555555555555	ind (in the Build Syst (x06_64 Iandard /x64 0_64 D_E_F_S	overs to show only packages with the selected status types. Wern NUSTALL 0 4 513 0 3 498 0 2 509 H I 2 K DISTALL	VEE: Peckape w Vee: Peckape w NO: Packape w Propositive dep propositive UNNEEDED: Pa the repositive of propagate by propag	vis propagate yed as not propaga endencies, or schage was no th this version vage name inc ci 4381 2 16 407 5 18 426 1 16 P Q B C	the cause and because version lo a propaga A versio loates the VECK	e it didr use of i wer the bed been n bump packs	IT previou a problem in what is cause it is b is required ge is gas BUILD 5	usby exit n s airead s airead ned in o arecate aw 1407 1426
ERRORARY BRADY AND	UT INVESTUL, BUAD CONCOLOR DE UN INVESTUL, BUAD DE UNID BIN CONCOLOR (BUAD BIN CONCOLOR DE UNID BIN BUAD CONCOLOR DE UNID BIN BUAD CONCOLOR DU DI BIN BUAD CONCOLOR DU DI BIN PUED CONCOLOR DU DI BIN CONCOLOR DE UNID BIN	ILD BIN 6 of packag amings (LD BIN o eded (clock ge was skip isuit is not is the deta	plackage book more than 40 mms     plackage book more than 40 mms     plackage was 0 K     corr dright to see why     post because the BUL/D alog book     post because the BUL/D alog book     post because the BUL/D alog book     post of cause     post of cause     post of cause	ind (in the Build Syst (x06_64 Iandard /x64 0_64 D_E_F_S	overs to show only packages with the selected status types. Wern Not 4 513 0 4 513 0 3 498 0 2 509 H I 3 K INSTALL NotNeede	International Constraints     Versional	visit propagatie speci as not propaga endencies, or r ickage was no th this version that the version the version that the version the version the versio	s because red because version lo a propaga A versio loates the seck 149 12 140 12 163 12 163 12 163 12 163 12 163 12 163 12 163 12 163 12 163 12	e it didr use of i wer the bed been n bump packs	IT previou a problem in what is cause it is is require ge is den BUALD & 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	asty ext n s airead s airead nid in o arecatis w 1407 1426 x x x x x x x x x x x x x x x x x x x
Environment ERRO Arrentino Colonies Skippe Biok on an Biok on Biok o	UT INSTALL BARD CHECKED ON UND BAY INSTALL BARD DE HOLD BAY INSTALL BARD DE HOLD BAY INSTALL STALL DE HOLD BAY INSTALL DE HOLD BAY BARD CHECKED BALD BAY INSTALL DE HOLD BALD BALD BAY INSTALL DE HOLD BALD BALD BALD BAY INSTALL DE HOLD BALD BALD BALD BALD INSTALL DE HOLD BALD BALD BALD BALD INSTALL DE HOLD BALD BALD BALD BALD INSTALL DE HOLD BALD BALD BALD BALD BALD INSTALL DE HOLD BALD BALD BALD BALD BALD BALD INSTALL DE HOLD BALD BALD BALD BALD BALD BALD BALD BA	ILD BIN 6 of packag mings ILD BIN o eded (clock ge was skip isuit is not is the deta	jackage bok more than 40 mm - jackage was CK.     corrections from 50 mm - or stype to see why .     corrections from 50 LLCD rise bok more corrections from 50 LLCD rise bok more who was the 50 LLCD rise bok more who was the 50 LLCD rise bok more who was served to 12 R2 or Core X 10 11 66 Lapiteria XI who was served 2012 R2 or Core X 10 11 66 Lapiteria XI who was served 2012 R2 or Core X 10 11 66 Lapiteria XI who was served 2012 R2 or Core X 10 11 66 Lapiteria XI who was served 2012 R2 or Core X 10 11 66 Lapiteria XI who was served 2012 R2 or core X 10 11 66 Lapiteria XI me _ OS / Arm of _ Limax (Uturula 16.04.1126)	In the Build Byst	overs to show only packages with the selected status types. Wern NUSTALL 0 4 513 0 3 498 0 2 509 H I 2 K DISTALL	International Constraints     Versional	visit propagatie speci as not propaga endencies, or r ickage was no th this version that the version the version that the version the version the versio	the cause and because version lo a propaga A versio loates the VECK	e it didr use of i wer the bed been n bump packs	rt previou a problem in what is cause it is o is require ge is <u>der</u> BLATLO E 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	asty ext n s airead s airead nid in o arecatis w 1407 1426 x x x x x x x x x x x x x x x x x x x
Invited ERRO Variante Schwards Schward	UT INCEAL BALD, CHECK EN ONLD BIN MICRALL, BALD, CHECK EN ONLD BIN MICRALL, BALD, CHECK EN ONLD BIN MICRALL, BALD, CHECK EN ON BIO CHECK EN ONLD BIN of paols of the CHECK END OF SHOLD BIN OF SHOLD BIN OF SHOLD BIN OF SHOLD BIN OF SHOLD BIN OF SHOLD BIN OF SHOLD BIN OF SHOLD BIN OF SHOLD BIN OF SHOLD BIN OF SHOLD BIN OF SHOLD BIN OF SHOLD BIN	ILD BIN o of packag amings ILD BIN o edied (clock pross skip isuit is not isuit is not is not isuit is not is n	jackage bok more than 40 mm - jackage was CK.     corrections from 50 mm - or stype to see why .     corrections from 50 LLCD rise bok more corrections from 50 LLCD rise bok more who was the 50 LLCD rise bok more who was the 50 LLCD rise bok more who was served to 12 R2 or Core X 10 11 66 Lapiteria XI who was served 2012 R2 or Core X 10 11 66 Lapiteria XI who was served 2012 R2 or Core X 10 11 66 Lapiteria XI who was served 2012 R2 or Core X 10 11 66 Lapiteria XI who was served 2012 R2 or Core X 10 11 66 Lapiteria XI who was served 2012 R2 or core X 10 11 66 Lapiteria XI me _ OS / Arm of _ Limax (Uturula 16.04.1126)	In the Build Syst in the Syst in th	overs to show only packages with the selected status (past. wm BVSTALL 0 4 513 0 3 498 0 2 509 H I 3 K BVSTALL NotNeede	International and a second secon	visit propagate special since propaga since propaga indendies, or i cotage was no th this version that this version the the this version the this version the this version the the the the the the the the the the	s because red because version lo a propaga A versio loates the seck 149 12 140 12 163 12 163 12 163 12 163 12 163 12 163 12 163 12 163 12 163 12	e it didr use of i wer the bed been n bump packs	IT previou a problem in what is cause it is is require ge is den BUALD & 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	asty ext s airead s airead ned in o preceite w 1407 1426 x x x x x x x x x x x x x x x
Invited ERRO Variante Schwards Schward	11 Rec14.12, BAAD ("CHECK et al.) WEXTALL, BLACK OF WALCH BIN WEXTALL, BLACK OF WALCH BIN WEXTALL BIN WEX	Hostna mage Hostna Hostna Hostna malbe tokay1 veracr	Jinkarge bok more than 40 mm 2     Jinkarge bok more than 50 mm 2     Jinkarge was OK     or dight to see why     or digh	In the Build Syst in the Syst in th	overs to show only packages with the askected status types. Arm 0 4 513 0 3 496 0 2 509 B 1 2 4 PSTALL NotNeede NotNeede	International and a second secon	nis propagale spect as not propaga as not propaga indencies, of t indencies, o	t because need because recision to a propaga c A viersio locates the neck 149 122 163 12 163 12 S I L NECK OK	e it didr use of i wer the bed been n bump packs	IT previou a problem in what is cause it is is require ge is gas BUILD & BUILD & BUILD &	usty con n s alread s arread s

→

#### Our contributors



• 1060 unique maintainers.

- < A

• • = • • = •

• 791 'first time' authors.

3

# Lessons learned from package reviews I

#### Interoperability

- Use feature × sample SummarizedExperiment, not sample × feature matrix.
- Use paradigms familiar to *Bioconductor* users.
- 2 Reuse
  - Use rtracklayer::import.bed(), not custom parser.
- 8 Robust code
  - Edge cases: seq\_len() / seq\_along(), not 1:n.
  - Code complexity: vapply(), not sapply().
- Performant code
  - Vectorize rather than iterate (for, lapply(), apply() are all iterative).
  - ▶ Reuse (e.g., *matrixStats*) before C / C++ implementation.

・ロト ・回ト ・ヨト ・ヨト

## Lessons learned from package reviews II

- Tested code
  - Essential: evaluated example and vignette code chunks.
  - Desireable: unit tests, e.g., *testthat*.
- Time and space limits.
  - Excessive computation may represent inefficient code.
  - Challenging to identify rich but modest data for illustration.
  - Experiment data packages, work flows, F1000 papers as venues for more expensive / comprehensive reproducible analysis.
- O Ambition
  - Implement essential features well.
  - Avoid dependencies on packages for marginal value.
- In Pretty
  - 'Poetry' with short lines, consistent and ample spacing, standard formatting.

- 4 週 ト - 4 ヨ ト - 4 ヨ ト - -

#### Recent developments

Git!

git clone https://git.bioconductor.org/packages/limma
git clone git@git.bioconductor.org:packages/DESeq2

#### Large Single Cell

- SingleCellExperiment
- HDF5Array

|山田 | 山田 | 山田 |

#### Large single-cell data

```
> sce = TENxBrainData::TENxBrainData()
snapshotDate(): 2017-10-30
> sce
class: SingleCellExperiment
dim: 27998 1306127
metadata(0):
assays(1): counts
rownames: NULL
rowData names(2): Ensembl Symbol
colnames(1306127): AAACCTGAGATAGGAG-1 AAACCTGAGCGGCTTC-1 ...
  TTTGTCAGTTAAAGTG-133 TTTGTCATCTGAAAGA-133
colData names(4): Barcode Sequence Library Mouse
reducedDimNames(0):
spikeNames(0):
```

(人間) とうき くうとう う

#### Large single-cell data

- Chunk-wise iteration (often transparent to the user / developer).
- Marginal summaries in rowData, colData.
- Supporting infrastructure: *ExperimentHub*, *rhdf5*, *HDF5Array*, *DelayedMatrixStats*, *beachmat*.

# Cloud computing

Possible visions

- As now, but 'in the cloud' https://rstudio.cloud.
- Exploit cloud services, e.g., BigQuery.
- Pay-as-you-play use existing *Bioconductor* AMIs or docker containers.
- Integrated with 'third party' compute efforts, e.g., NCI, NIH in the United States.
- Federated data access.

#### **Events**

- CSAMA (training), Brixen / Bressanone, Italy, 8 13 July.
- Bioc2018 Toronto, Canada, 25 27 July.

3

- 4 同 6 4 日 6 4 日 6

# Acknowledgments

Core team: Qian Liu, Valerie Obenchain, Hervé Pagès, Marcel Ramos, Lori Shepherd, Nitesh Turaga, Daniel van Twisk.

Technical advisory board: Vincent Carey, Kasper Hansen, Wolfgang Huber, Robert Gentleman, Rafael Irizzary, Levi Waldron, Michael Lawrence, Sean Davis, Aedin Culhane

Scientific advisory board: Vincent Carey (Brigham & Women's), Wolfgang Huber (EBI), Rafael Irizzary (Dana Farber), Jan Vitek (Northeastern University), Robert Gentleman (23andMe).

Research reported in this presentation was supported by the National Human Genome Research Institute and the National Cancer Institute of the National Institutes of Health under award numbers U41HG004059 and U24CA180996. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

・ロト ・回ト ・ヨト ・ヨト