### Bioc 2014

Sonali Arora

Developer Day, July 30, 2014

Sonali Arora

Developer Day, July 30, 2014 1 / 14

- Website organization
- biocViews
- Videos
- GenomeInfoDb

э

- 4 squares for 'install' ,'learn', 'use', 'develop' motivated by the idea of different Bioconductor Users.
- Another motivation was to make the most accessed material easily accessed from the main landing page/ Home page.

- YouTube Channel http://www.youtube.com/user/bioconductor
- Instructional videos which complement traditional vignettes
- Package Contributors can add a Video: Tag in DESCRIPTION file containing link of their video.

- enrich the biocViews vocabulary
- autocomplete biocViews search box find packages faster
- find by package name / author name / keywords.

- This package provides an interface to access seqlevelsStyle() and their supported mappings for various organisms.
- It supports 10 organisms which can be found with

library(GenomeInfoDb)
names(genomeStyles())

- ## [1] "Arabidopsis\_thaliana"
- ## [3] "Cyanidioschyzon\_merolae"
- ## [5] "Homo\_sapiens"
- ## [7] "Oryza\_sativa"
- ## [9] "Saccharomyces\_cerevisiae" '

"Caenorhabditis\_elegans" "Drosophila\_melanogaster" "Mus\_musculus" "Populus\_trichocarpa" "Zea\_mays"

6 / 14

- Subsetting operation that returns only the 'standard' Chromosomes.
- We define 'standard Chromosomes' as those Chromosomes which represent sequences in the assembly that are not scaffolds.
- Applicable when X has a Seqinfo object.
- multiple organisms supported.

### demo

# library(parathyroidSE) data(exonsByGene)

GRangesList of length 100: SENSG0000000003 GRanges with 17 ranges and 2 seqnames (Rle> [1] X [99883667, 9 [2] X [99885756, 9 [3] X [99885756, 9 [3] X [99887482, 9 [4] X [99887538, 9 [5] X [9988402, 9  [13] X [99890555, 9 [14] X [99891188, 9 [15] X [99891165, 9 [16] X [99891790, 9 [17] X [99891790, 9 [18] X [9989170, 9 [18] X [9980170, 9 [18] X [9980170, 9 [18] X [9080100, 9 [18] X [900000000000000000000000000	ranges sti XIRanges> <br 99884983] 99885863] 99887565] 99887565] 99887565] 99887565] 99887565] 99887565] 99887565] 9988743] 99891686] 99891803]	rand	664096 664097 664098 664099  664106 664108 664109	exon_nam <character: ENSE0000145932; ENSE0000040107; ENSE0000184913; ENSE0000184913; ENSE0000185540; ENSE0000185548; ENSE0000186339</character: 	> 2 2 2 2 2 5	
<pre>[17] X [99894942, 9 &lt;99 more elements&gt; seqlengths:</pre>	-	-   2 3199373 Bioc 20		-	3	

#### std\_exons <- keepStandardChromosomes(exonsByGene)</pre>

GRangesList of length 100: \$ENSG0000000003	
GRanges with 17 ranges and 2 metadata columns:	
segnames ranges strand exon_id exon_name	
<pre><rle> <iranges> <rle>   <integer> <character></character></integer></rle></iranges></rle></pre>	
[1] X [99883667, 99884983] -   664095 ENSE00001459322	
[2] X [99885756, 99885863] - 664096 ENSE00000868868	
[3] X [99887482, 99887565] -   664097 ENSE00000401072	
[4] X [99887538, 99887565] -   664098 ENSE00001849132	
[5] X [99888402, 99888536] -   664099 ENSE00003554016	
[13] X [99890555, 99890743] -   664106 ENSE00003512331	
[14] X [99891188, 99891686] -   664108 ENSE00001886883	
[15] X [99891605, 99891803] -   664109 ENSE00001855382	
[16] X [99891790, 99892101] -   664110 ENSE00001863395	
[17] X [99894942, 99894988] -   664111 ENSE00001828996	
 <99 more elements>	
<a> more elements&gt;</a>	
sealenaths:	
1 2 3 4 5 MT X 18 19 20 21	22 Y
249250621 243199373 198022430 191154276 180915260 16569 155270560 78077248 59128983 63025520 48129895	51304566 59373566
	51501500 55575500

포 🛌 포

#### std\_exons <- sortSeqlevels(std\_exons)</pre>

GRangesList of length 100: \$ENSG0000000003							
GRanges with 17 ranges and	d 2 metadata columns:						
segnames	ranges strand	exon_id	exon_name				
<r1e></r1e>	<iranges> <rle></rle></iranges>	<pre> <integer></integer></pre>	<character></character>				
[1] X [99883667,	, 99884983] -	664095	ENSE00001459322				
[2] X [99885756.	, 99885863] –	664096	ENSE00000868868				
[3] X [99887482,		664097	ENSE00000401072				
[4] X [99887538,			ENSE00001849132				
[5] X [99888402,	, 99888536] -	664099	ENSE00003554016				
[13] X [99890555.			ENSE00003512331				
[14] X [99891188,			ENSE00001886883				
[15] X [99891605.			ENSE00001855382				
[16] X [99891790.			ENSE00001863395				
[17] X [99894942.	, 99894988] -	664111	ENSE00001828996				
<99 more elements>							
<99 more erements>							
seqlengths:							
1 2	3 4	5	6 7	20	21 22	х ү	MT
249250621 243199373 19802	22430 191154276 18091	5260 1711150	067 159138663		29895 51304566 1		16569
2,020022 24020000 2000				1011			20000

电

イロト イロト イヨト イヨト

#### 

ENSG0000000003 ENSG0000000005 ENSG0000000419 ENSG0000000457 ENSG0000000460 ENSG00000000938 ENSG00000000971 ENSG00000001036 ENSG00000								
0 0 0 0 0 0 0 0	0							
ENSG0000001460 ENSG0000001461 ENSG00000001497 ENSG0000001561 ENSG0000001617 ENSG00000001626 ENSG00000001629 ENSG00000001630 ENSG00000	01631							
0 0 0 0 0 0 0 0	0							
ENSG0000002079 ENSG0000002330 ENSG0000002549 ENSG0000002586 ENSG00000002587 ENSG00000002726 ENSG00000002745 ENSG00000002746 ENSG00000	02822							
0 0 0 0 0 0 0 0 0	0							
ENSG0000002919 ENSG0000002933 ENSG0000003056 ENSG0000003096 ENSG0000003137 ENSG00000003147 ENSG00000003249 ENSG00000003393 ENSG00000	03400							
0 0 0 0 0 0 0 0 0	0							
ENSG0000003436 ENSG0000003509 ENSG0000003756 ENSG0000003987 ENSG0000003989 ENSG00000004059 ENSG00000004139 ENSG00000004142 ENSG00000	04399							
0 0 0 0 0 0 0 0	0							
ENSG0000004468 ENSG00000004478 ENSG00000004487 ENSG0000004534 ENSG0000004660 ENSG0000004700 ENSG00000004766 ENSG00000004776 ENSG000000	04777							
	0							
ENSG0000004799 ENSG0000004809 ENSG0000004838 ENSG0000004846 ENSG0000004848 ENSG0000004864 ENSG0000004866 ENSG0000004867 ENSG0000004897 ENSG00000	04939							
0 0 0 0 0 0 0	0							
ENSG0000004961 ENSG00000004975 ENSG00000005001 ENSG0000005007 ENSG00000005020 ENSG0000005022 ENSG00000005059 ENSG00000005073 ENSG00000	05075							
	0							
ENSG0000005102 ENSG0000005108 ENSG00000005156 ENSG00000005175 ENSG00000005187 ENSG00000005189 ENSG00000005194 ENSG00000005206 ENSG00000	J3238							
ENSG0000005249 ENSG0000005302 ENSG0000005339 ENSG0000005379 ENSG0000005381 ENSG0000005421 ENSG00000005436 ENSG00000005448 ENSG00000	05460							
EN200000002543 EN200000002302 EN200000002333 EN200000002373 EN2000000002381 EN2000000002421 EN2000000002439 EN2000000003448 EN2000000	J5469							
warning message:	0							
warning message: In . Seqinfo.mergexy(x, y) :								
Each of the 2 combined objects has sequence levels not in the other:								
a in 'x': 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, X, Y, MT								
- in 'y': chr1, chr2, chr3, chr3, chr5, chr5, chr6, chr7, chr8, chr9, chr10, chr11, chr12, chr14, chr14, chr15, chr16, chr16, chr17, chr18, chr19, chr20								

 - in 'y': chr1, chr2, chr3, chr4, chr5, chr6, chr7, chr8, chr9, chr10, chr11, chr12, chr13, chr14, chr15, chr16, chr17, chr18, chr19, chr20 Make sure to always combine/compare objects based on the same reference genome (use suppresswarrings() to suppress this warring).

< ロ > < 同 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ >

# seqlevelsStyle(std\_exons) <- seqlevelsStyle(reads) cnt2 <- countOverlaps(std\_exons, reads)</pre>

ENSG0000000003 E	NSG0000000005	EN5G0000000419	EN5G0000000457	EN5G0000000460	EN5G0000000938	ENSG0000000971	EN5G0000001036
0	0	0	0	0	0	0	0
ENSG0000001461 E	N5G0000001497	ENSG0000001561	EN5G0000001617	EN5G0000001626	ENSG0000001629	ENSG0000001630	EN5G0000001631
0	0	0	0	0	0	0	0
EN5G0000002549 E	N5G0000002586	EN5G0000002587	EN5G0000002726	EN5G0000002745	ENSG0000002746	ENSG0000002822	EN5G0000002834
0	0	0	0	0	0	0	0
ENSG0000003096 E	NSG0000003137	EN5G0000003147	EN5G0000003249	EN5G0000003393	EN5G0000003400	ENSG0000003402	EN5G0000003436
0	0	0	0	0	0	0	0
ENSG0000003989 E	NSG0000004059	EN5G0000004139	EN5G0000004142	EN5G0000004399	ENSG0000004455	EN5G0000004468	EN5G0000004478
0	0	0	0	0	0	0	0
ENSG0000004700 E	N5G0000004766	EN5G0000004776	EN5G0000004777	EN5G0000004779	ENSG0000004799	ENSG0000004809	EN5G0000004838
0	0	0	0	0	0	0	0
ENSG0000004866 E	N5G0000004897	EN5G0000004939	EN5G0000004948	EN5G0000004961	ENSG0000004975	ENSG0000005001	ENSG0000005007
0	0	0	0	0	0	0	0
ENSG0000005073 E	N5G0000005075	EN5G0000005100	EN5G0000005102	EN5G0000005108	ENSG0000005156	ENSG0000005175	EN5G0000005187
0	0	0	0	0	0	0	0
ENSG0000005238 E	N5G0000005243	EN5G0000005249	EN5G0000005302	EN5G0000005339	ENSG0000005379	ENSG0000005381	EN5G0000005421
0	0	0	0	0	0	0	0

- Arora S, Morgan M, Carlson M and Pages H. GenomeInfoDb: Utilities for manipulating chromosome and other 'seqname' identifiers. R package version 1.1.17.
- Haglund, F., et al (2012). Evidence of a functional estrogen receptor in parathyroid adenomas. J. Clin. Endocrinol. Metab., 97(12), pp. 4631-4639
- Carey V, Harshfield B and Falcon S (2014). biocViews: Categorized views of R package repositories. R package version 1.32.1, http://www.bioconductor.org/packages/release/BiocViews.html

Bioconductor is supported in part by grants NHGRI / NIH U41HG004059 and NSF 1247813

・ロト ・聞 ト ・ ヨ ト ・ ヨ ト



## **Questions?**



## Thank you!