# Cirrus-NGS: Cloud-optimized next generation sequencing primary analysis pipeline



Guorong Xu, Mustafa Guler, Mengyi Liu, Roman Sasik, Amanda Birmingham, Kathleen M. Fisch\*

Center for Computational Biology & Bioinformatics, Department of Medicine

University of California, San Diego, La Jolla, CA

Contact: g1xu@ucsd.edu



#### **Motivation**

Bioinformatics analysis of large-scale next-generation sequencing (NGS) data requires significant compute resources. While cloud computing makes such processing power available on-demand, the administration of dynamic compute clusters is a daunting task for most working biologists. To address this pain point, we have developed cirrus-ngs, a turn-key solution for common NGS analyses using Amazon Web Services (AWS).

## **Pipelines**

Cirrus-ngs currently supports RNA-Seq, miRNA-Seq, ChIP-Seq and WGS/WES data with multiple version of genomes.

#### **Features:**

WGS/WES pipelines include bwa gatk and bwa mutect workflows for germline and somatic variants calling. The analysis steps include fastqc, trim, align, multiqc, sort, dedup, split, postalignment, haplotype, somatic\_variant\_calling, merge, combine\_vcf.

RNA-seq pipelines include star\_rsem, star\_htseq, kallisto and star\_gatk workflows. The analysis steps include fastqc, trim, align\_count, multiqc, merge\_counts and variant\_calling.

ChiP-seq pipeline includes Homer workflow. The analysis steps include fastqc, trim, align, multiqc, make\_tag\_directory, make\_UCSC\_file, find\_peaks, annotate\_peaks, pos2bed, find\_motifs\_genome.

miRNA-seq pipeline includes bowtie2 workflow. The analysis steps include fastqc, trim, cut\_adapt, align\_and\_count, multiqc.

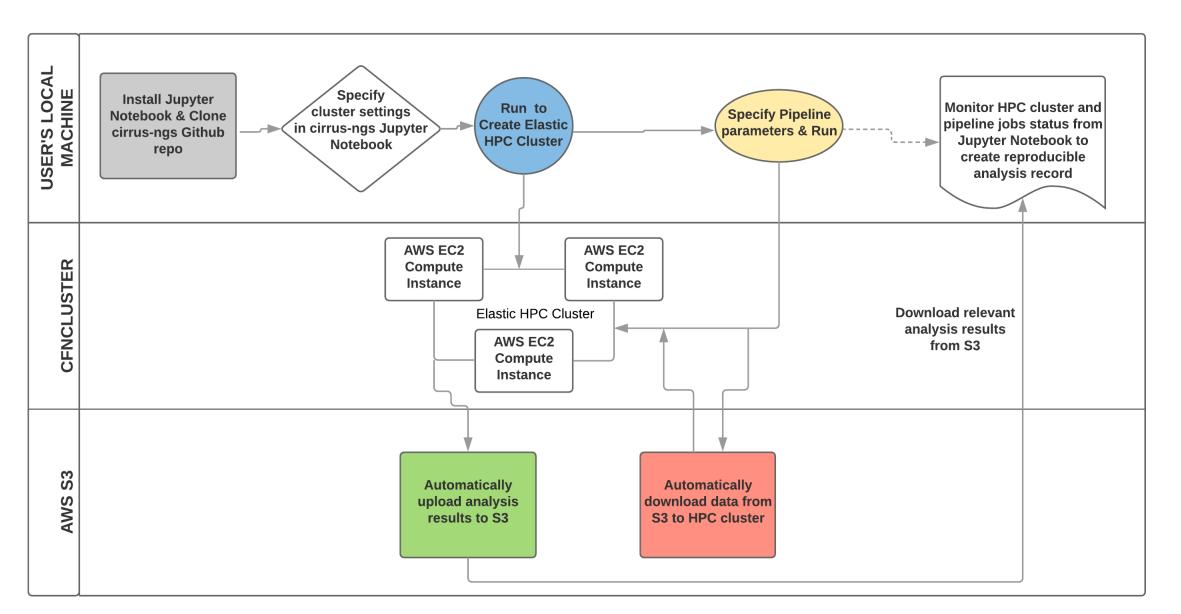
## Reliability:

To avoid data loss, the intermediate results of each step are saved to S3 and users can rerun any step of the pipeline if one step is failed because of spot instances loss or unknown reasons.

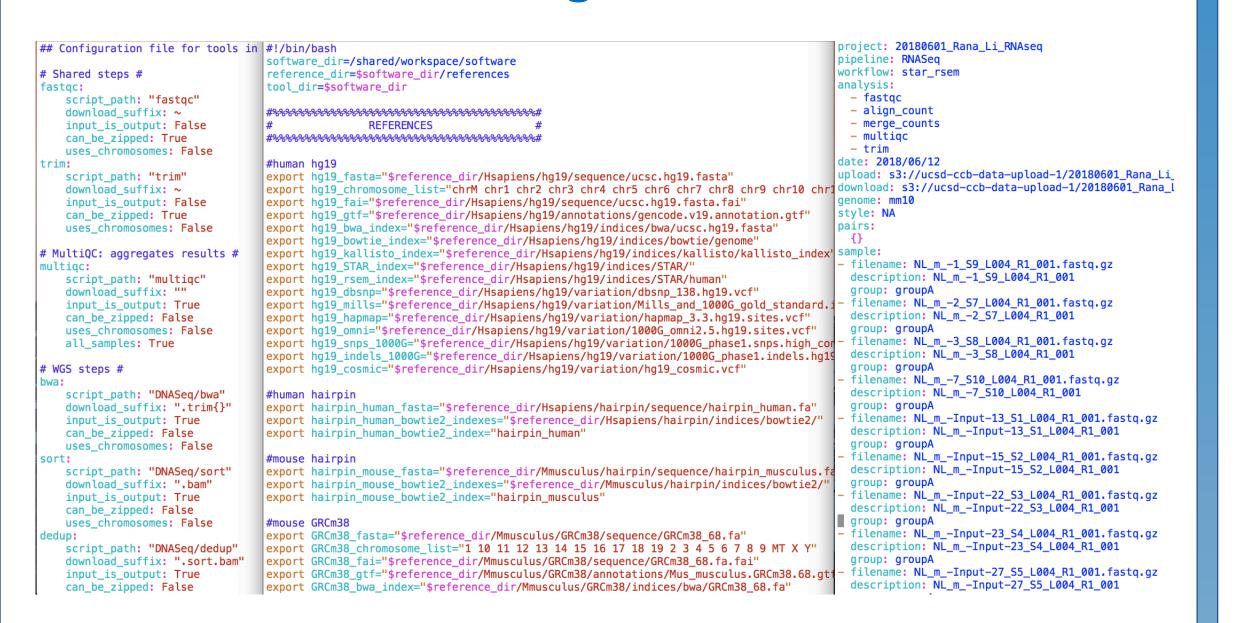
## **Extensibility:**

Each workflow is designed independently but shares the common steps. The developers/users can easily modify or integrate their own codes to Cirrus-ngs by following the design and configuration of framework.

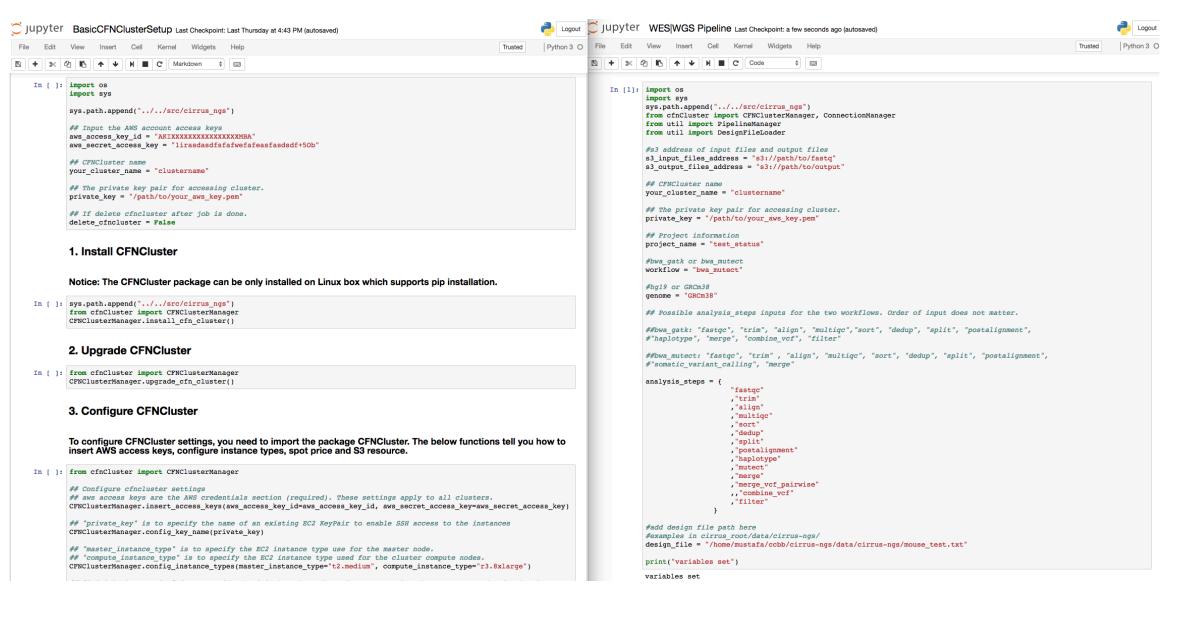
## **Schematic**



# **Configurations**



## **Notebooks**



# **Analysis**

eneral Statistic	s Plot Showing 6/6 ro	vs and 8/10 columns.							
mple Name	Error rate	M Non-Primary	M Reads Mapped		% Mapped	M Total seqs	% Dups	% GC	M Seqs
_S1_L007_R1_001	0.00%	27.6	56.0		97.1%	57.7	43.8%	49%	54.1
_S2_L007_R1_001	0.00%	29.9	60.7		97.0%	62.6	38.9%	49%	59.7
_S3_L007_R1_001	0.00%	28.8	58.1		96.4%	60.3	48.1%	49%	56.4
_S4_L007_R1_001	0.00%	28.5	57.7		96.7%	59.6	41.8%	49%	56.7
_S5_L007_R1_001 _S6_L007_R1_001	0.00%	30.3	59.6		97.0%	61.4	41.7%	49%	57.6 61.1
amtools  attools is a suite of programs for  arcent Mapped  ment metrics from samtools									<b>❷</b> Help
nber of Reads Percentages	stats, mapped vs. unmapp	ed reads.							
Torontages									
7			Samtools s	tats: Alignment	Scores				<b>.</b> Export Plot
1_S1_L007_R1_001									
3_S3_L007_R1_001									
_									-
1_S4_L007_R1_001									
1_S4_L007_R1_001 									
2_S5_L007_R1_001									
-									
2_S5_L007_R1_001	5M 7.5M 10M	12.5M 15M 17.5	M 20M 22.5M 25M 27.5M	30M 32.5M #Read	35M 37.5M 40M	42.5M 45M 47.5	SM 50M 52.5M 55M	4 57.5M 60I	M 62.5M 65M 67
2_S5_L007_R1_001 	5M 7.5M 10M	12.5M 15M 17.5	M 20M 22.5M 25M 27.5M	# Read	S	42.5M 45M 47.5	SM 50M 52.5M 55M	и 57.5М 601	M 62.5M 65M 67
2_S5_L007_R1_001 	M3_S6_L007_R1 M3_S6_L00	7_l M3_S6_L007 M3_S6_L00	Mai     transcript_id gene_id	# Read  pped • Unmapped  M3_S6_L007 M3_S6_L	s d .007 M3_S6_L007 M3_S6_L007	1 item BA2	SM 50M 52.5M 55M  _S2_L007_FBA1_S1_L00_M1_S4_L06780 48664 1659502 447 1941302	007 M3_S6_L007 BA3_!	S3_L00 M2_S5_L007_R1_001.cnt
2_SS_L007_R1_001 3_S6_L007_R1_001 0 2.5M	M3_S6_L007_R1 M3_S6_L00 00001 3262 3188. 00003 799.5 726.	7 M3 S6 L007 M3 S6 L00 52 11580 120.00 02 0 (	Maj	# Read  pped Unmapped  M3_S6_L007 M3_S6_L  01 3262 3188  03 902 828	s d .007 M3_S6_L007 M3_S6_L007	1 item BA2 2 alignment_statistics 1900 3 alignment_certainty 4250	_S2_L007_F BA1_S1_L00 M1_S4_Li 6780 48664 1659502 447 1941302 69810 6094 39008992 57 40132817	007 M3_S6_L007 BA3_ 459 2094120 495 21666 58 43329187 62 39971	53_L00 M2_55_L007_R1_001.cnt 610 460 1866043 47498159 0 493 1456 60 41438084 6060075 29019
2_S5_L007_R1_001  3_S6_L007_R1_001  0 2.5M  ene_id transcript_id(s)  VSMUSG00000000001 ENSMUST000000  VSMUSG000000000028 ENSMUST000000  VSMUSG000000000028 ENSMUST000000  VSMUSG000000000028 ENSMUST000000	M3_S6_L007_R1 M3_S6_L00 00001 3262 3188. 00003 799.5 726. 00028 2014.62 1941. 32294 1708.71 1635.	7_ M3_S6_L007 M3_S6_L00 52 11580 120.08 02 0 0 13 3953 67.33 23 11 0.22	transcript_id   gene_id     ENSMUST0000000001   ENSMUSG00000000000000000000000000000000000	# Read  pped Unmapped  M3_S6_L007 M3_S6_L  101 3262 3188  103 902 828  103 697 623  128 2143 2069	s di	1 item BA2 2 alignment_statistics 1900 3 alignment_certainty 4256 4 alignment_Hits 1359	_S2_L007_F BA1_S1_L00 M1_S4_L0 6780 48664 1659502 447 1941302 69810 6094 39008992 57 40132817 532222 1 124451241 12831512 1906780 1659502 19413	007 M3_S6_L007 BA3_ 159 2094120 495 21666 58 43329187 62 3997 4 1 137292456 1 1286 02 2094120 25	S3_L00 M2_S5_L007_R1_001.cnt 610 466 1866043 47498159 0 493 1456 60 41438084 6060075 29019 14064 1 132324410 1 166610 1866043
2_S5_L007_R1_001  3_S6_L007_R1_001  0 2.5M  ene_id transcript_id(s)  NSMUSG00000000001 ENSMUST000000 NSMUSG0000000003 ENSMUST000000 ENSMUST000000 ENSMUST000000 ENSMUST000000 ENSMUST000000 ENSMUST000000 ENSMUST000000 ENSMUST000000 ENSMUST000000 ENSMUST00000000037 ENSMUST000000 ENSMUST000000 ENSMUST0000000 ENSMUST0000000 ENSMUST0000000 ENSMUST00000000000000000000000000000000000	M3_S6_L007_R1 M3_S6_L00 00001 3262 3188. 00003 799.5 726. 000028 2014.62 1941. 32294 1708.71 1635. 19101 3654.2 3580. 00049 665.5 592.	7_ M3_56_L007 M3_56_L00 52	transcript_id   gene_id	# Read  pped Unmapped  M3_S6_L007 M3_S6_L  01 3262 3188  03 902 828  03 697 623  128 2143 2069  128 1747 1673  128 832 758	s di	1 item BA2 2 alignment_statistics 1900 3 alignment_certainty 4250 4 alignment_Hits 1350 5 0 6 1 7 2	_S2_L007_F BA1_S1_L00 M1_S4_Li 6780 48664 1659502 447 1941302 69810 6094 39008992 57 40132817 532222 1 124451241 12831512 1906780 1659502 19413 18919498 17428562 178898 10256399 9419374 96667	007 M3_S6_L007 BA3_t 459 2094120 495 21666 58 43329187 62 3997: 41 137292456 12861 02 2094120 2: 33 19438339 176 97 10441682 96	S3_L00 M2_S5_L007_R1_001.cnt 610 460 1866043 47498159 0 4930 1456 60 41438084 6060075 29019 14064 1 132324410 1 166610 1866043 881171 18478954 697364 10017658
2_S5_L007_R1_001  B_S6_L007_R1_001  0 2.5M  ne_id transcript_id(s)  ISMUSG00000000001 ENSMUST000000  ISMUSG0000000003 ENSMUST000000  ISMUSG00000000031 ENSMUST000000  ISMUSG000000000037 ENSMUST000000  ISMUSG000000000006 ENSMUST000000  ISMUSG00000000006 ENSMUST000000  ISMUSG000000000006 ENSMUST000000  ISMUSG000000000005 ENSMUST000000	M3_S6_L007_R1 M3_S6_L00 00001 3262 3188. 00003 799.5 726. 000028 2014.62 1941. 32294 1708.71 1635. 19101 3654.2 3580. 00049 665.5 592. 03015 3438.5 3365. 00058 2215.51 2142.	7_ M3_S6_L007 M3_S6_L00 52	1         transcript_id         gene_id           2         ENSMUST00000000001         ENSMUSG000000000           3         ENSMUST0000000003         ENSMUSG000000000           4         ENSMUST00000114041         ENSMUSG000000000           5         ENSMUST00000096990         ENSMUSG000000000           6         ENSMUST00000115585         ENSMUSG000000000           8         ENSMUST00000132294         ENSMUSG000000000           9         ENSMUST00000136359         ENSMUSG0000000000	# Read  pped Unmapped  M3_S6_L007 M3_S6_L  01 3262 3188  03 902 828  03 697 623  28 2143 2069  28 2143 2069  28 1747 1673  28 832 758  31 452 378  31 2286 2212	s di	1 item BA2 2 alignment_statistics 1903 3 alignment_certainty 4904 4 alignment_Hits 1355 5 0 6 1 7 2 8 3 9 4	_S2_L007_F BA1_S1_L00 M1_S4_L0 6780 48664 1659502 447 1941302. 69810 6094 39008992 57 40132817 532222 1 124451241 12831512 1906780 1659502 19413 18919498 17428562 178898 10256399 9419374 96667 6798330 6228177 64288 4460638 4079788 42072	007 M3_S6_L007 BA3_1 459 2094120 495 21666 58 43329187 62 3997 4 1137292456 11286 02 2094120 2: 33 19438339 176 97 10444682 96 96 6892358 65 73 4486541 42	S3_L00 M2_S5_L007_R1_001.cnt 610 460 1866043 47498159 0 493 1456 60 41438084 6060075 29019 14064 1132324410 1 166610 1866043 881171 18478954 697364 10017658 396395 6644609 217579 4340240
0 2.5M  ne_id transcript_id(s) ISMUSG00000000001 ENSMUST000000 ISMUSG000000000031 ENSMUST000000 ISMUSG00000000037 ENSMUST000000 ISMUSG000000000038 ENSMUST000000 ISMUSG000000000058 ENSMUST000000 ISMUSG0000000000058 ENSMUST000000 ISMUSG0000000000058 ENSMUST000000	M3_S6_L007_R1 M3_S6_L00 00001 3262 3188. 00003 799.5 726. 00028 2014.62 1941. 32294 1708.71 1635. 19101 3654.2 3580. 00049 665.5 592. 00049 665.5 592. 00058 2215.51 2142. 00080 4217 4143.	7_ M3_S6_L007 M3_S6_L00 52	transcript_id   gene_id	# Read  pped Unmapped  M3_S6_L007 M3_S6_L  01 3262 3188  03 902 828  03 697 623  228 2143 2069  228 1747 1673  228 832 758  331 452 378  331 2286 2212  331 817 743	s di	1 1 item BA2 2 alignment_statistics 1900 3 alignment_certainty 4256 4 alignment_Hits 1359 5 0 6 1 7 2 8 3 9 4 10 5	_S2_L007_F BA1_S1_L00 M1_S4_L0 6780 48664 1659502 447 1941302 69810 6094 39008992 57 40132817 53222 1 124451241 12831512 1906780 1659502 178898 10256399 9419374 96667 6798330 6228177 64288	007 M3_S6_L007 BA3_: 159 2094120 495 21666 58 43329187 62 3997: 41 137292456 1 1286; 02 2094120 2: 33 19438339 178 97 10441682 99 96 6892358 63 73 4486541 44; 39 2889107 22	53_L00 M2_55_L007_R1_001.cnt 610 460 1866043 47498159 0 493 1456 60 41438084 6060075 29019 14064 1 132324410 1 166610 1866043 881171 18478954 697364 10017658 396395 6644609
ne_id transcript_id(s) SMUSG0000000003 ENSMUST000000 SMUSG000000003 ENSMUST000000 SMUSG0000000003 ENSMUST000000 SMUSG0000000003 ENSMUST000000 SMUSG00000000049 ENSMUST000000 SMUSG00000000049 ENSMUST000000 SMUSG00000000008 ENSMUST000000 SMUSG0000000008 ENSMUST000000 SMUSG00000000008 ENSMUST000000 SMUSG00000000008 ENSMUST000000 SMUSG00000000008 ENSMUST000000 SMUSG000000000088 ENSMUST000000	M3_S6_L007_R1 M3_S6_L00 00001 3262 3188. 00003 799.5 726. 000028 2014.62 1941. 32294 1708.71 1635. 19101 3654.2 3580. 00049 665.5 592. 03015 3438.5 3365. 00058 2215.51 2142. 00080 4217 4143. 00087 1764.06 1690.	7_ M3_S6_L007 M3_S6_L00 52	transcript_id   gene_id     ENSMUST00000000001   ENSMUSG00000000000000000000000000000000000	#Read  pped Unmapped  M3_S6_L007 M3_S6_L  01 3262 3188  03 902 828  03 697 623  128 2143 2069  128 1747 1673  128 832 758  131 452 378  131 2286 2212  131 817 743  131 935 861  131 1853 1779	1007 M3_S6_L007 M3_S6_L007 3.52 11580 120.08 3.52 0 0 5.52 0 0 5.52 2848.13 45.51 5.52 1104.87 21.83 5.52 0 0 5.52 0 0	1 1 item BA2 2 alignment_statistics 1904 3 alignment_certainty 4254 4 alignment_Hits 1355 5 0 6 1 7 2 8 3 9 4 10 5 11 6	_S2_L007_F BA1_S1_L00 M1_S4_L0 6780 48664 1659502 447 1941302 69810 6094 39008992 57 40132817 532222 1 124451241 112831512 1906780 1659502 19413 18919498 17428562 178898 10256399 9419374 96667 6798330 6228177 64288 4460638 4079788 42072 2881646 2638070 27077	007 M3_S6_L007 BA3_t 459 2094120 495 21666 58 43329187 62 3997: 41 137292456 1 1286: 02 2094120 2: 33 19438339 176: 97 10441682 96: 96 6892358 63: 73 4486541 42: 39 2889107 2: 35 1914660 18:	\$3_L00 M2_\$5_L007_R1_001.cnt 610 466 1866043 47498159 0 493 1456 66 41438084 6060075 29019 14064 1 132324410 1 166610 1866043 881171 18478954 697364 10017658 396395 6644609 217579 4340240 714782 2786992
2_SS_L007_R1_001  3_S6_L007_R1_001  0 2.5M  ine_id transcript_id(s) ISMUSG00000000001 ENSMUST000000 ISMUSG0000000003 ENSMUST000000 ISMUSG0000000003 ENSMUST000000 ISMUSG0000000003 ENSMUST000000 ISMUSG0000000003 ENSMUST000000 ISMUSG00000000005 ENSMUST000000 ISMUSG00000000005 ENSMUST000000 ISMUSG00000000008 ENSMUST000000 ISMUSG00000000008 ENSMUST000000 ISMUSG00000000008 ENSMUST000000 ISMUSG00000000088 ENSMUST000000 ISMUSG000000000088 ENSMUST000000 ISMUSG000000000093 ENSMUST000000 ISMUSG000000000093 ENSMUST000000 ISMUSG000000000093 ENSMUST000000 ISMUSG000000000093 ENSMUST000000 ISMUSG000000000093 ENSMUST000000 ISMUSG000000000093 ENSMUST000000	M3_S6_L007_R1 M3_S6_L00 00001 3262 3188. 00003 799.5 726. 000028 2014.62 1941. 32294 1708.71 1635. 19101 3654.2 3580. 00049 665.5 592. 00049 665.5 1343.5 00058 2215.51 2142. 00080 4217 4143. 00087 1764.06 1690. 00090 690 616. 00095 3626 3552. 00096 2959 2885.	7_ M3_S6_L007 M3_S6_L00 52	transcript_id	# Read  pped Unmapped  M3_S6_L007 M3_S6_L  01 3262 3188  03 902 828  03 697 623  28 2143 2069  28 1747 1673  28 832 758  31 452 378  31 452 378  31 817 743  31 935 861  31 1853 1779  337 4842 4768  337 3440 3366	1 1500 M3_S6_L007 M3_S	1 item BA2 2 alignment_statistics 1900 3 alignment_certainty 4256 4 alignment_Hits 1353 5 0 6 11 7 2 8 3 9 4 10 5 11 6 12 7 13 8 14	_S2_L007_F BA1_S1_L00 M1_S4_L0 6780 48664 1659502 447 1941302.6 69810 6094 39008992 57 40132817 532222 1 124451241 112831512 1906780 1659502 19413 18919498 17428562 178898 10256399 9419374 96667 6798330 6228177 64288 4460638 4079788 42072 2881646 2638070 27077 1892559 1738359 17930 1176969 1073808 1136 740805 681500 7118 471352 436480 4510	007 M3_S6_L007 BA3_: 155 2094120 495 21666 158 43329187 62 3997: 14 1 137292456 1 1286: 102 2094120 25: 133 19438339 178 197 10441682 99 196 6892358 63: 173 4486541 44: 173 2889107 27: 175 1914660 18: 179 1182121 11: 175 745507 75: 175 1478348 44:	\$3_L00 M2_S5_L007_R1_001.cnt 610 460 1866043 47498159 0 493 1456 60 41438084 6060075 29019 14064 1 132324410 1 166610 1866043 881171 18478954 697364 10017658 396395 6644609 217579 4340240 714782 2786992 804641 1850496 118876 1145253 714263 722610 457476 459088
2_S5_L007_R1_001  3_S6_L007_R1_001  0 2.5M  transcript_id(s)	M3_S6_L007_R1 M3_S6_L00 00001 3262 3188. 00003 799.5 726. 000028 2014.62 1941. 32294 1708.71 1635. 19101 3654.2 3580. 00049 665.5 592. 03015 3438.5 3365. 00058 2215.51 2142. 00080 4217 4143. 00087 1764.06 1690. 00090 690 616. 00095 3626 3552. 00096 2959 2885. 00122 3446 3372. 00127 3076 3002.	7_ M3_56_L007 M3_56_L00 52	transcript_id   gene_id	# Read  pped Unmapped  M3_S6_L007 M3_S6_L  01 3262 3188  03 902 828  03 697 623  28 2143 2069  28 1747 1673  28 832 758  31 452 378  31 2286 2212  31 817 743  31 935 861  31 1853 1779  37 4842 4768  37 3440 3366  37 3550 3476  37 2953 2879	5 dd	1 1 item BA2 2 alignment_statistics 1904 3 alignment_certainty 4254 4 alignment_Hits 1355 5 0 6 1 7 2 2 8 3 3 9 4 4 10 5 5 11 6 6 12 7 7 13 8 8 14 9 9 15 10 16	_S2_L007_F BA1_S1_L00 M1_S4_L0 6780 48664 1659502 447 1941302 69810 6094 39008992 57 40132817 532222 1 124451241 1 12831512 1906780 1659502 19413 18919498 17428562 178898 10256399 9419374 96667 6798330 6228177 64288 4460638 4079788 42072 2881646 2638070 27077 1892559 1738359 17930 1176969 1073808 11136 740805 681500 7118 471352 436480 4510 303788 279938 2905 191250 176535 1829	007 M3_S6_L007 BA3_ 459 2094120 495 21666 58 43329187 62 3997; 41 137292456 1 1286; 02 2094120 2: 33 19438339 178 97 10441682 96 96 6892358 6; 73 4486541 4; 39 2889107 2: 35 1914660 18 19 1182121 1: 58 745507 7; 58 478348 4; 31 305546 2;	S3_L00 M2_S5_L007_R1_001.cnt 610 460 1866043 47498159 0 4930 1456 60 41438084 6060075 29019 14064 1 132324410 1 166610 1866043 881171 18478954 697364 10017658 396395 6644609 217579 4340240 714782 2786992 804641 1850496 118876 1145253 714263 722610 457476 459088 289990 296204 181517 18985
2_S5_L007_R1_001  3_S6_L007_R1_001  0 2.5M  ene_id transcript_id(s)  SSMUSG0000000001  SSMUSG0000000003  SSMUSG0000000003  SSMUSG0000000003  SSMUSG0000000003  SSMUSG0000000005  SSMUSG0000000005  SSMUSG0000000005  SSMUSG0000000005  SSMUSG0000000008  SSMUSG0000000008  SSMUSG0000000008  SSMUSG0000000008  SSMUSG0000000008  SSMUSG0000000008  SSMUSG0000000008  SSMUSG00000000098  SSMUSG00000000098  SSMUST000000  SSMUSG00000000098  SSMUST000000  SSMUST000000  SSMUST000000  ENSMUST000000  ENSMUST0000000  ENSMUST000000  ENSMUST000000  ENSMUST000000  ENSMUST0000000  ENSMUST000000  ENSMUST0000000  ENSMUST0000000  ENSMUST000000  ENSMUST0000000  ENSMUST0000000  ENSMUST0000000  ENSMUST0000000  ENSMUST0000000  ENSMUST0000000  ENSMUST000000  ENSMUST000000  ENSMUST000000  ENSMUST0000000  ENSMUST0000000  ENSMUST0000000  ENSMUST0000000  ENSMUST00000000000000000000000000000000000	M3_S6_L007_R1 M3_S6_L00 00001 3262 3188. 000028 2014.62 1941. 32294 1708.71 1635. 19101 3654.2 3580. 00049 665.5 592. 03015 3438.5 3365. 00058 2215.51 2142. 00080 4217 4143. 00087 1764.06 1690. 00090 690 616. 00095 3626 3552. 00096 2959 2885. 000122 3446 3372. 00127 3076 3002. 00128 3318.06 3244.	7_ M3_S6_L007 M3_S6_L00 52	transcript_id	# Read  pped Unmapper  M3_56_L007 M3_56_L  01 3262 3188  003 902 828  003 697 623  228 2143 2069  228 2143 2069  228 31747 1673  228 832 758  331 452 378  331 2286 2212  331 817 743  331 935 861  331 935 861  331 935 861  331 935 863  337 3440 3366  337 3440 3366  337 3550 3476  337 2953 2879  337 2931 2857	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 item BA2 2 alignment_statistics 1900 3 alignment_certainty 4254 4 alignment_Hits 1355 5 0 6 11 7 2 8 3 9 4 10 5 11 6 12 7 13 8 14 9 15 10 16 11 17 12	_S2_L007_F BA1_S1_L007 M1_S4_L0 6780 48664 1659502 447 1941302. 69810 6094 39008992 57 40132817 532222 1 124451241 1 12831512 1906780 1659502 19413 18919498 17428562 178898 10256399 9419374 96667 6798330 6228177 64288 4460638 4079788 42072 2881646 2638070 27077 1892559 1738359 17930 1176969 1073808 11136 740805 681500 7118 471352 436480 4510 303788 279938 2905	007 M3_S6_L007 BA3_155 2094120 495 21666 1584 3329187 62 3997: 4 1137292456 11286 02 2094120 2: 33 19438339 176 997 10441682 96 6892358 65 73 4486541 42 39 2889107 2: 335 1914660 18 19 1182121 11 58 745507 58 478348 43 305546 22 194318 558 138035 1	\$3_L00 M2_\$5_L007_R1_001.cnt 610 460 1866043 47498159 0 493 1456 60 41438084 6060075 29019 1466610 1866043 881171 18478954 697364 10017658 396395 6644609 217579 4340240 714782 2786992 804641 1850496 118876 1145253 714263 722610 457476 459088 289990 296204
2_SS_L007_R1_001  3_SS6_L007_R1_001  0 2.5M  ene_id transcript_id(s)  SSMUSG0000000001  SSMUSG0000000003  SSMUSG0000000003  SSMUSG0000000003  SSMUSG0000000003  SSMUSG00000000036  SSMUSG00000000056  SSMUSG00000000056  SSMUSG0000000008  SSMUSG0000000008  SSMUSG0000000008  SSMUSG0000000008  SSMUSG0000000008  SSMUSG0000000008  SSMUSG0000000008  SSMUSG0000000008  SSMUSG0000000009  SSMUSG0000000009  SSMUSG00000000009  SSMUSG000000000125  SSMUSG00000000127  SSMUSG00000000121  SSMUSG00000000121  SSMUST000000  SSMUSG000000000125  SSMUST000000  SSMUST000000  SSMUSG000000000125  SSMUST000000  SSMUSG000000000127  SSMUSG000000000127  SSMUSG00000000121  SSMUST000000  SSMUSG000000000127  SSMUSG000000000127  SSMUST000000  ENSMUST000000  ENSMUST000	M3_S6_L007_R1 M3_S6_L00 00001 3262 3188. 00003 799.5 726. 000028 2014.62 1941. 32294 1708.71 1635. 19101 3654.2 3580. 00049 665.5 592. 03015 3438.5 3365. 00058 2215.51 2142. 00080 4217 4143. 00087 1764.06 1690. 00090 690 616. 00095 3626 3552. 00096 2959 2885. 00122 3446 3372. 00127 3076 3002. 00128 3318.06 3244. 00129 2993.38 2915.	7_ M3_S6_L007 M3_S6_L00 52	Transcript_id	# Read  pped Unmapped  M3_S6_L007 M3_S6_L  101 3262 3188  103 902 828  103 697 623  128 1747 1673  128 832 758  131 452 378  131 2286 2212  131 817 743  131 935 861  131 1853 1779  137 4842 4768  137 3550 3476  137 2953 2879  137 2931 2857  137 4847 4773  137 4847 4773  137 4847 4773  137 4443	5 di	1 1 item BA2 2 alignment_statistics 1900 3 alignment_certainty 4250 4 alignment_Hits 1350 5 0 6 1 7 2 2 8 3 3 9 4 4 10 5 5 11 6 6 12 7 1 13 8 8 14 9 9 15 10 16 11 17 12 18 13	S2_L007_F BA1_S1_L00 M1_S4_L0 6780 48664 1659502 447 1941302 69810 6094 39008992 57 40132817 532222 1 124451241 112831512 1906780 1659502 19413 18919498 17428562 178898 10256399 9419374 96667 6798330 6228177 64288 4460638 4079788 42072 2881646 2638070 27077 1892559 1738359 17930 1176969 1073808 1136 740805 681500 7118 471352 436480 4510 303788 279938 2905 191250 176535 1829 138405 129411 1310 114583 104936 1079 97937 92028 946	007 M3_S6_L007 BA3_: 159 2094120 495 21666 58 43329187 62 3997: 41 137292456 1 1286: 02 2094120 2: 33 19438339 178 97 10441682 99 96 6892358 63 73 4486541 44 39 2889107 2: 35 1914660 18: 19 1182121 1: 58 745507 - 58 478348 4 31 305546 2: 20 194318 3: 58 138035 1: 67 115205 97	\$3_L00 M2_S5_L007_R1_001.cnt 610 466 1866043 47498159 0 4931 1456 66 41438084 6060075 29019 146641 132324410 1 166610 1866043 881171 18478954 697364 10017658 396395 6644609 217579 4340240 714782 2786992 804641 1850496 118876 1145253 714263 722610 457476 459088 289990 296204 181517 189885 134604 135981 109976 112151 95053 97247
2_SS_L007_R1_001  3_S6_L007_R1_001  0 2.5M  transcript_id(s) ENSMUSG0000000001 ENSMUSG00000000001 ENSMUSG00000000000000000000000000000000000	M3_S6_L007_R1 M3_S6_L00 00001 3262 3188. 000028 2014.62 1941. 32294 1708.71 1635. 19101 3654.2 3580. 00049 665.5 592. 03015 3438.5 3365. 00058 2215.51 2142. 00080 4217 4143. 00087 1764.06 1690. 00090 690 616. 00095 3626 3552. 00096 2959 2885. 00096 2959 2885. 00092 3446 3372. 00122 3446 3372. 00122 3446 3272. 00128 3318.06 3244. 00129 2993.38 2915 09344 3836.92 3763. 77680 3142.19 3068.	7_ M3_S6_L007 M3_S6_L00 52	1         transcript_id         gene_id           2         ENSMUST00000000001         ENSMUSG000000000           3         ENSMUST0000000003         ENSMUSG000000000           4         ENSMUST00000114041         ENSMUSG000000000           5         ENSMUST000000128         ENSMUSG000000000           6         ENSMUST00000015585         ENSMUSG000000000           7         ENSMUST00000132294         ENSMUSG000000000           8         ENSMUST00000140716         ENSMUSG000000000           10         ENSMUST00000149774         ENSMUSG000000000           11         ENSMUST00000127754         ENSMUSG000000000           12         ENSMUST00000074802         ENSMUSG000000000           14         ENSMUST00000077375         ENSMUSG000000000           15         ENSMUST00000077375         ENSMUSG000000000           16         ENSMUST000000121345         ENSMUSG000000000           17         ENSMUST00000124775         ENSMUSG000000000           18         ENSMUST00000124775         ENSMUSG000000000           19         ENSMUST000000124775         ENSMUSG000000000           20         ENSMUST000000124775         ENSMUSG000000000           21         ENSMUST000000124775         ENSMUSG0000000000	#Read  pped Unmapper  M3_56_L007 M3_56_L  01 3262 3188  003 902 828  003 697 623  228 2143 2069  228 2143 2069  228 31747 1673  228 832 758  331 452 378  331 935 861  337 3440 3366  337 3440 3366  337 2931 2857  347 4847 473  349 1190 1116  49 367 293	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1   item	_S2_L007_F BA1_S1_L00 M1_S4_L0 6780 48664 1659502 447 1941302 . 69810 6094 39008992 57 40132817 532222 1 124451241 112831512 1906780 1659502 19413 18919498 17428562 178898 10256399 9419374 96667 6798330 6228177 64288 4460638 4079788 42072 2881646 2638070 27077 1892559 1738359 17930 1176969 1073808 11136 740805 681500 7118 471352 436480 4510 303788 279938 2905 191250 176535 1829 1138405 129411 1310 114583 104936 1079 97937 92028 946 45116 41523 436	007 M3_S6_L007 BA3_1 155 2094120 495 21666 158 43329187 62 3997: 14 1 137292456 1 12866 102 2094120 2: 133 19438339 176 196 6892358 66 173 4486541 42 196 389107 2: 191 1182121 1: 158 745507 7: 158 478348 42 191 1182121 1: 158 138035 62 191 1182121 1: 158 138035 63 191 15205 63 194 118205 63 195 118205 63 197 100094 194 45573 199 34701	\$3_L00 M2_\$5_L007_R1_001.cnt 610 460 1866043 47498159 0 493 14064 1132324410 1 166610 1866043 881171 18478954 697364 10017658 396395 6644609 217579 4340240 714782 2786992 804641 1850496 118876 1145253 714263 722610 457476 459088 289990 296204 181517 189885 134604 135981 109976 112151 95053 97247 442551 44655 34494 33933
2_SS_L007_R1_001  3_SS6_L007_R1_001  0 2.5M  ene_id transcript_id(s)  SSMUSG0000000001  SSMUSG00000000001  SSMUSG00000000001  SSMUSG00000000001  SSMUSG0000000000000000000000000000000000	M3_S6_L007_R1 M3_S6_L00 00001 3262 3188. 00003 799.5 726. 000028 2014.62 1941. 32294 1708.71 1635. 19101 3654.2 3580. 00049 665.5 592. 00015 3438.5 3365. 00058 2215.51 2142. 00080 4217 4143. 00087 1764.06 1690. 00090 690 616. 00095 3626 3552. 00096 2959 2885. 00122 3446 3372. 00127 3076 3002. 00128 3318.06 3244. 00129 2993.38 291. 00129 2993.48 291. 00129 399.38 291. 00344 3836.92 3763. 77680 3142.19 3068. 52915 1512.79 1439. 41588 3227.46 3153.	7_ M3_S6_L007 M3_S6_L00 52	Transcript_id	# Read  pped Unmapped  M3_S6_L007 M3_S6_L  01 3262 3188  03 902 828  03 697 623  28 2143 2069  28 2143 2069  28 832 758  31 452 378  31 452 378  31 935 861  31 935 861  31 1853 1779  337 3480 3366  337 350 3476  337 2953 2879  337 2951 2857  337 4847 473  337 517 443  449 1190 1116  449 367 293  449 374 300  449 374 300  449 374 300	5 di	1 item BA2 2 alignment_statistics 1903 3 alignment_certainty 425 4 alignment_Hits 135: 5 0 6 11 7 2 8 3 9 4 10 5 11 6 12 7 13 8 14 9 15 10 16 11 17 12 18 13 19 14 20 15 21 16 22 17 23 18	_S2_L007_F BA1_S1_L00 M1_S4_L0 6780 48664 1659502 447 1941302 69810 6094 39008992 57 40132817 532222 1 124451241 112831512 1906780 1659502 19413 18919498 17428562 178898 10256399 9419374 96667 6798330 6228177 64288 4460638 4079788 42072 2881646 2638070 27077 1892559 1738359 17930 1176969 1073808 1136 740805 681500 7118 471352 436480 4510 303788 279938 2905 191250 176535 1829 138405 129411 1310 114583 104936 1079 97937 92028 946 45116 41523 432 28358 26628 284 21576 20137 216	007 M3_S6_L007 BA3_: 159 2094120 495 21666: 158 43329187 62 3997: 41 137292456 1 1286: 02 2094120 22: 33 19438339 176: 97 10441682 96: 96 6892358 63: 73 4486541 44: 39 2889107 22: 35 1914660 18: 19 1182121 11: 58 745507 7: 58 478348 4: 31 305546 22: 194318 1588 138035 1: 67 115205 97 100094 40 45573 97 100094 40 45573 92 34701 13 28435 26 22884	\$3_L00 M2_S5_L007_R1_001.cnt 610 46C 1866043 47498159 0 4931 1456 6C 41438084 6060075 29019 14064 1 132324410 1 166610 1866043 881171 18478954 697364 10017658 396395 6644609 217579 4340240 714782 2786992 804641 1850496 118876 1145253 714263 722610 457476 459088 289990 296204 181517 189885 134604 135981 109976 112151 95053 97247 42951 44655 34494 33933 27665 28546 21880 22583
Re_id transcript_id(s) SS6_L007_R1_001  0 2.5M  Re_id transcript_id(s) SIMUSG00000000001 SIMUSG00000000031 SIMUSG0000000031 SIMUSG0000000032 SIMUSG0000000035 SIMUSG00000000058 SIMUSG00000000058 SIMUSG00000000058 SIMUSG00000000058 SIMUSG00000000058 SIMUSG00000000058 SIMUSG00000000058 SIMUSG00000000058 SIMUSG000000000058 SIMUSG0000000008 SIMUSG0000000008 SIMUSG00000000008 SIMUSG00000000008 SIMUSG000000000000000000000000000000000000	M3_S6_L007_R1 M3_S6_L00 00001 3262 3188. 00003 799.5 726. 000028 2014.62 1941. 32294 1708.71 1635. 19101 3654.2 3580. 00049 665.5 592. 03015 3438.5 33655. 00058 2215.51 2142. 00080 4217 4143. 00087 1764.06 1690. 00090 690 616. 00090 690 616. 00095 3626 3552. 00095 2959 2885. 00122 3446 3372. 00127 3076 3002. 00128 3318.06 3244. 00129 2993.38 2914. 00129 2993.38 2919. 00124 3836.92 3763. 77680 3142.19 3068. 52915 1512.79 1439. 41588 3227.46 3153. 00153 3416 3342.	7_ M3_S6_L007 M3_S6_L00 52	■ Maj           1         transcript_id         gene_id           2         ENSMUST00000000001         ENSMUSG000000000           3         ENSMUST00000000003         ENSMUSG000000000           4         ENSMUST00000114041         ENSMUSG000000000           5         ENSMUST000000096990         ENSMUSG000000000           6         ENSMUST00000115585         ENSMUSG000000000           8         ENSMUST00000132294         ENSMUSG000000000           9         ENSMUST00000149716         ENSMUSG000000000           10         ENSMUST00000149974         ENSMUSG000000000           12         ENSMUST00000152754         ENSMUSG000000000           13         ENSMUST00000074802         ENSMUSG000000000           14         ENSMUST00000077375         ENSMUSG000000000           15         ENSMUST00000077375         ENSMUSG000000000           16         ENSMUST0000011113         ENSMUSG000000000           17         ENSMUST00000124775         ENSMUSG000000000           18         ENSMUST00000012475         ENSMUSG000000000           20         ENSMUST00000124775         ENSMUSG000000000           21         ENSMUST00000133383         ENSMUSG0000000000           21         ENSMUST00000133383<	#Read  pped Unmapped  M3_S6_L007 M3_S6_L 01 3262 3188 03 902 828 03 697 623 28 2143 2069 28 1747 1673 28 832 758 31 452 378 31 2286 2212 31 817 743 31 935 861 31 1853 1779 31 1853 1779 317 4842 4768 317 3550 3476 317 2953 2879 317 2953 2879 317 4847 4773 317 4849 1190 1116 49 367 293 499 374 300 499 374 300 499 374 305 556 4395 4321	5 dd    10.007 M3_S6_L007 M3_S6_L007   1.52	1 item BA2 2 alignment_statistics 1900 3 alignment_certainty 4254 4 alignment_Hits 1353 5 0 6 11 7 2 2 8 3 3 9 4 4 10 5 5 11 6 6 12 7 7 13 8 8 14 9 9 15 10 16 11 17 12 18 13 19 19 14 20 15 21 16 22 17 23 18	_S2_L007_F BA1_S1_L00 M1_S4_L0 6780 48664 1659502 447 1941302 69810 6094 39008992 57 40132817 532222 1 124451241 12831512 1906780 1659502 19413 18919498 17428562 178898 10256399 9419374 96667 6798330 6228177 64288 4460638 4079788 42072 2881646 2638070 27077 1892559 1733859 17930 1176969 1073808 11136 740805 681500 7118 471352 436480 4510 303788 279938 2905 191250 176535 1829 138405 129411 1310 114583 104936 1079 97937 92028 946 45116 41523 430 34222 31335 322 28358 26828 2844	007 M3_S6_L007 BA3_ 459 2094120 495 21666 58 43329187 62 3997; 41 137292456 1 1286; 02 2094120 2: 33 19438339 178 97 10441682 99 96 6892358 63 73 4486541 44; 39 2889107 2: 35 1914660 18; 19 1182121 1: 58 745507 7; 58 478348 4; 31 305546 2: 20 194318 1; 56 115205 1; 97 100094 40 45573 92 34701 13 28435 26 22884 42 16103	\$3_L00 M2_S5_L007_R1_001.cnt 610 46C 1866043 47498159 0 493 1456 6C 41438084 6060075 29019 1456 6C 41438084 6060075 29019 166610 1866043 881171 18478954 697364 10017658 396395 6644609 217579 4340240 714782 2786992 804641 1850496 118876 1145253 714263 722610 457476 459088 289990 296204 181517 189885 134604 135981 109976 112151 95053 97247 42951 44655 34494 33933 27665 28546
2_S5_L007_R1_001  3_S6_L007_R1_001  0 2.5M  Ene_id transcript_id(s)  NSMUSG00000000001  NSMUSG00000000000  NSMUSG0000000003  NSMUSG0000000003  NSMUSG0000000003  NSMUSG0000000003  NSMUSG0000000005  NSMUSG0000000005  NSMUSG0000000005  NSMUSG0000000005  NSMUSG0000000005  NSMUSG0000000005  NSMUSG0000000008  NSMUSG0000000008  NSMUSG0000000008  NSMUSG0000000008  NSMUSG00000000000  NSMUSG00000000000000000000000000000000000	M3_S6_L007_R1 M3_S6_L00 00001 3262 3188. 00003 799.5 726. 000028 2014.62 1941. 32294 1708.71 1635. 19101 3654.2 3580. 00049 665.5 592. 03015 3438.5 33655. 00058 2215.51 2142. 00080 4217 4143. 00087 1764.06 1690. 00090 690 616. 00090 690 616. 00095 3626 3552. 00095 2959 2885. 00122 3446 3372. 00127 3076 3002. 00128 3318.06 3244. 00129 2993.38 2914. 00129 2993.38 2919. 00124 3836.92 3763. 77680 3142.19 3068. 52915 1512.79 1439. 41588 3227.46 3153. 00153 3416 3342.	7_ M3_S6_L007 M3_S6_L00 52	Transcript_id	# Read    M3_S6_L007 M3_S6_L   D1   3262   3188     D1   3268   2143   2069     28   2143   2069     28   322   758     31   452   378     31   452   378     31   935   861     31   935   861     31   1857   743     31   935   861     31   355   3476     37   4842   4768     37   3550   3476     37   2953   2857     38   2931   2857     38   2931   2857     39   37   443     49   367   293     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     56   4395   4321     56   525   451	1 1580 12007 M3_S6_L007 M3_S6_L00	1 1 item BA2 2 alignment_statistics 1900 3 alignment_certainty 4254 4 alignment_Hits 1355 5 0 6 1 7 2 8 3 3 9 4 4 10 5 5 11 6 6 12 7 7 13 8 8 14 9 9 15 10 16 11 17 12 18 13 19 14 20 15 21 16 22 17 23 18 24 19	S2_L007_F BA1_S1_L00 M1_S4_L0 6780 48664 1659502 447 1941302.6 69810 6094 39008992 57 40132817 532222 1 124451241 112831512 1906780 1659502 19413 18919498 17428562 178898 10256399 9419374 96667 6798330 6228177 64288 4460638 4079788 42072 2881646 2638070 27077 1892559 1738359 17930 1176969 11073808 11136 740805 681500 7118 471352 436480 4510 303788 279938 2905 191250 176535 1829 138405 129411 310 114583 104936 1079 97937 92028 946 45116 41523 436 34222 31335 322 28358 26828 284 21576 20137 216 15586 14369 151 31087 29506 309	007 M3_S6_L007 BA3_ 459 2094120 495 21666 58 43329187 62 3997; 41 137292456 1 1286; 02 2094120 2: 33 19438339 178 97 10441682 99 96 6892358 63 73 4486541 44; 39 2889107 2: 35 1914660 18; 19 1182121 1: 58 745507 7; 58 478348 4; 31 305546 2: 20 194318 1; 56 115205 1; 97 100094 40 45573 92 34701 13 28435 26 22884 42 16103	\$3_L00 M2_S5_L007_R1_001.cnt 610 466 1866043 47498159 0 493 1456 66 41438084 6060075 29019 1466610 1866043 881171 18478954 697364 10017658 396395 6644609 217579 4340240 714782 2786992 804641 1850496 118876 1145253 714263 722610 457476 459088 289990 296204 181517 189885 134604 135981 199976 112151 95053 97247 42951 44655 34494 33933 27665 28546 21880 22583 15683 15298
2_S5_L007_R1_001  3_S6_L007_R1_001  0 2.5M  Ene_id transcript_id(s)  ISMUSG00000000001  ISMUSG00000000003  ISMUSG0000000003  ISMUSG0000000003  ISMUSG0000000003  ISMUSG0000000003  ISMUSG0000000003  ISMUSG0000000003  ISMUSG0000000003  ISMUSG0000000005  ISMUSG0000000005  ISMUSG0000000005  ISMUSG0000000005  ISMUSG0000000005  ISMUSG0000000008  ISMUSG0000000008  ISMUSG0000000003  ISMUSG0000000003  ISMUSG0000000003  ISMUSG0000000003  ISMUSG0000000003  ISMUSG0000000003  ISMUSG00000000012  ISMUSG0000000012  ISMUSG0000000012  ISMUSG0000000012  ISMUSG0000000012  ISMUSG0000000013  ISMUSG0000000014  ISMUSG0000000015  ISMUSG00000000015  ISMUSG00000000015  ISMUSG00000000000000000000000000000000000	M3_S6_L007_R1 M3_S6_L00 00001 3262 3188. 000003 799.5 726. 000028 2014.62 1941. 32294 1708.71 1635. 19101 3654.2 3580. 00049 665.5 395. 00058 2215.51 2142. 00080 4217 4143. 00087 1764.06 1690. 00090 690 616. 00095 3626 3552. 00096 2959 2885. 00122 3446 3372. 00127 3076 3002. 00128 3318.06 3244. 00129 2993.38 2919 00124 3456 3142.19 3068. 52915 1512.79 1439. 41588 3227.46 3153. 00153 3416 3342. 52348 1336.08 1266. 00161 2648.93 2575.	7_ M3_S6_L007 M3_S6_L00 52	transcript_id	# Read    M3_S6_L007 M3_S6_L   D1   3262   3188     D1   3268   2143   2069     28   2143   2069     28   322   758     31   452   378     31   452   378     31   935   861     31   935   861     31   1857   743     31   935   861     31   355   3476     37   4842   4768     37   3550   3476     37   2953   2857     38   2931   2857     38   2931   2857     39   37   443     49   367   293     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     56   4395   4321     56   525   451	1 1580 12007 M3_S6_L007 M3_S6_L00	1 1 item BA2 2 alignment_statistics 1900 3 alignment_certainty 4254 4 alignment_Hits 1355 5 0 6 1 7 2 8 3 3 9 4 4 10 5 5 11 6 6 12 7 7 13 8 8 14 9 9 15 10 16 11 17 12 18 13 19 14 20 15 21 16 22 17 23 18 24 19	S2_L007_F BA1_S1_L00 M1_S4_L0 6780 48664 1659502 447 1941302.6 69810 6094 39008992 57 40132817 532222 1 124451241 112831512 1906780 1659502 19413 18919498 17428562 178898 10256399 9419374 96667 6798330 6228177 64288 4460638 4079788 42072 2881646 2638070 27077 1892559 1738359 17930 1176969 11073808 11136 740805 681500 7118 471352 436480 4510 303788 279938 2905 191250 176535 1829 138405 129411 310 114583 104936 1079 97937 92028 946 45116 41523 436 34222 31335 322 28358 26828 284 21576 20137 216 15586 14369 151 31087 29506 309	007 M3_S6_L007 BA3_: 159 2094120 495 21666 58 43329187 62 3997: 41 137292456 1 1286: 02 2094120 2: 33 19438339 178 97 10441682 99 96 6892358 63 73 4486541 44 39 2889107 2: 35 1914660 18 19 1182121 1: 58 745507 7 58 478348 4 31 305546 2: 20 194318 3: 56 115205 2: 97 100094 40 45573 92 34701 13 28435 26 22884 42 16103 68 32660	\$3_L00 M2_S5_L007_R1_001.cnt 610 466 1866043 47498159 0 493 1456 66 41438084 6060075 29019 1466610 1866043 881171 18478954 697364 10017658 396395 6644609 217579 4340240 714782 2786992 804641 1850496 118876 1145253 714263 722610 457476 459088 289990 296204 181517 189885 134604 135981 199976 112151 95053 97247 42951 44655 34494 33933 27665 28546 21880 22583 15683 15298
2_SS_L007_R1_001  3_S6_L007_R1_001  0 2.5M  transcript_id(s) ENSMUSG00000000001 ENSMUSG0000000003 ENSMUST000000 ENSMUST000000 ENSMUST000000003 ENSMUST000000 ENSMUST0000000003 ENSMUST000000 ENSMUST00000000000000000000000000000000000	M3_S6_L007_R1 M3_S6_L00 00001 3262 3188. 000003 799.5 726. 000028 2014.62 1941. 32294 1708.71 1635. 19101 3654.2 3580. 00049 665.5 395. 00058 2215.51 2142. 00080 4217 4143. 00087 1764.06 1690. 00090 690 616. 00095 3626 3552. 00096 2959 2885. 00122 3446 3372. 00127 3076 3002. 00128 3318.06 3244. 00129 2993.38 2919 00124 3456 3142.19 3068. 52915 1512.79 1439. 41588 3227.46 3153. 00153 3416 3342. 52348 1336.08 1266. 00161 2648.93 2575.	7_ M3_S6_L007 M3_S6_L00 52	Transcript_id	# Read    M3_S6_L007 M3_S6_L   D1   3262   3188     D1   3268   2143   2069     28   2143   2069     28   322   758     31   452   378     31   452   378     31   935   861     31   935   861     31   1857   743     31   935   861     31   355   3476     37   4842   4768     37   3550   3476     37   2953   2857     38   2931   2857     38   2931   2857     39   37   443     49   367   293     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     56   4395   4321     56   525   451	1 1580 12007 M3_S6_L007 M3_S6_L00	1 1 item BA2 2 alignment_statistics 1900 3 alignment_certainty 4254 4 alignment_Hits 1355 5 0 6 1 7 2 8 3 3 9 4 4 10 5 5 11 6 6 12 7 7 13 8 8 14 9 9 15 10 16 11 17 12 18 13 19 14 20 15 21 16 22 17 23 18 24 19	S2_L007_F BA1_S1_L00 M1_S4_L0 6780 48664 1659502 447 1941302.6 69810 6094 39008992 57 40132817 532222 1 124451241 112831512 1906780 1659502 19413 18919498 17428562 178898 10256399 9419374 96667 6798330 6228177 64288 4460638 4079788 42072 2881646 2638070 27077 1892559 1738359 17930 1176969 11073808 11136 740805 681500 7118 471352 436480 4510 303788 279938 2905 191250 176535 1829 138405 129411 310 114583 104936 1079 97937 92028 946 45116 41523 436 34222 31335 322 28358 26828 284 21576 20137 216 15586 14369 151 31087 29506 309	007 M3_S6_L007 BA3_155 2094120 495 21666 158 43329187 62 3997: 4 1137292456 11286; 02 2094120 2: 33 19438339 176 997 10441682 96 6892358 63 4486541 42 39 2889107 2: 355 1914660 18 19 1182121 1: 58 745507 58 478348 42 31 305546 22 194318 158 138035 167 115205 1997 100094 40 45573 99 34701 13 28435 26 22884 42 16103 68 32660   CKD_KineretvsVehicle	\$3_L00 M2_S5_L007_R1_001.cnt 610 466 1866043 47498159 0 493 1456 66 41438084 6060075 29019 1466610 1866043 881171 18478954 697364 10017658 396395 6644609 217579 4340240 714782 2786992 804641 1850496 118876 1145253 714263 722610 457476 459088 289990 296204 181517 189885 134604 135981 199976 112151 95053 97247 42951 44655 34494 33933 27665 28546 21880 22583 15683 15298
2_SS_L007_R1_001  3_S6_L007_R1_001  0 2.5M  transcript_id(s) ENSMUSG00000000001 ENSMUSG0000000003 ENSMUST000000 ENSMUST000000 ENSMUST000000003 ENSMUST000000 ENSMUST0000000003 ENSMUST000000 ENSMUST00000000000000000000000000000000000	M3_S6_L007_R1 M3_S6_L00 00001 3262 3188. 000028 2014.62 1941. 32294 1708.71 1635. 19101 3654.2 3580. 00049 665.5 3595. 00058 2215.51 2142. 00080 4217 4143. 00087 1764.06 1690. 00095 3626 3552. 00096 2959 2885. 000122 3446 3372. 00122 3446 3372. 00122 3496 3352. 00128 3318.06 3244. 00129 2993.38 2919 00344 3836.92 3763. 77680 3142.19 3068. 52915 1512.79 1439. 41588 3227.46 3153. 00153 3416 3342. 52348 1336.08 1267.  ■ **Correspondent Correspondent Corresponden	7_ M3_S6_L007 M3_S6_L00 52	transcript_id	# Read    M3_S6_L007 M3_S6_L   D1   3262   3188     D1   3268   2143   2069     28   2143   2069     28   322   758     31   452   378     31   452   378     31   935   861     31   935   861     31   1857   743     31   935   861     31   355   3476     37   4842   4768     37   3550   3476     37   2953   2857     38   2931   2857     38   2931   2857     39   37   443     49   367   293     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     49   374   300     56   4395   4321     56   525   451	1 1580 12007 M3_S6_L007 M3_S6_L00	1 1 item BA2 2 alignment_statistics 1900 3 alignment_certainty 4254 4 alignment_Hits 1355 5 0 6 1 7 2 8 3 3 9 4 4 10 5 5 11 6 6 12 7 7 13 8 8 14 9 9 15 10 16 11 17 12 18 13 19 14 20 15 21 16 22 17 23 18 24 19		007 M3_S6_L007 BA3_155_2094120 495_21666 158_43329187 62 3997: 4 1137292456 11286: 02 2094120 2: 33 19438339 176 996 6892358 6: 73 4486541 44: 39 2889107 2: 35 1914660 18: 19 1182121 11: 58 745507 : 58 478348 4: 31 305546 2: 20 194318 1: 58 138035 : 67 115205 1: 97 100094 40 45573 92 34701 13 28435 26 22884 42 16103 68 32660	\$3_L00 M2_S5_L007_R1_001.cnt 610 460 1866043 47498159 0 493 14064 1132324410 1 166610 1866043 881171 18478954 697364 10017658 396395 6644609 217579 4340240 714782 2786992 804641 1850496 118876 1145253 714263 722610 457476 459088 298990 296204 181517 189885 134604 135981 109976 112151 95053 97247 42951 44655 34494 33933 27665 28546 21880 22583 15683 15298 31045 31866

Sill4

Axin2

Gnal3

Gpr107

Ed.

Sill4

Sil

X1794, SS
X1795, S8
X1811, S13
X1807, S12
X1806, S10
X1806, S10
X1806, S2
X1789, S7
X1789, S7
X1789, S7
X1789, S7
X1789, S7
X1789, S2
X1

0.20 - 0.20 - 0.20 - 0.15 - 0.15 - 0.05 - 0.00 - 0.

#### Performance

Data Type	Num of Reads (M)	Running Time (H)	Instance Type	Running Cost (\$)
WGS	680	30	r3.8xlarge	15
WES	43	3.6	r3.8xlarge	1.7
RNA-seq	35	2	r3.8xlarge	1
ChiP-seq	45	2	r3.8xlarge	1
miRNA-seq	15	0.5	r3.4xlarge	0.25

Note: the numbers are per sample

## Conclusion

Cirrus-ngs is a lightweight, reproducible tool to perform scalable NGS primary analyses on the cloud. Cirrus-ngs has been optimized for use on AWS HPC clusters, which dynamically scale depending on the compute resources required for the various steps in the pipelines to minimize the per sample compute cost. All computation is performed using AWS EC2 compute instances and all results are uploaded to AWS S3 storage.

# **Availability**

Cirrus-ngs is developed in Python and is available for free use and extension under the MIT License. Source code and extensive documentation are on GitHub at

https://github.com/ucsd-ccbb/cirrus-ngs

# Funding/Disclosures

Supported by the UC San Diego Clinical and Translational Research Institute Grant UL1TR001442.

#### References

- [1] 1000 Genomes Project Consortium. 2015. A global reference for human genetic variation. **Nature** 526:68-74. doi: 10.1038/nature15393
- [2] cfncluster (https://github.com/awslabs/cfncluster)
- [3] H. Li, R. Durbin Fast and accurate short read alignment with Burrows-Wheeler transform **Bioinformatics**, 25 (2009), pp. 1754-1760
- [4] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G.Abecasis, R. Durbin, 1000 Genome Project Data Processing Subgroup The Sequence Alignment/Map format and SAMtools. **Bioinformatics**, 25 (2009), pp. 2078-2079
- [5] A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K.Garimella, D. Altshuler, S. Gabriel, M. Daly, M.A. DePristoThe Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data **Genome Res.**, 20 (2010), pp. 1297-1303
- [6] D. Koboldt, D. Larson, R. Wilson. Using VarScan 2 for Germline Variant Calling and Somatic Mutation Detection. Curr Protoc **Bioinformatics**. 2013;44:15.14.11–17. doi: 10.1002/0471250953.bi1504s44.
- [7] Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat. Biotechnol. 31, 213–219 (2013).
- [8] Z. Lai, A. Markovets, M. Ahdesmaki, et al. VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research **Nucleic Acids Res**, 44 (2016), p. E108
- [9] K. Wang, M. Li & H. Hakonarson ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. **Nucleic Acids Res**. 38, e164 (2010)
- [10] S. Heinz et al. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. **Mol Cell** 38(4):576-589. PMID: 20513432 (2010)
- [11] N. Bray et al. Near-optimal probabilistic RNA-seq quantification, **Nature Biotechnology 34**, 525–527 (2016), doi:10.1038/nbt.3519