Single Cell Sequencing

Callum J. Bell, Ph.D.

New Mexico INBRE
IDeA Networks of Biomedical Research Excellence

NCGR
National Center for Genome Resources

Single cell transcriptomics in plants

New Mexico INBRE
IDeA Networks of Biomedical Research Excellence

WPI

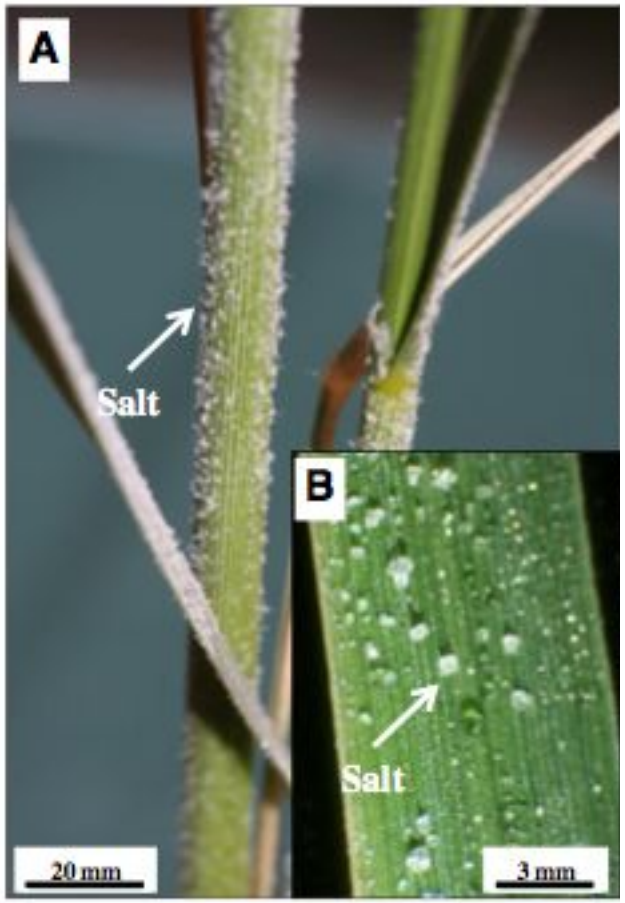
NCGR
National Center for Genome Resources

Salinity Effects on Germination and Plant Growth of Prairie Cordgrass and Switchgrass

Sumin Kim · A. Lane Rayburn · Thomas Voigt · Allen Parrish · D. K. Lee

S. Kim · A. L. Rayburn · T. Voigt · A. Parrish · D. K. Lee (✉)
Department of Crop Sciences, University of Illinois,
Urbana, IL 61801, USA
e-mail: leedk@illinois.edu

Bioenerg. Res. (2012) 5:225–235
DOI 10.1007/s12155-011-9145-3

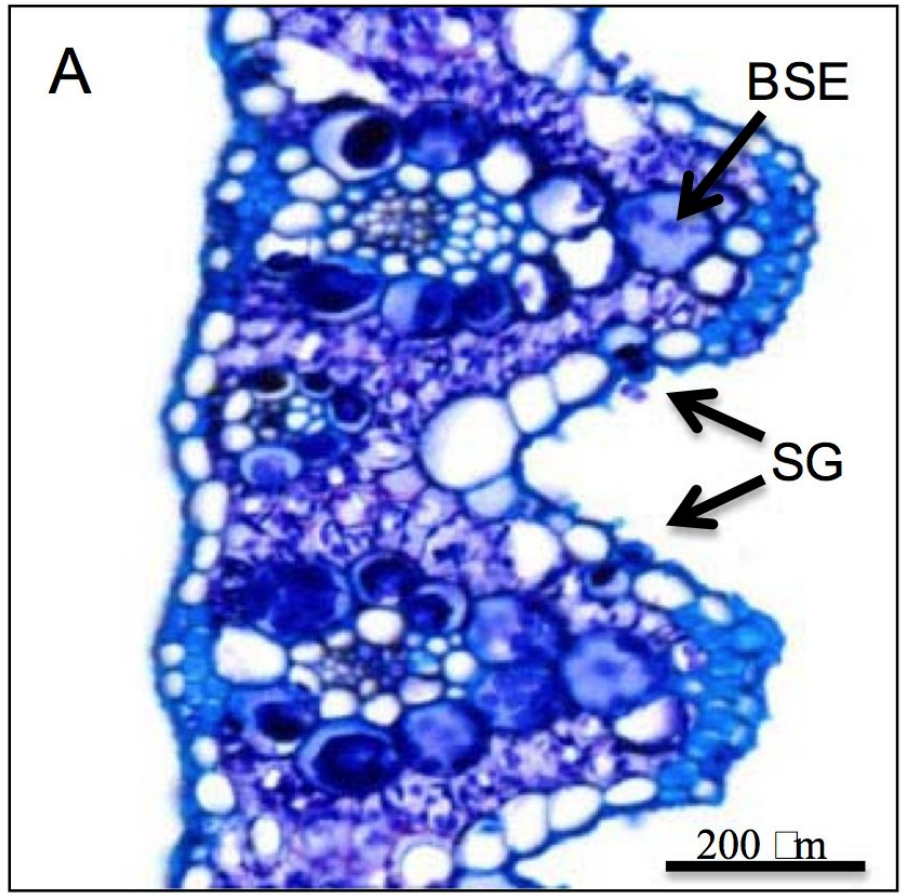


Salinity Effects on Germination and Plant Growth of Prairie Cordgrass and Switchgrass

Sumin Kim · A. Lane Rayburn · Thomas Voigt · Allen Parrish · D. K. Lee

S. Kim · A. L. Rayburn · T. Voigt · A. Parrish · D. K. Lee (✉)
Department of Crop Sciences, University of Illinois,
Urbana, IL 61801, USA
e-mail: leedk@illinois.edu

Bioenerg. Res. (2012) 5:225–235
DOI 10.1007/s12155-011-9145-3



> Nat Methods. 2009 May;6(5):377-82. doi: 10.1038/nmeth.1315. Epub 2009 Apr 6.

mRNA-Seq whole-transcriptome analysis of a single cell

Fuchou Tang [1], Catalin Barbacioru, Yangzhou Wang, Ellen Nordman, Clarence Lee, Nanlan Xu, Xiaohui Wang, John Bodeau, Brian B Tuch, Asim Siddiqui, Kaiqin Lao, M Azim Surani

Affiliations + expand

PMID: 19349980 DOI: 10.1038/nmeth.1315

Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 1663–1667, March 1990
Biochemistry

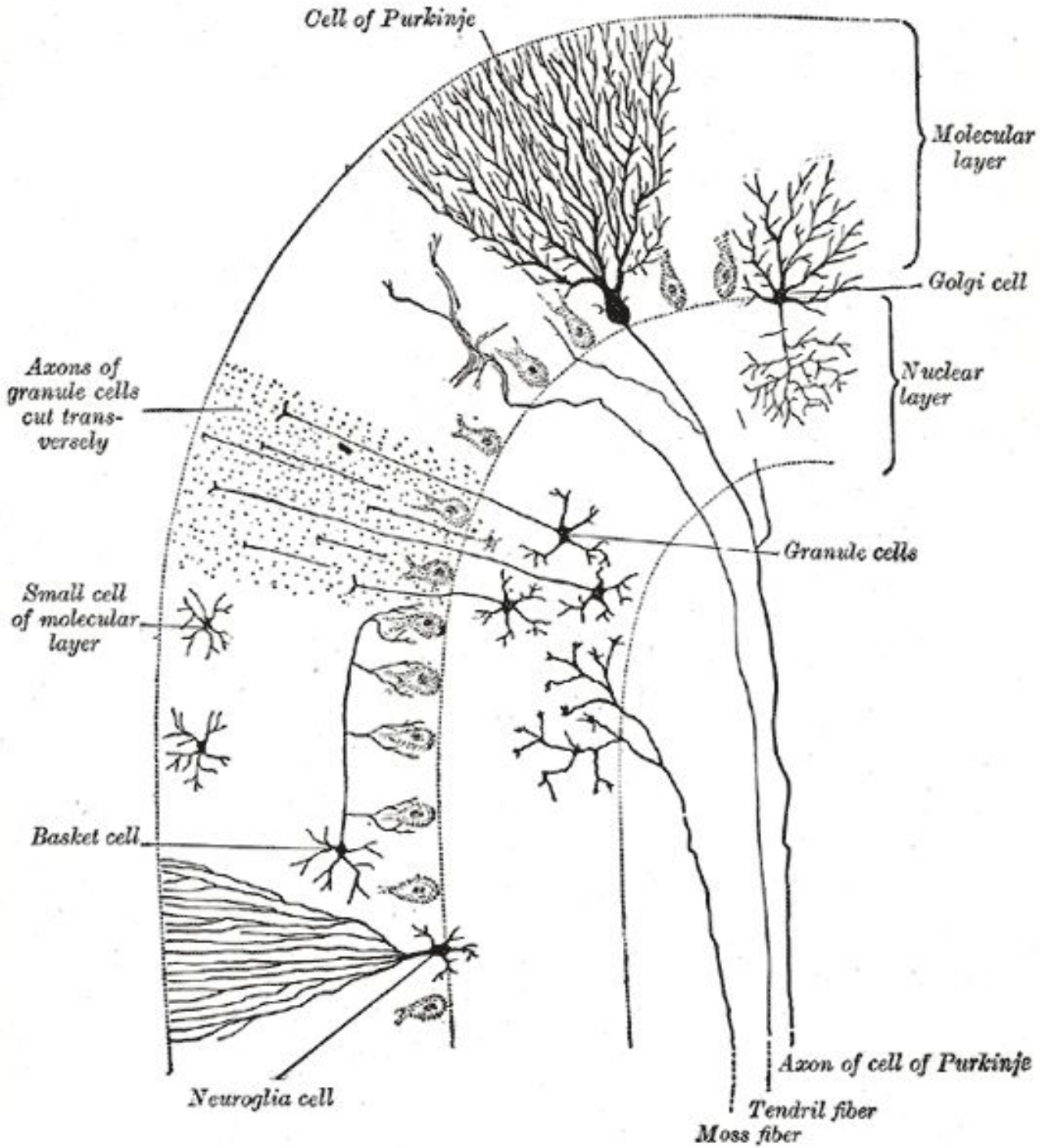
Amplified RNA synthesized from limited quantities of heterogeneous cDNA

(cerebellum/guanine nucleotide-binding protein/T7 RNA polymerase/Purkinje cell)

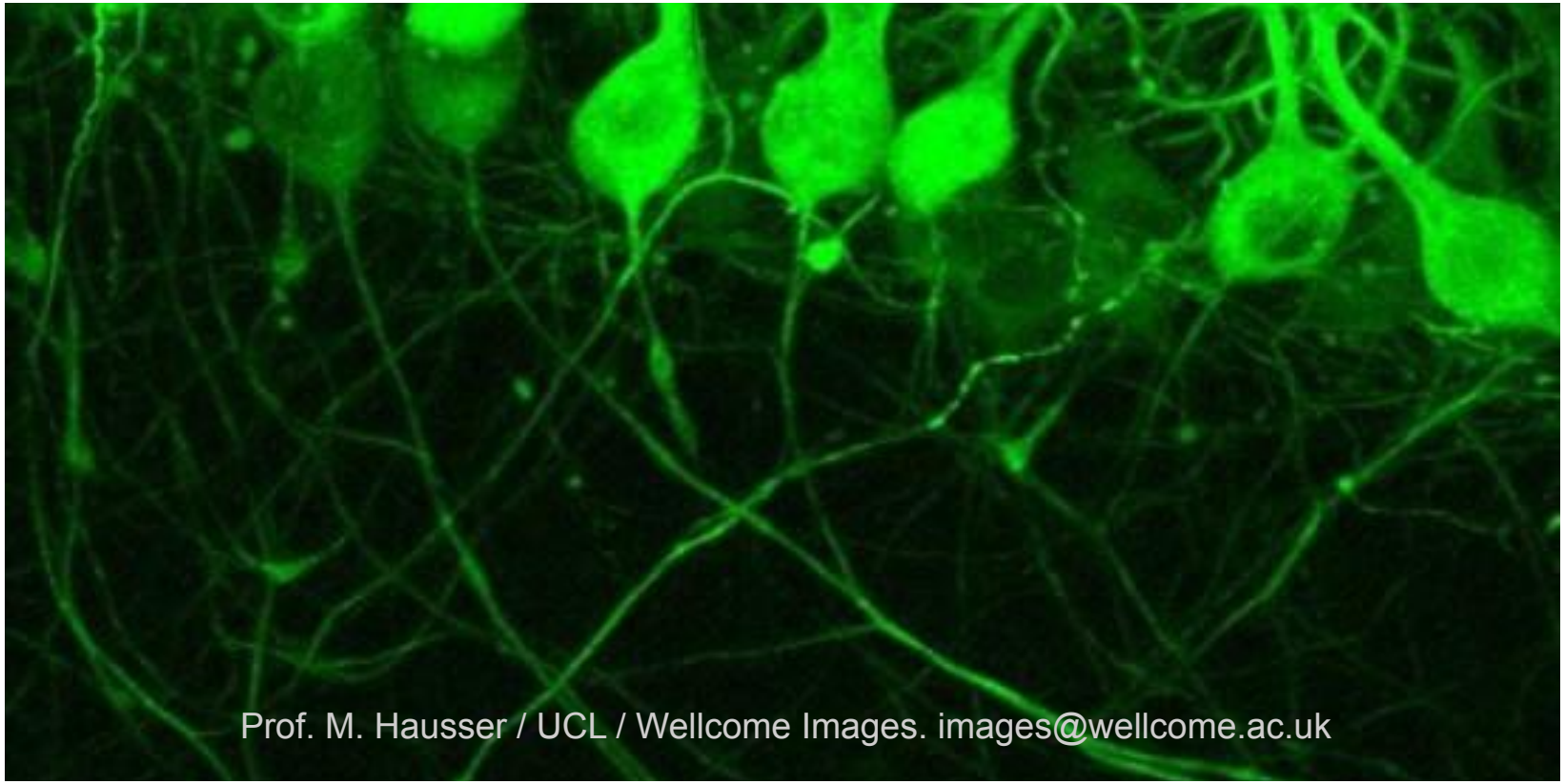
RUSSELL N. VAN GELDER*†, MARK E. VON ZASTROW‡, ANDREA YOOL§, WILLIAM C. DEMENT*, JACK D. BARCHAS‡, AND JAMES H. EBERWINE‡¶

‡Nancy Pritzker Laboratory of Behavioral Neurochemistry, *Sleep Research Laboratory, Department of Psychiatry, and §Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, CA 94305

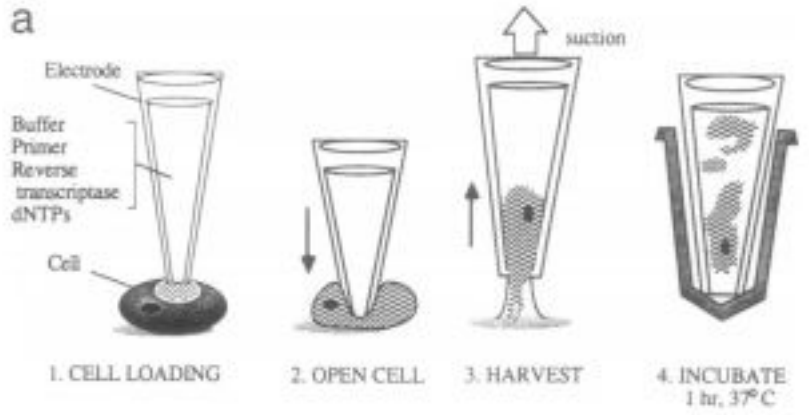
Communicated by Seymour Benzer, November 13, 1989



Purkinje cell bodies



Van Gelder *et al.* 1990 Fig. 5a

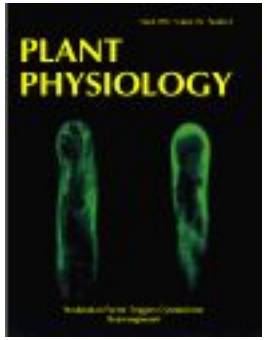


Plant Physiol. (1998) 116: 871–877

Rearrangement of Actin Microfilaments in Plant Root Hairs Responding to *Rhizobium etli* Nodulation Signals[1]

Luis Cárdenas, Luis Vidali, Jimena Domínguez, Héctor Pérez[2], Federico Sánchez, Peter K. Hepler, and Carmen Quinto*

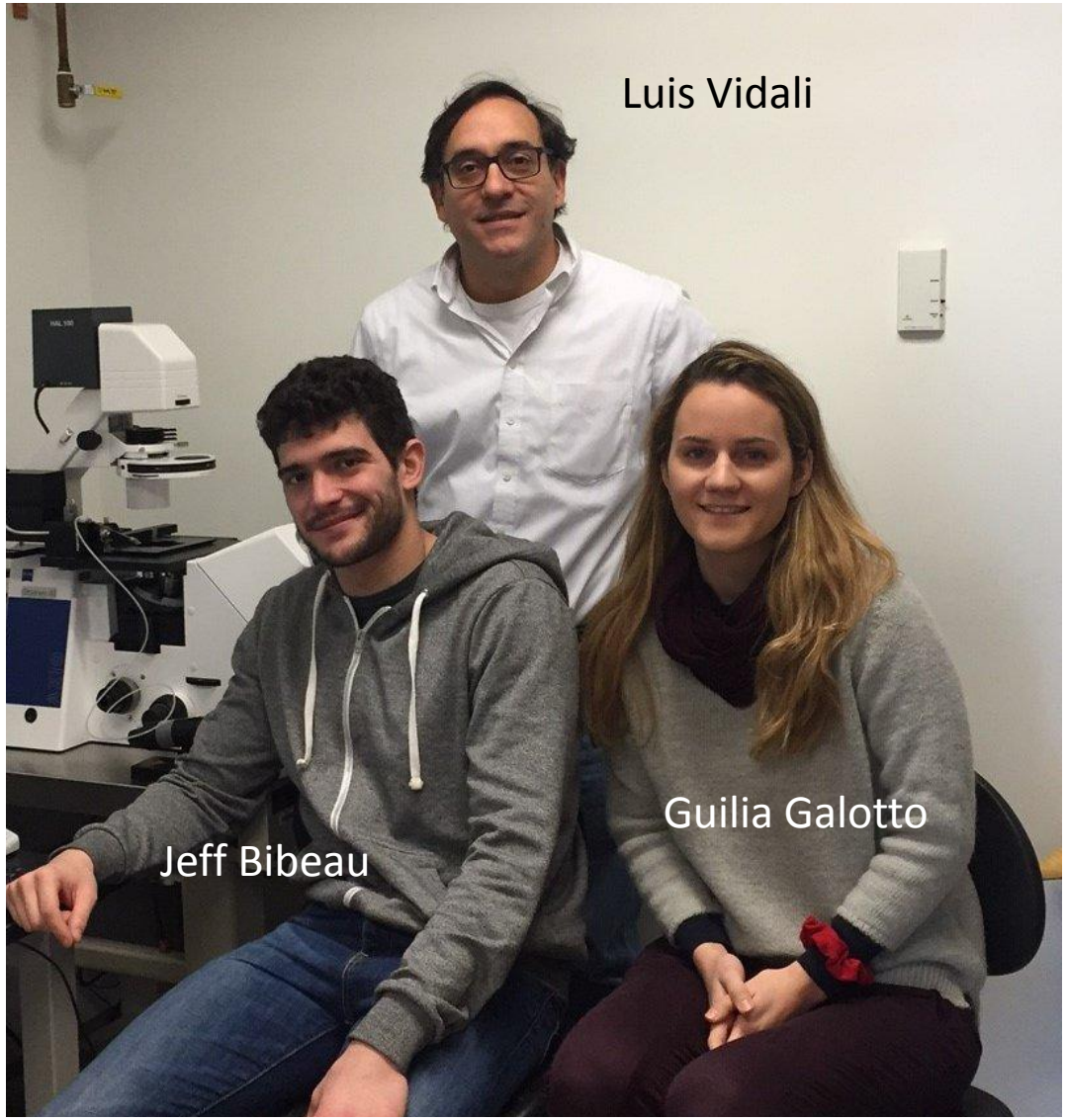
Departamento de Biología Molecular de Plantas, Instituto de Biotecnología, Universidad Nacional Autonoma de Mexico, Apartado Postal 510–3, Cuernavaca Morelos 62271, Mexico (L.C., J.D., H.P., F.S., C.Q.); and Biology Department, Morrill Science Center, University of Massachusetts, Box 35810, Amherst, Massachusetts 01003–5810 (L.V., P.K.H.)



Department of Biology and Biotechnology

Worcester Polytechnic Institute

WPI

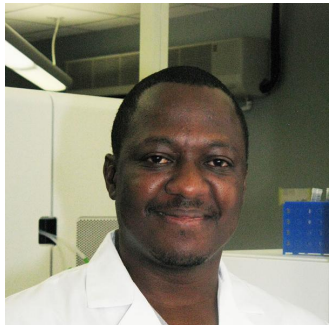


New Mexico INBRE
IDeA Networks of Biomedical Research Excellence

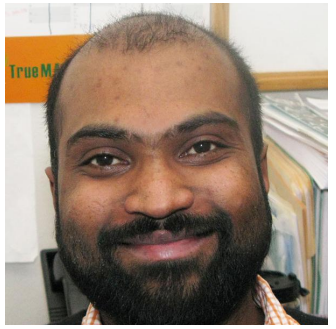
NCGR
National Center for Genome Resources

Jennifer Jacobi

Pooja Umale

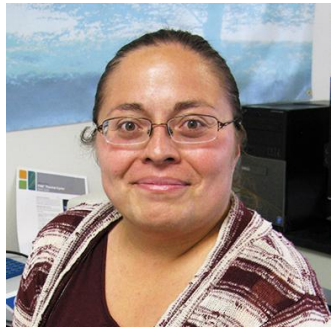


Peter Ngam

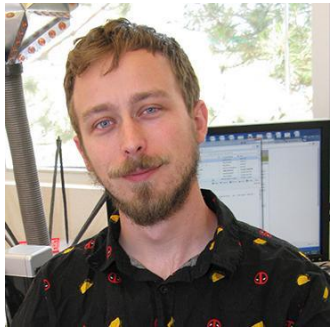


Thiru Ramaraj

Johnny Sena



Melanie Connick



Nico Devitt

Callum Bell

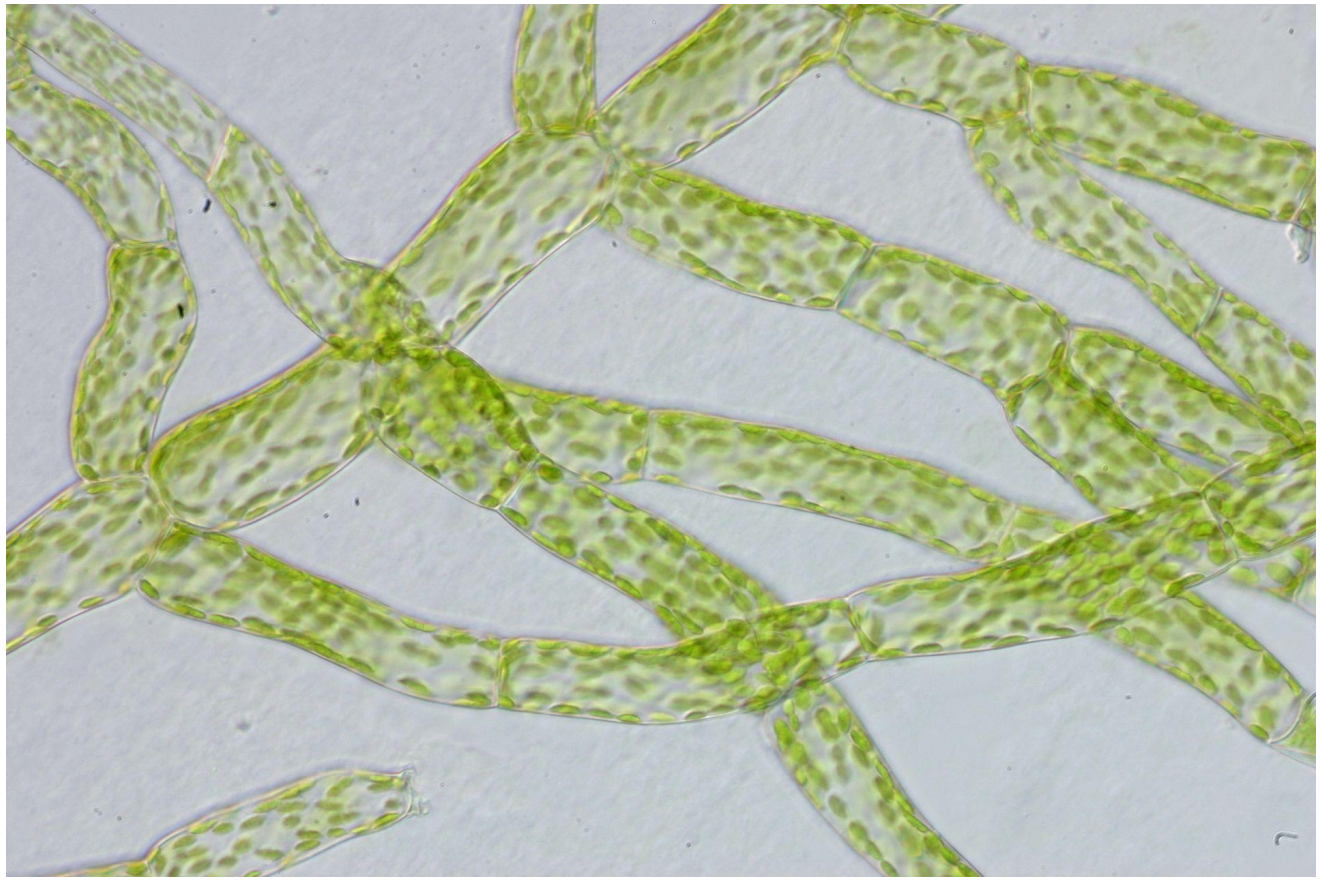
New Mexico INBRE
IDeA Networks of Biomedical Research Excellence

NCGR
National Center for Genome Resources


GORDON AND BETTY
MOORE
FOUNDATION


New Mexico
INBRE
IDeA Networks of Biomedical Research Excellence


NCGR
National Center for Genome Resources

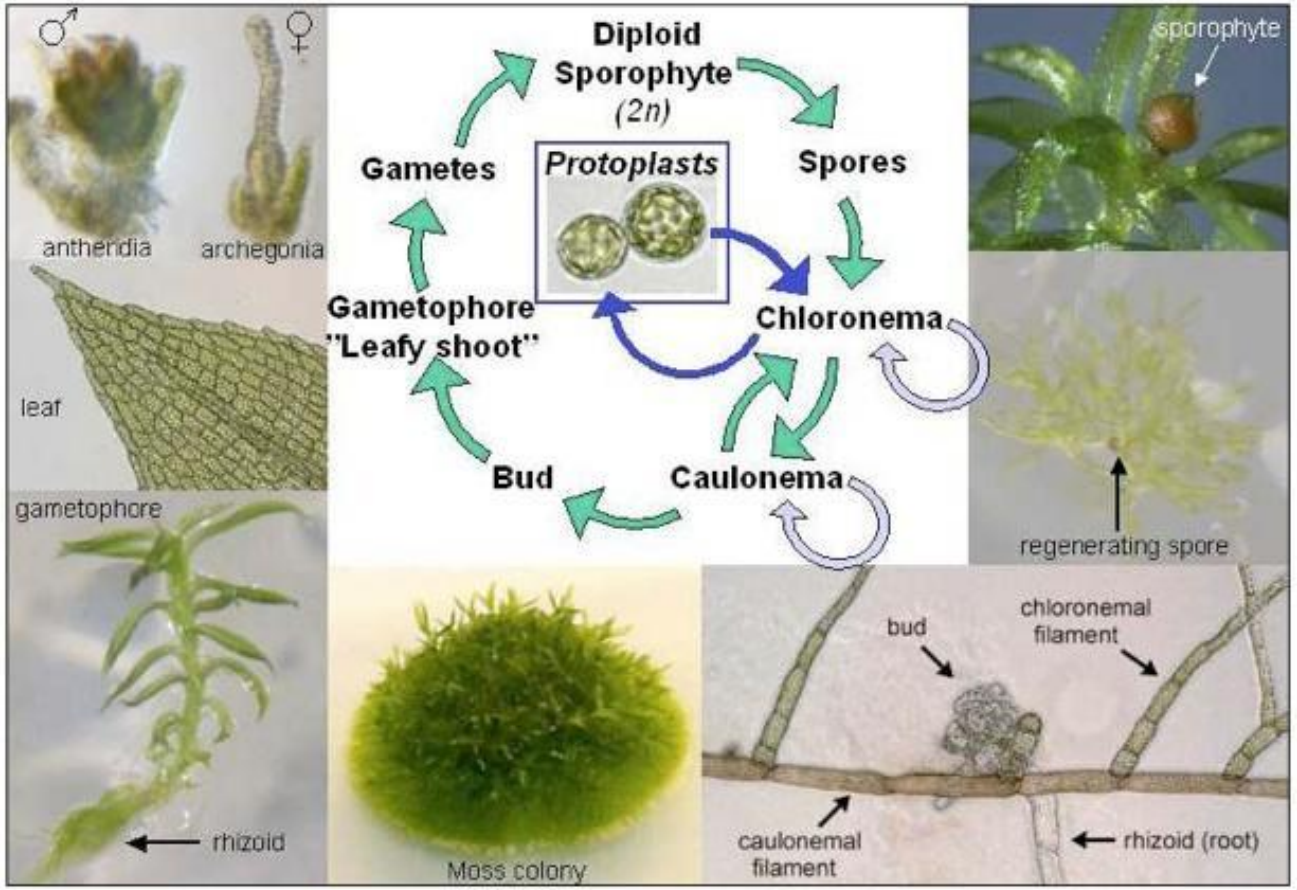


"Physcomitrella Protonema" by Anja Martin, Labor Ralf Reski
(http://en.wikipedia.org/wiki/Ralf_Reski)
Reski Lab, University of Freiburg

New Mexico INBRE
IDeA Networks of Biomedical Research Excellence

NCGR
National Center for Genome Resources

Physcomitrella patens life cycle



The Plant Journal (1999) **20**(2), 245–250

TECHNICAL ADVANCE

A rapid method for detection of plant gene transcripts from single epidermal, mesophyll and companion cells of intact leaves

Stephan Brandt[1], Julia Kehr[1], Christina Walz[1], Astrid Imlau[2], Lothar Willmitzer[1] and Joachim Fisahn[1,*]
[1]Max Planck Institut für molekulare Pflanzenphysiologie, Karl Liebknecht Str. 25, 14476 Golm, Germany, and
[2]Universität Erlangen-Nürnberg, Lehrstuhl Botanik II, Staudtstr. 5, 91058 Erlangen, Germany

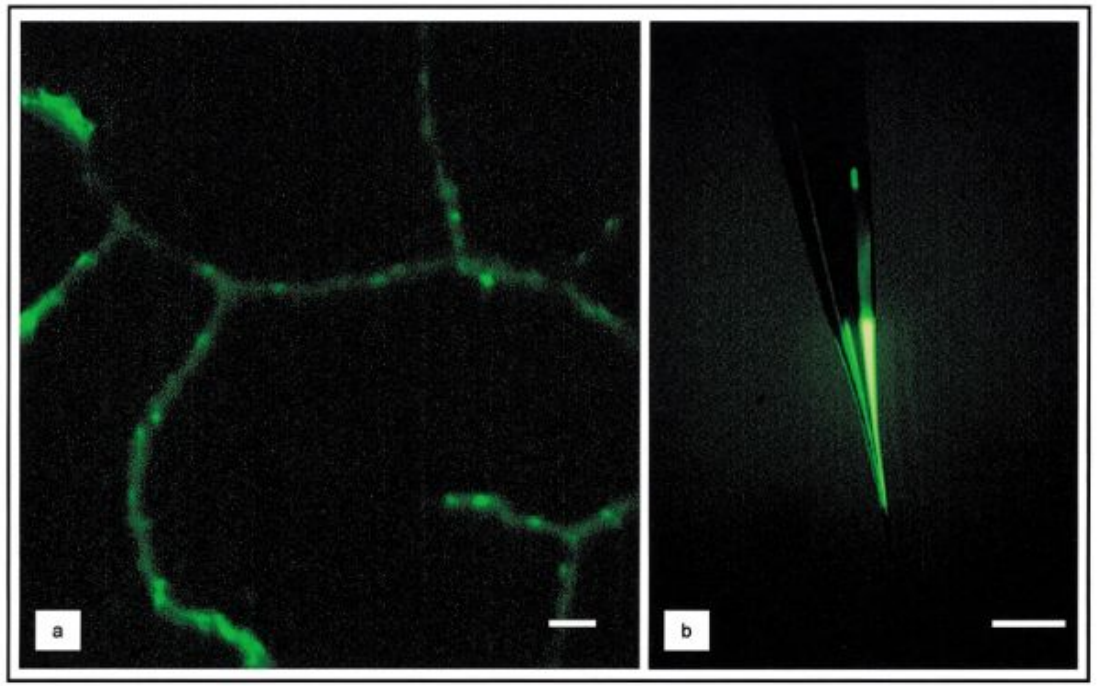


Figure 3. Fluorescence micrographs of (a) top view of a transgenic tobacco leaf expressing GFP under control of a companion cell specific promoter. The GFP fluorescence patterns represent the venation of the leaf, (b) a microcapillary filled with cell extract from a GFP labelled companion cell. Bars indicate 100 µm.

Journal of Plant Physiology 165 (2008) 1530–1544

Available online at www.sciencedirect.com

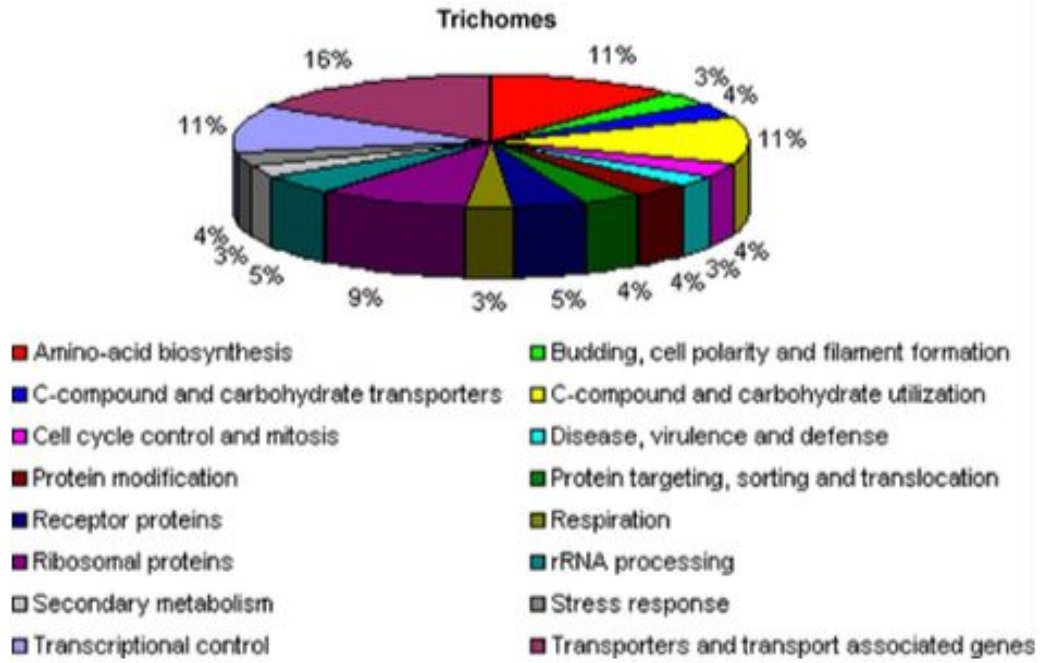
ScienceDirect

JOURNAL OF PLANT PHYSIOLOGY

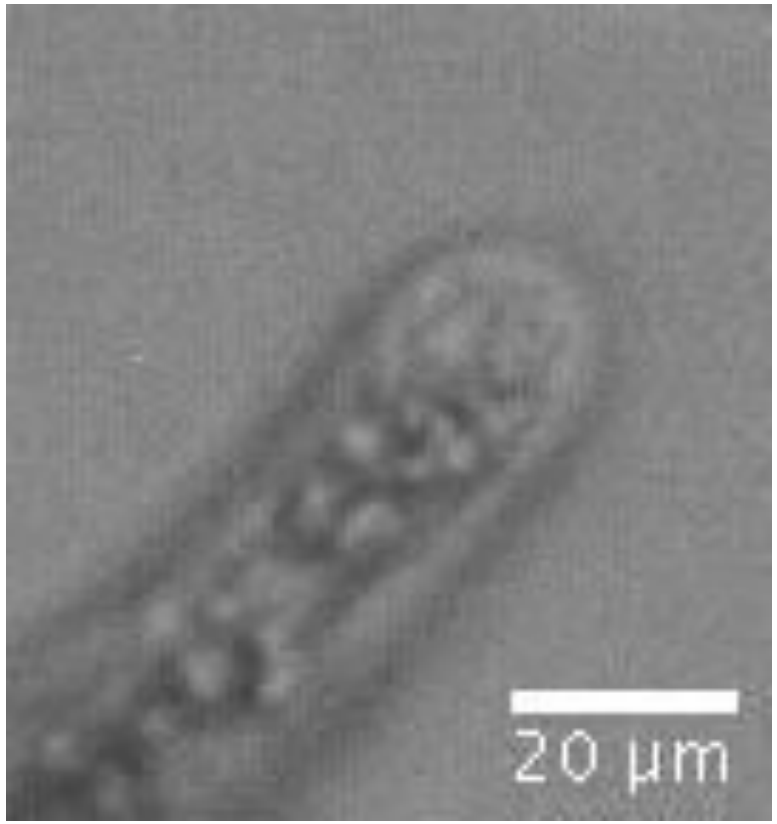
www.elsevier.de/jplph

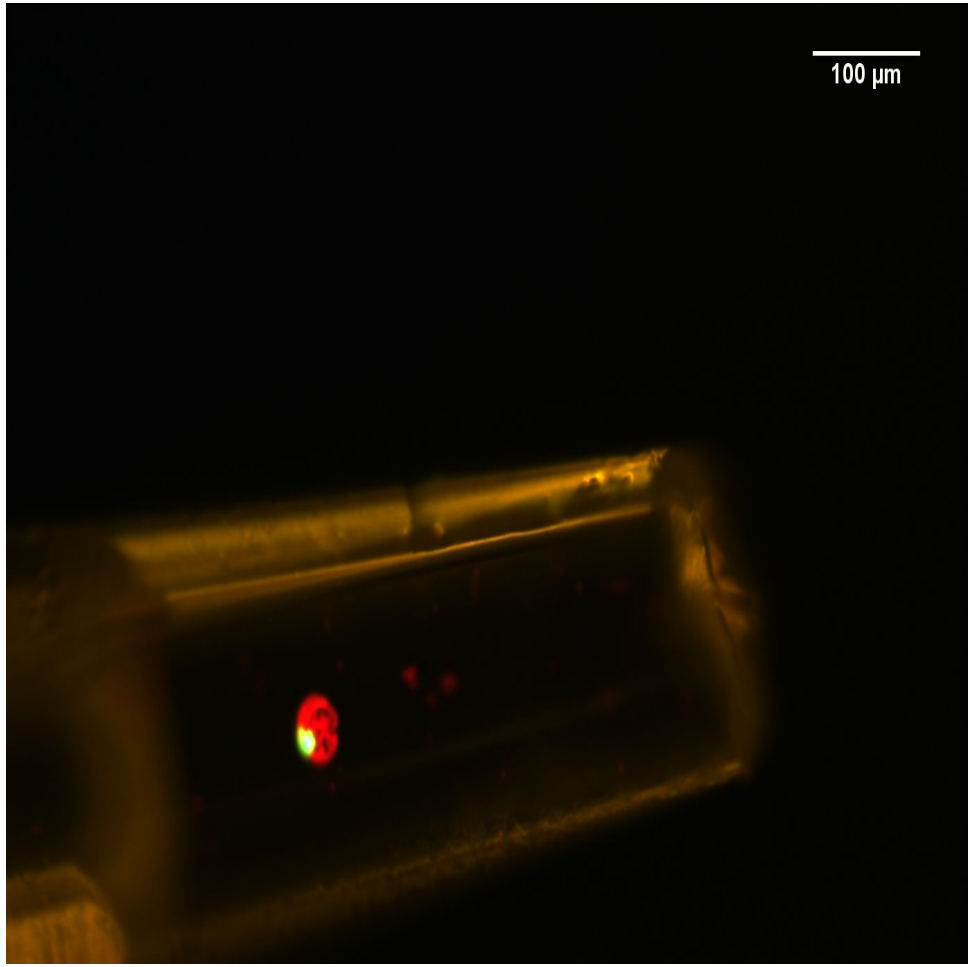
Gene expression profiling of single epidermal, basal and trichome cells of *Arabidopsis thaliana*

Elke Lieckfeldt[1], Ulrike Simon-Rosin[1], Frank Kose, Daniela Zoeller, Martin Schliep, Joachim Fisahn*



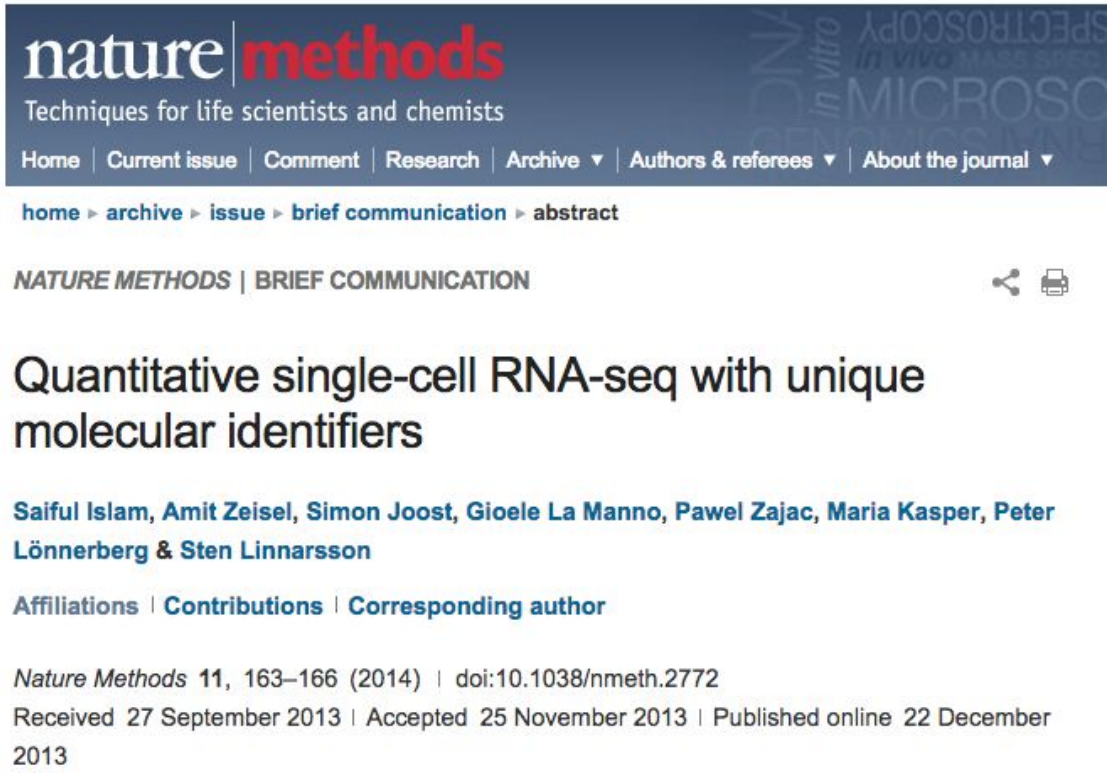
P. Patens protoplast preparation





Protoplast expressing nuclear GFP

Counting mRNA molecules is better than counting reads



nature methods
Techniques for life scientists and chemists
Home | Current issue | Comment | Research | Archive | Authors & referees | About the journal
home ▸ archive ▸ issue ▸ brief communication ▸ abstract
NATURE METHODS | BRIEF COMMUNICATION

Quantitative single-cell RNA-seq with unique molecular identifiers

Saiful Islam, Amit Zeisel, Simon Joost, Gioele La Manno, Pawel Zajac, Maria Kasper, Peter Lönnerberg & Sten Linnarsson

Affiliations | Contributions | Corresponding author

Nature Methods **11**, 163–166 (2014) | doi:10.1038/nmeth.2772
Received 27 September 2013 | Accepted 25 November 2013 | Published online 22 December 2013

NNNNNNNNNNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

mRNA

NNNNNNNNNNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

AATGATACGGCGACCACCGATCGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT

Anneal oligo-dT primer

NNNNNNNNNNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

AATGATACGGCGACCACCGATCGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNNNNNNNNNNCCC

First strand cDNA

TSO-AAUGAUACGGCGACCACCGAGAUCUACACXXXXXXXXXXGGGNNNNNNNNNNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

AATGATACGGCGACCACCGATCGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNNNNNNNNNNCCC

Anneal template switching oligo

TSO-AAUGAUACGGCGACCACCGAGAUCUACACXXXXXXXXXXGGGNNNNNNNNNNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

AATGATACGGCGACCACCGATCGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNNNNNNNNNNCCCXXXXXXXXXXGTGTAGATCTCGGTGGACGCCGTATCATT

Finish first strand cDNA

TSO-AAUGAUACGGCGACCACCGAGAUCUACACXXXXXXXXXXGGGNNNNNNNNNNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
TTACTATGCCGCAGGTGGCTCTAGATGTGXXXXXXXXXXCCCNNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGCTAGCCACCAGCGGCATAGTAA

After 1st strand

GAATGATACGGCGACCACCGA
TSO-AAUGAUACGGCGACCACCGAGAUCUACACXXXXXXXXXXGGGNNNNNNNNNNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
TTACTATGCCGCAGGTGGCTCTAGATGTGXXXXXXXXXXCCCNNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGCTAGCCACCAGCGGCATAGTAA
AGCCACCAGCGGCATAGTAAG

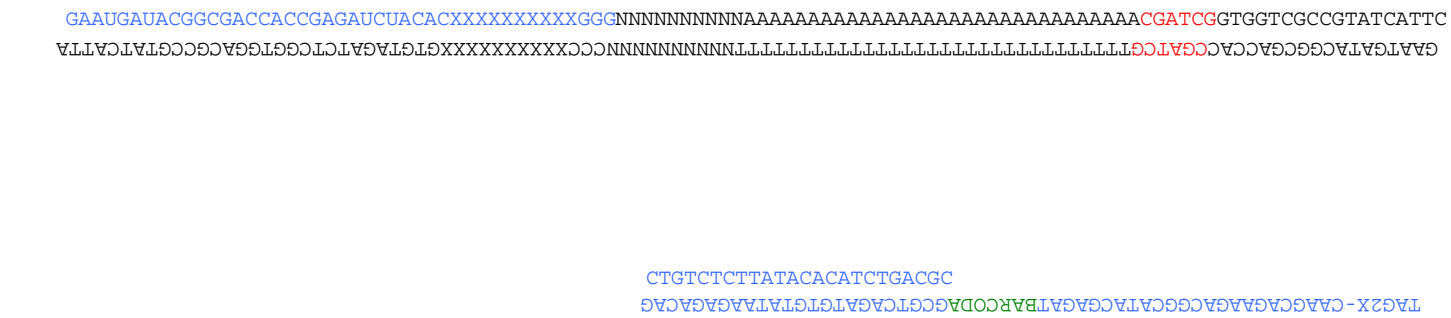
One primer PCR

GAAUGAUACGGCGACCACCGAGAUCUACACXXXXXXXXXXGGGNNNNNNNNNNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACGATCGGTGGTCGCCGTATCATTC
ATTACTATGCCGCAGGTGGCTCTAGATGTGXXXXXXXXXXCCCNNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGCTAGCCACCAGCGGCATAGTAAG

After one primer PCR

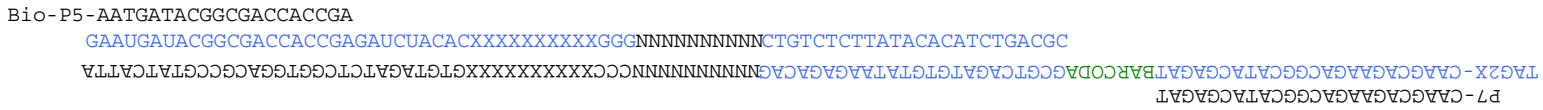
New Mexico INBRE
IDeA Networks of Biomedical Research Excellence

NCGR
National Center for Genome Resources

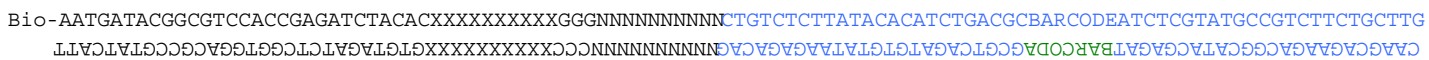


New Mexico INBRE
IDeA Networks of Biomedical Research Excellence

NCGR
National Center for Genome Resources



PCR

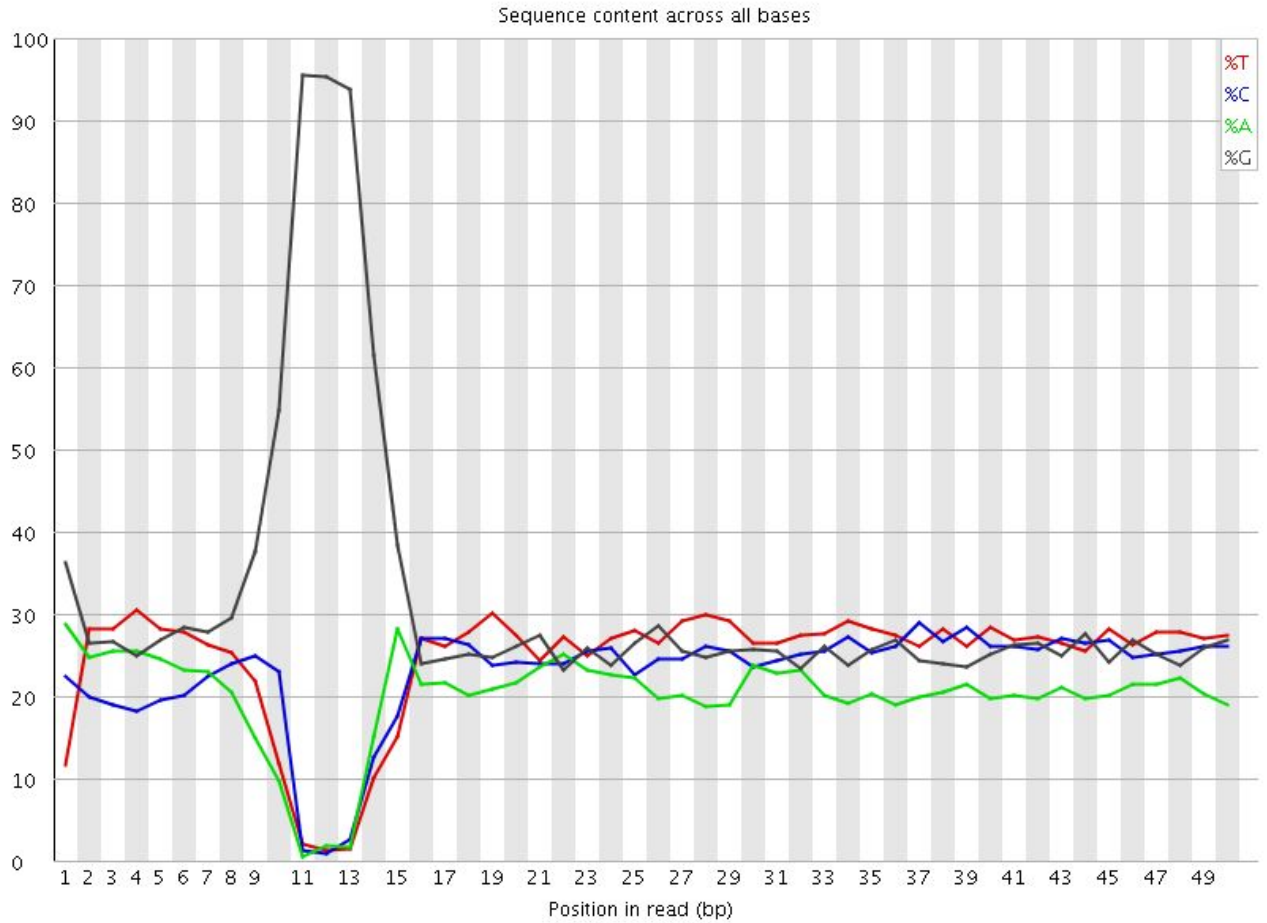


Recover the bottom strand with streptavidin coated beads

New Mexico INBRE
IDeA Networks of Biomedical Research Excellence

NCGR
National Center for Genome Resources

FASTQC of reads passing QC



New Mexico INBRE
IDeA Networks of Biomedical Research Excellence

NCGR
National Center for Genome Resources

```
@HISEQ2:1063:CAWEWACXX:4:1101:1285:1999#TAATTGCGCT
ACACGATGAAACAAGTACTTCCAGATGCCTACAAGG
+
IIJJJJJJJJIIJGIGIJIJJJJJJJJJJJIIJJJ
@HISEQ2:1063:CAWEWACXX:4:1101:3142:1995#GTTACTCATG
NCAACTCGTGGAGGGTGGAAAGCGCGTGCTAGTGCT
+
#3AFHIIIIIIGHIIBGIIIFHIIIIIHGHHFEFFF
@HISEQ2:1063:CAWEWACXX:4:1101:3245:1998#TTGGGATAGG
TCGTCGTCTCTTTGCAAATTCTAAGCTCCCGCTTCT
+
GHIHIGGIJJIJJIGEHHGHIIJGEIIJJJG@FHJJ
@HISEQ2:1063:CAWEWACXX:4:1101:3355:1999#GCAGCACATT
GGATTCCCCACCCGAGCGTGGACAGTGTGGCTGGTG
+
JDDDDDEDDDDDDDDDDD@BDDDDDDDDDDDDDD##
@HISEQ2:1063:CAWEWACXX:4:1101:3720:1996#GATTTCAATG
NCCGCTCATTTGCCATTGCGCGTCTGGTTCACGAGG
+
#2AEGHJJIJJJJICHHHHIJJIJJJJIJGIJJIHH
@HISEQ2:1063:CAWEWACXX:4:1101:3688:2000#TCCCGATCGG
CTCTTATACACATCTGACGCTTACAAGATCTCGTAT
+
JHIIIFHGGEGIIIIIIGFEBFDHGIIIIGGJJIII
@HISEQ2:1063:CAWEWACXX:4:1101:3790:1997#GTGTTTCACG
NGGTGCGCATGGCCCCCGTCGTCCGTGCCTCGGCCA
+
#1?GHIIJJJHJJJJJJJJHHFFFDDDDDDDDDDDD
@HISEQ2:1063:CAWEWACXX:4:1101:3916:1998#GTCATAGGGG
GTATCGAGGAGCTGCGAGCCAGGCGTGGTCGAAGTC
+
D<BCDDDDD8BDDDDDDDDD?BDDDDDDBCBDDDDD
```

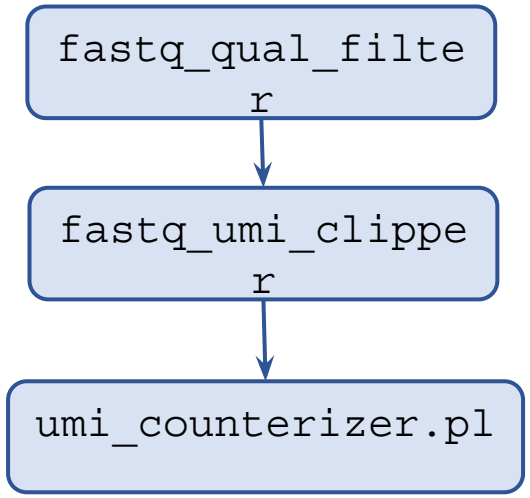
New Mexico INBRE
IDeA Networks of Biomedical Research Excellence

NCGR
National Center for Genome Resources

Alignment to genome

- Physcomitrella patens genome version 3.3
- STAR with no introns:

```
/sw/compbio/STAR/build/STAR-STAR_2.4.0j/bin/Linux_x86_64_static/STAR \
   --runThreadN 8 \
   --readFilesIn /home/projects/adhoc/vulcanite/run_170420/good_clipped.fq \
   --genomeDir /home/projects/adhoc/vulcanite/ref_3.3/idx \
   --alignIntronMax 1 \
   --alignEndsType EndToEnd
```




```
Chr01:219194 1 CAAGGTATAG
Chr01:219196 6 TGATGGAGGG TCAACCGGGT TGGGGAAAGG AAGTAGAAGG TATGTAACAG GCTTCCAGGG
Chr01:219204 1 GTGGGGTCCT
Chr01:219221 1 TCCTTTGGTT
Chr01:219226 2 GGTCAAACCA TGGTCATCCC
Chr01:219243 2 TGGTTTGTGA CAGTTTGTGC
Chr01:219245 3 GGGGCCCACT TCTGTATGTA CGCGGGTCCC
etc.
```

- Should be enough to map UMI counts to annotations
- But...

Something strange is going on with the UMIs

```
AAAAGAAGAG   Chr12:7900304 Chr26:3619711 Chr26:3619712
AAAAGAAGCG   Chr02:3872221 Chr02:3872222 Chr13:4315394
AAAAGAAGGC   Chr20:14859923 Chr20:14859924
AAAAGAAGGG   Chr19:903186 Chr19:903187
AAAAGAATCG   Chr04:81040 Chr15:11239144
AAAAGAATGG   Chr04:7886664 Chr04:7886665 Chr04:7886666 Chr04:20229236 Chr20:9736571
AAAAGACCGG   Chr03:13013503 Chr03:13013504 Chr16:4510628
AAAAGACCTG   Chr05:9674490 Chr05:9674491 Chr12:9536972
AAAAGACGGC   Chr04:81040 Chr08:9268889 Chr08:9268890 Chr09:2418217 Chr13:16920618
AAAAGACGGG   Chr09:12743652 Chr12:13598785 Chr16:3730921 Chr16:3730922 Chr22:3241130
AAAAGACTGG   Chr16:7754118 Chr16:7754119 Chr21:2462127 Chr22:9433984
etc.
```

- Same UMI
- Not quite the same position
- Different read sequence!

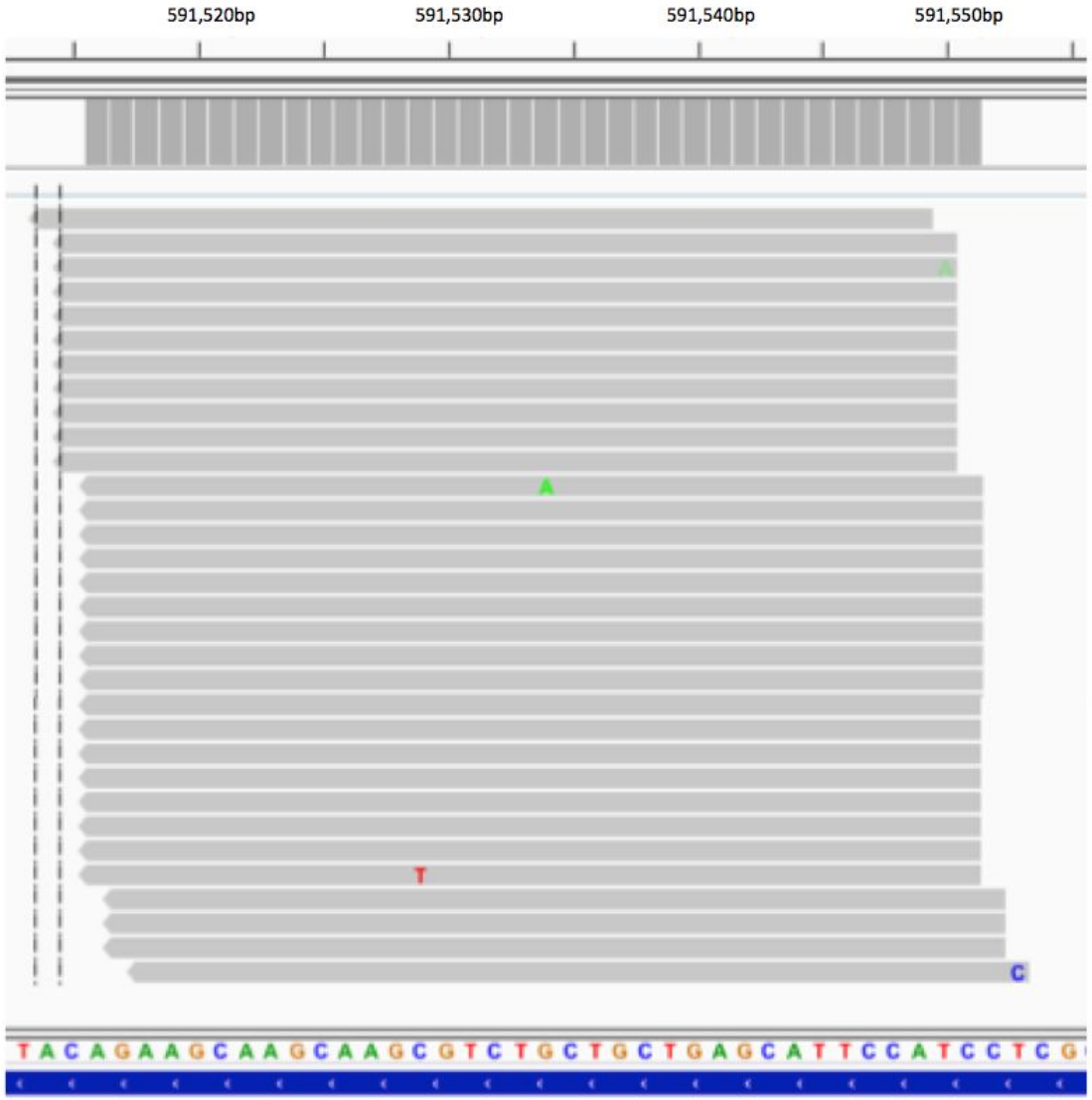


Figure 1. An example of reads sharing a UMI mapping to adjacent locations on the genome. The numbers of reads mapping to the five coordinates shown were 1, 10, 796, 3, and 1, respectively. The 796 alignments at the mode were edited to make it possible to see reads at each position in one figure. All reads in this UMI-read cluster were tagged with UMI AAAGGAGTGG. The central mapping location is Chr03:591516. The vertical dashed line was inserted by the application and indicates the coordinate used to search.

What are the chances?

```
Total UMIs:     78207
Median spacing: 1
Mean spacing:   544420
Mode spacing:   1
Min spacing:    1
Max spacing:    29063882

1   66354
2   1905
3   636
4   177
5   93
6   41
7   15
8   16
9   16
```

- Spacing of identical UMIs
- Identical UMIs are seldom distant
- Identical UMIs are usually next to one another

SCIENTIFIC REPORTS

OPEN

Unique Molecular Identifiers reveal a novel sequencing artefact with implications for RNA-Seq based gene expression analysis

Received: 4 May 2018
Accepted: 7 August 2018
Published online: 03 September 2018

Johnny A. Sena[1], Giulia Galotto[2], Nico P. Devitt[1], Melanie C. Connick[1], Jennifer L. Jacobi[1], Pooja E. Umale[1], Luis Vidali[2] & Callum J. Bell[1]

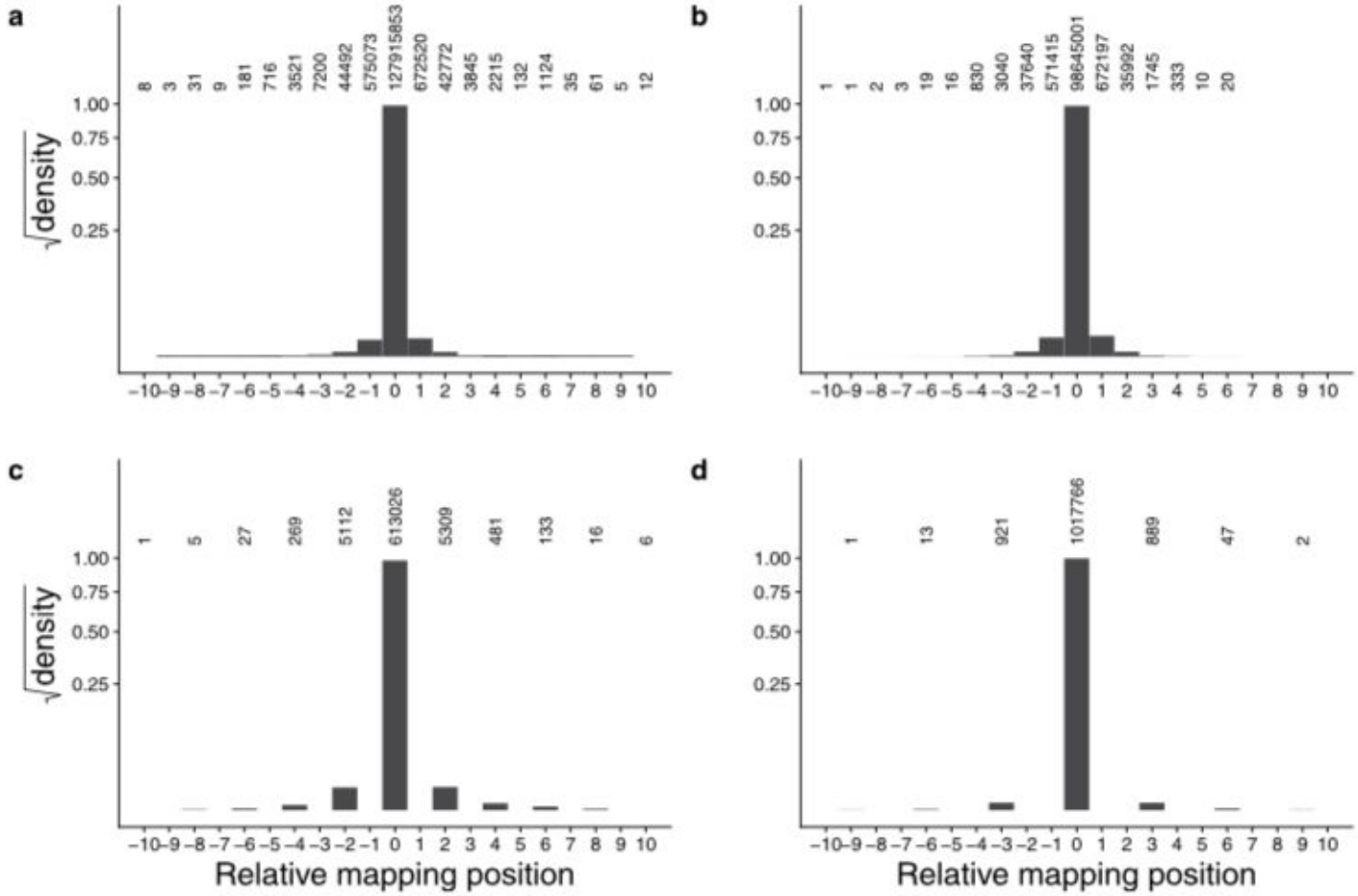


Figure 2. Mapping shifts of reads sharing a UMI in data set run_171108. (**a**) All UMI, including those having no mapping shifts. (**b**) UMI having adjacent mapping shifts of strictly 1 bp. (**c**) UMI having adjacent mapping shifts of strictly 2 bp. (**d**) UMI having adjacent mapping shifts of strictly 3 bp. The Y axis shows the square root of the probability density (summing to 1 for each plot), to make smaller values more visible. The position at which most of the reads map is position zero, with upstream mapping positions taking negative values, and downstream mapping positions taking positive values. Numbers of reads are indicated above each bar. The largest read cluster in this data set consisted of nine adjacent positions.

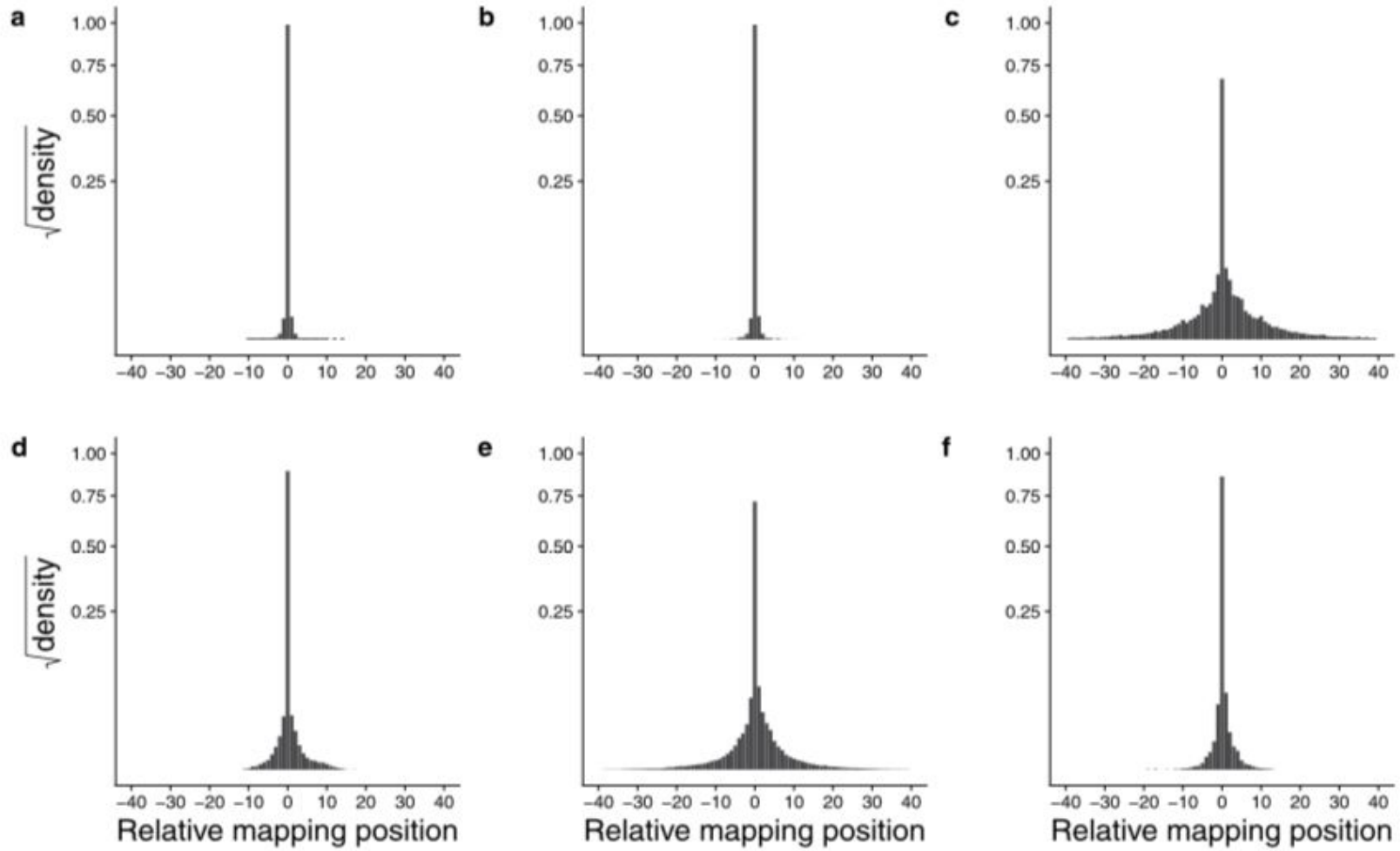


Figure 3. Mapping shifts of reads sharing a UMI in six data sets. Mapping shifts of all sizes are shown. The Y axis shows the square root of the probability density (summing to 1 for each plot), to make smaller values more visible. (**a**) run_171108. (**b**) run_170420. (**c**) SCRB. (**d**) La Manno. (**e**) Yanai1. (**f**) Yanai2. The numbers of reads in each category are shown in supplementary data file UMI Position Read Counts.

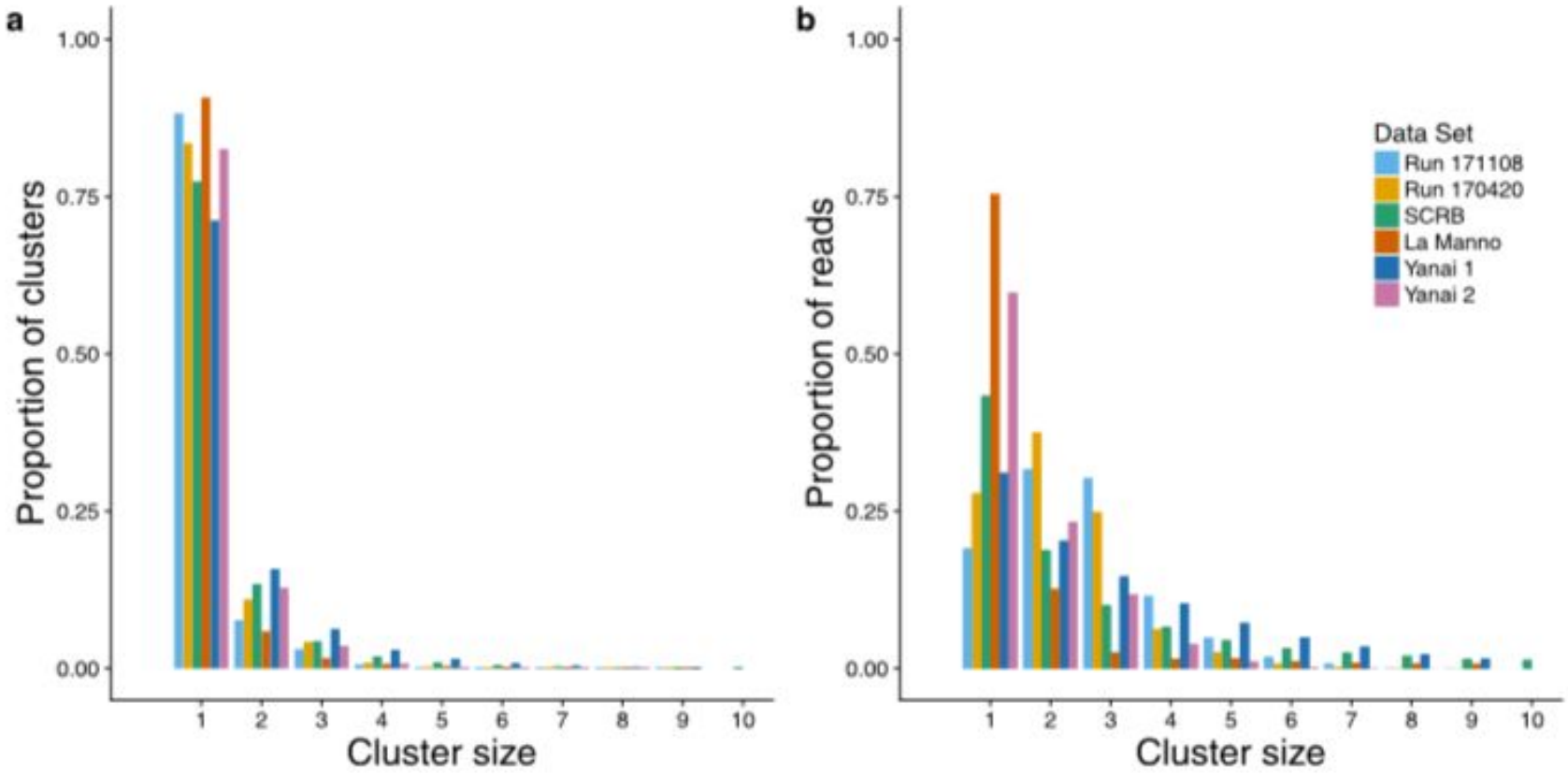


Figure 4. Proportions of clusters of different size and proportions of reads contained in clusters of different size. Cluster sizes range from 1, in which all reads having the same UMI map to the same coordinate, to a case in which reads having the same UMI map to a series 57 adjacent coordinates. The numbers of very large clusters are relatively small. Accordingly, cluster sizes up to 10 adjacent coordinates are shown here. (**a**) Proportions of clusters of different size. (**b**) Proportions of reads found in clusters of different size.

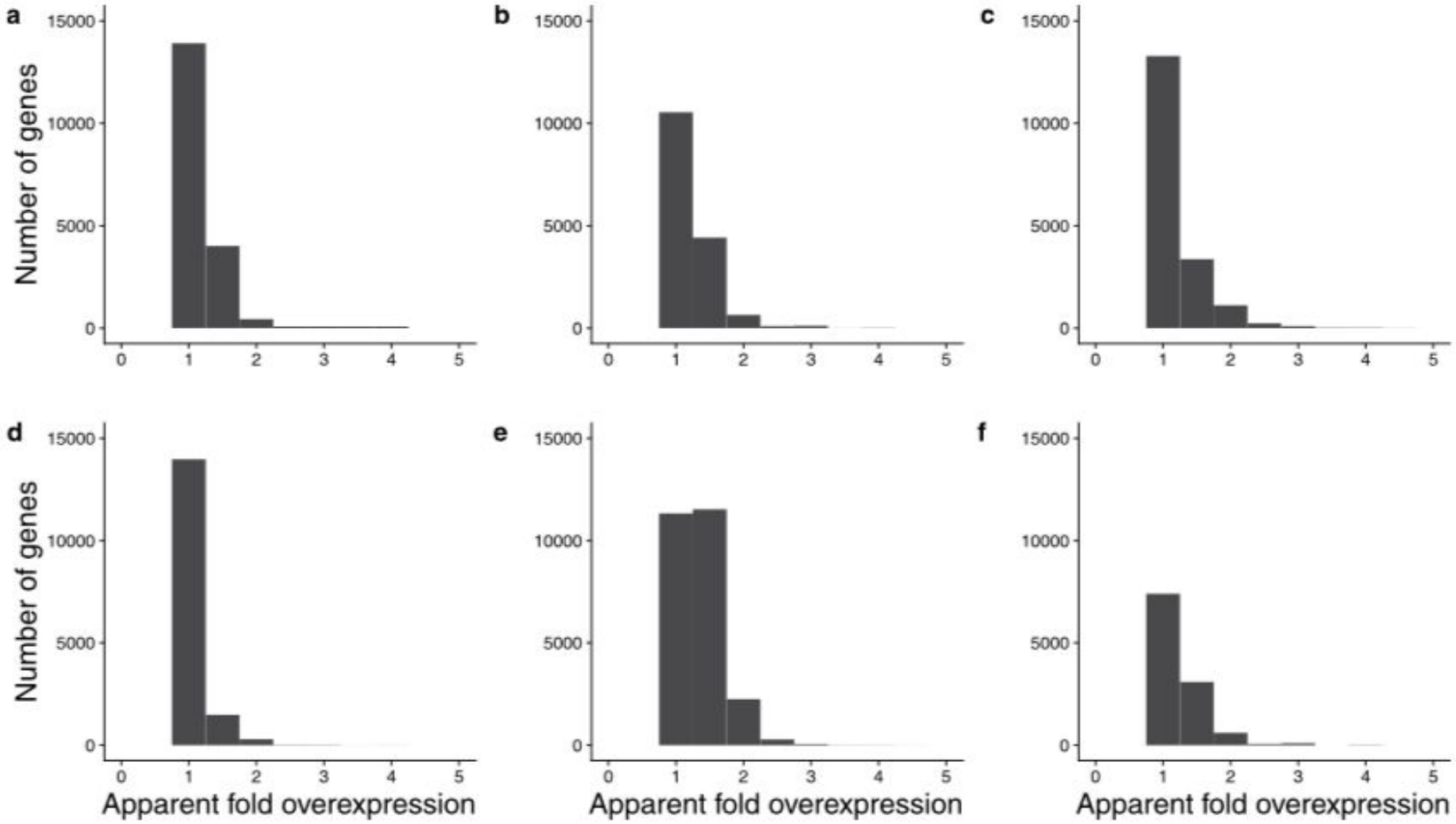
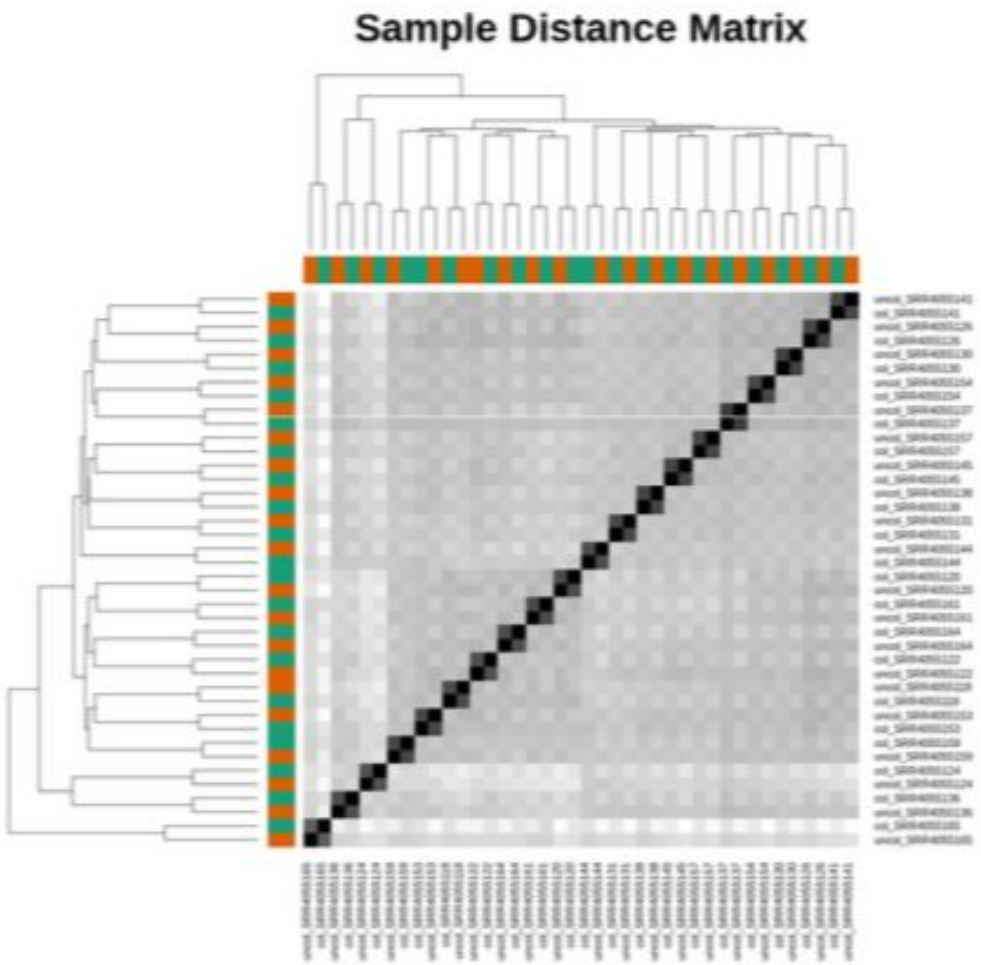


Figure 7. Apparent relative overexpression of genes if mapping shifts are not taken into account. UMI mapping to genes were counted with or without collapsing UMI-read clusters into single observation. The numbers of genes and their ratios of non-collapsed to collapsed UMI-read clusters are plotted. (**a**) run_171108. (**b**) run_170420. (**c**) SCRB. (**d**) La Manno. (**e**) Yanai1. (**f**) Yanai2.



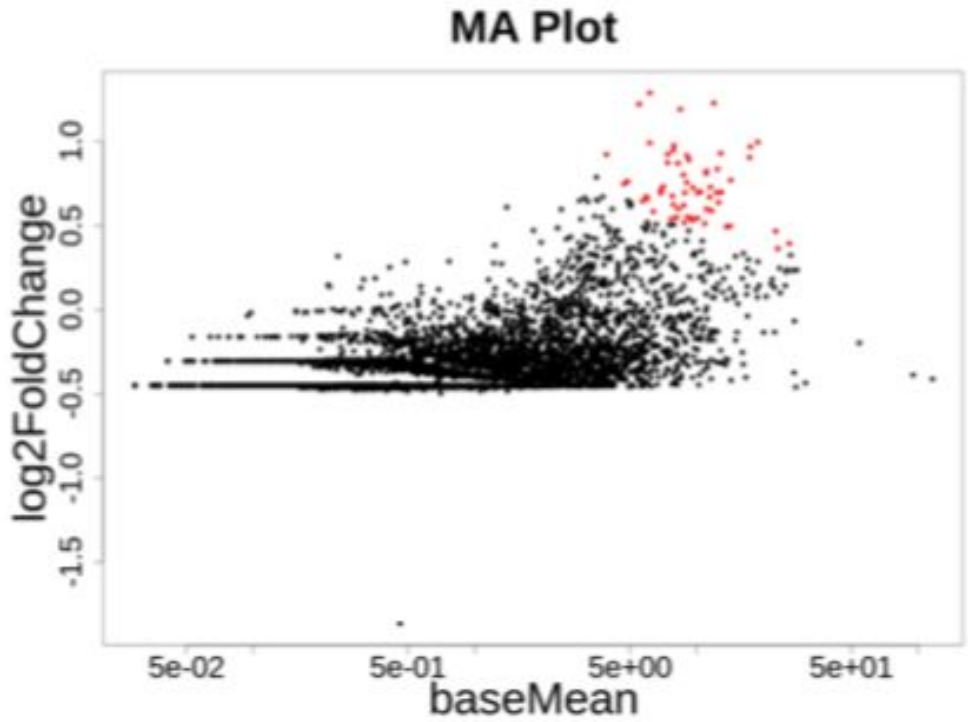


Figure S 4. DESeq2 analysis of 20 mouse RNA-seq biological replicates (single cells) from the La Manno data set in which UMI-read clusters were collapsed (col_) or not collapsed (uncol_) into single observations. Left: pairwise distance matrix. Right: MA plot with log_2 fold *apparent* expression differences on the Y axis and mean expression level on the X. Genes having adjusted P values of less than 0.05 have red dots.

CAAGCGCGAGAGAGCGAGAGAGAGAGAGAGAGAGAG
CTCTCTCTCTCTCTCTCGTAGTAGTACTCTGAATTT
GGACGGGGACAGAGAGAGAGAGAGTGGCATATCCTC
AGAGAGAGAGAGAGAGTGTGTGTGAGTGTGTGTGTT
TGTCTCTGCAGTGTGTGTGTGTGTGAAAATGAAGCT
GGAGAGAGAGAGAGAGAGAGAGGCAACGTCGTGATC
AGAGAGAGAGTGTGTGTGAGTGAGAGATTGAGAGAG
GTGTGTGTATGTGTGAGAGAGAGAGAGAGCAAGAGA
TCTCTCTCTCTCTCTCTCTCAGTTTTGGTTGATGCC
ACTGAGACACCAAGAGAGAGAGAGAGAGAGAGAGCC
CAAGCGCGAGAGAGCGAGAGAGAGAGAGAGAGAGAG
ATCTCTCTCTCTCTCGTAGTAGTACTCTGAATTTGT
TGTGTATCACACACTCTCTCTCTCTCTCTCTCGCTC
TACCACCAGGACTTGCAAACACACACACACACACAC
GAGAGAGAGAGAGAGAGAGAGAGCGAGAACGGAGGG
CTCTCTCTCTCTCGTAGTAGTACTCTGAATTTGTGT
AGGAGGAAAACGGGAGAGAGAGAGAGAGAGAGAGAC

Table S 2. Run_171108 reads found in clusters greater than size 3 with shifts of strictly 2.

Shannon Entropy:

$$H = \sum_{x \in X} p(x) log_2 p(x)$$

Mutual Information

$$I(X, Y) = \sum_{x,y} p(x, y) log_2 (p(x, y)/(p(x)p(y)))$$

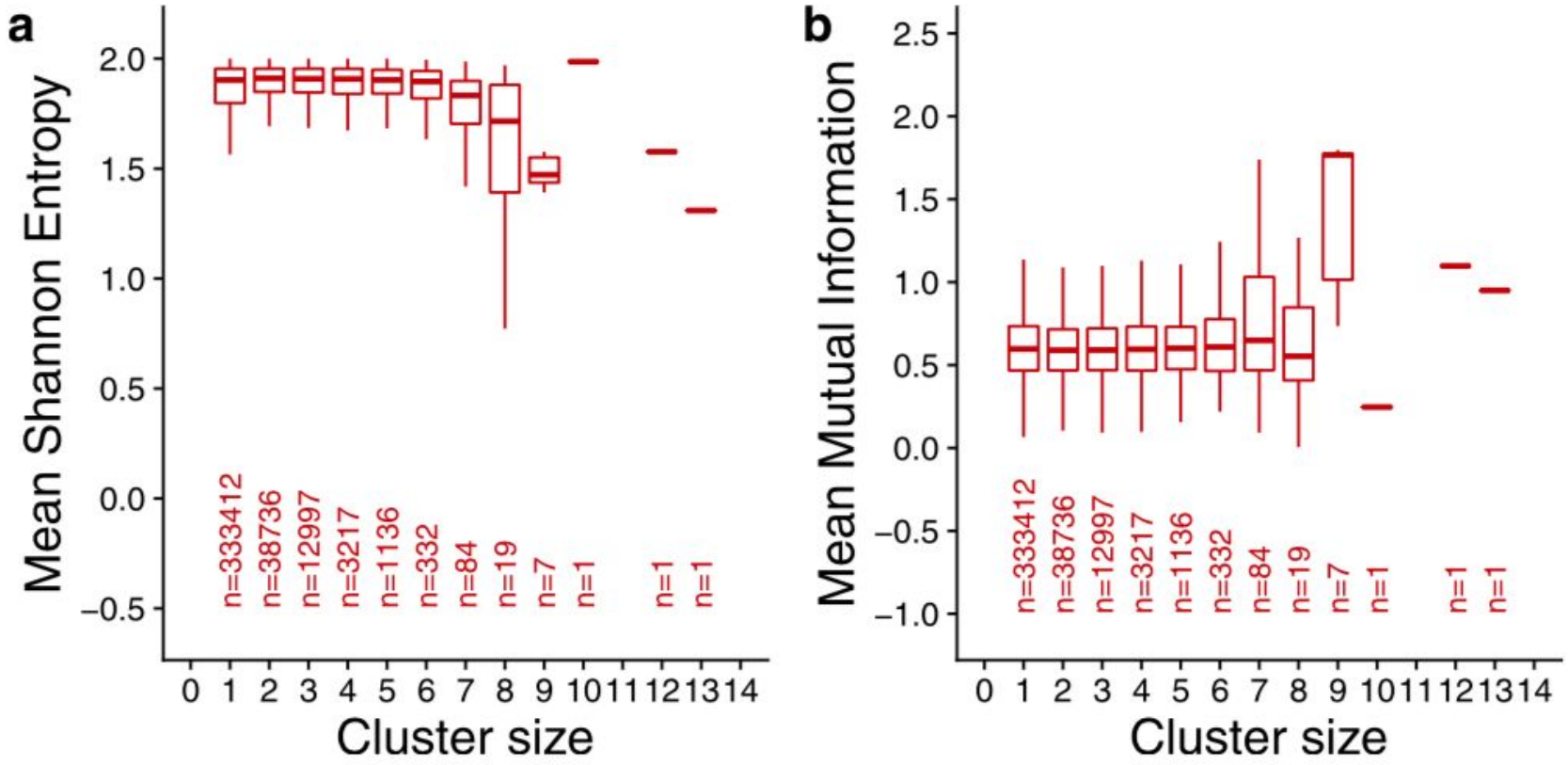
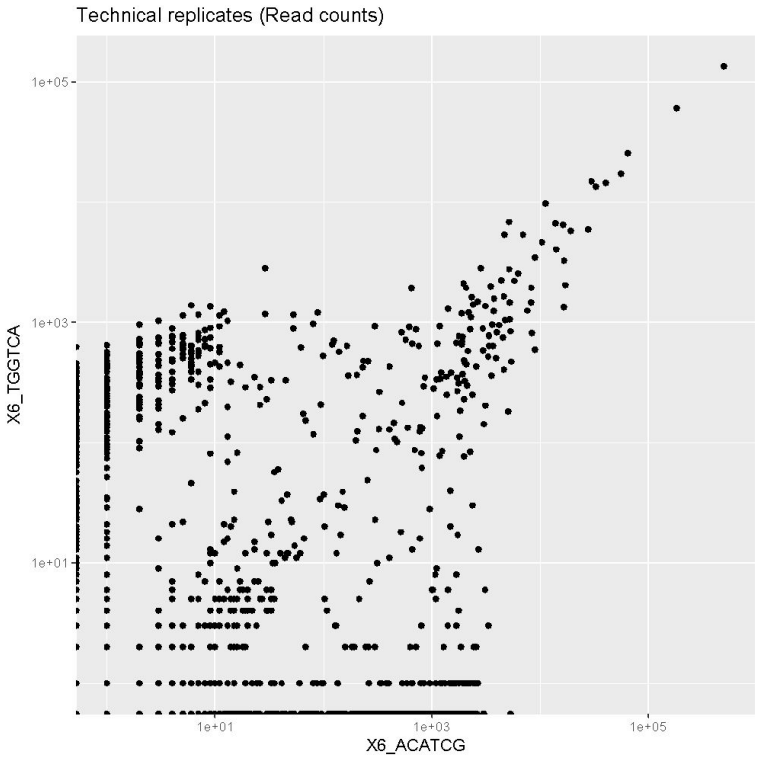
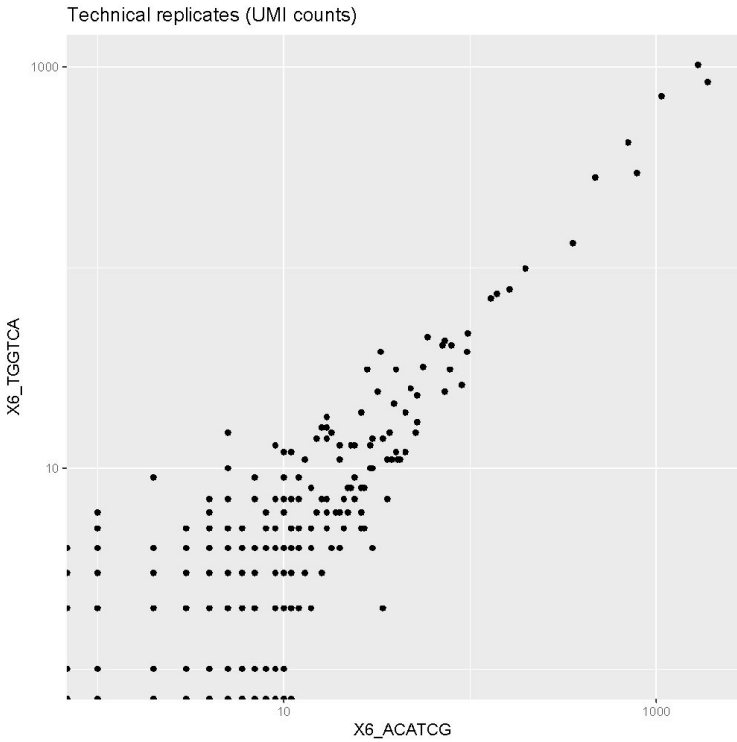


Figure 8. Shannon Entropy (**a**) and Mutual Information (**b**) of run_171108 reads belonging to clusters of increasing size. Each number on the X axis is the number of successive mapping coordinates in which reads share the same UMI. Cluster size = 1 represents reads mapping to only one location. Box plot hinges represent the first and third quartiles. Whiskers extend no more than 1.5 times the inter-quartile range.





The bioinformatics workshop is supported by New Mexico INBRE through an Institutional Development Award (IDeA) from National Institute of General Medical Sciences of the NIH grant
number P20GM103451

New Mexico INBRE
IDeA Networks of Biomedical Research Excellence

NCGR
National Center for Genome Resources