1

TEAL-seq: Targeted Expression Analysis Sequencing

3

4

Georgia Doing, Priya Shanbhag¶, Isaac Bell, Sara Cassidy, Efthymios Motakis, Elizabeth

Aiken, Julia Oh, Mark D. Adams*

The Jackson Laboratory for Genomic Medicine

10 Discovery Dr

Farmington, CT 06032

*Addressee for correspondence at:

Mark.Adams@jax.org

¶current address - TÜV SÜD, Ridlerstraße 65, 80339 Munich, Germany

14

15   **ABSTRACT**

16   Metagenome sequencing enables genetic characterization of complex microbial

17   communities. However, determining the *activity* of isolates within a community presents

18   several challenges including the wide range of organismal and gene expression

19   abundances, the presence of host RNA, and low microbial biomass at many sites. To

20   address these limitations, we developed "targeted expression analysis sequencing" or

21   TEAL-seq, enabling sensitive species-specific analyses of gene expression using highly

22   multiplexed custom probe pools. For proof-of-concept we targeted about 1,700 core and

23   accessory genes of *Staphylococcus aureus* and *S. epidermidis*, two key species of the

24   skin microbiome. Two targeting methods were applied to laboratory cultures and human

25   nasal swab specimens. Both methods showed a high degree of specificity, with >90%

26   reads on target, even in the presence of complex microbial or human background

27   DNA/RNA. Targeting using molecular inversion probes demonstrated excellent

28   correlation in inferred expression levels with bulk RNA-seq. Further, we show that a linear

29   pre-amplification step to increase the amount of nucleic acids for analysis yielded

30   consistent and predictable results when applied to complex samples and enabled profiling

31   of expression from as little as 1 ng of total RNA. TEAL-seq is much less expensive than

32   bulk metatranscriptomic profiling, enables detection across a greater dynamic range, and

33   uses a strategy that is readily configurable for determining the transcriptional status of

34   organisms in any microbial community.

35

36

37

IMPORTANCE

The gene expression patterns of bacteria in microbial communities reflect their activity and interactions with other community members. Measuring gene expression in complex microbiome contexts is challenging, however, due to the large dynamic range of microbial abundances and transcript levels. Here we describe an approach to assessing gene expression for specific species of interest using highly multiplexed pools of targeting probes. We show that an isothermal amplification step can allow profiling of low biomass samples. TEAL-seq should be widely adaptable to the study of microbial activity in natural environments.

INTRODUCTION

As microbiome research matures, studies are shifting from those that establish correlations between microbial profiles and disease states to those that define the mechanisms by which microbes impact host physiology (1, 2). Metagenome shotgun sequencing (mWGS) analysis has been a powerful discovery tool, but it cannot address the functional state of constituent microbes. Measurement of microbial gene expression in a natural context is key to understanding the mechanistic forces driving commensalism and pathogenesis within community and host relationships (3–6).

Changes in diet, drug treatment, introduction of a pathogen, or alteration in host pathways may change the activity of the microbiome, not just its composition (6, 7). Metatranscriptome (metaTx) profiling can uncover which organisms are transcriptionally active, and which genes are expressed. However, the efficacy of metaTx is limited by the

61 large dynamic range of organismal abundance, gene expression levels, and—in some

62 cases—the presence of host transcripts. Thus, quantitation of gene expression in lower

63 abundance organisms by untargeted metaTx sequencing remains infeasible due to the

64 cost of sequencing to sufficient coverage depth. A few metaTx studies have been

65 published that illustrate this point. One study evaluated the contribution of vitamin B12

66 production by skin microbes to the development of acne (8). Another examined

67 transcription of the fecal microbiome of adult men (9). Both studies required ~100 million

68 reads/sample. New, more accessible methods are needed to transcriptionally profile

69 microbes of interest in their natural milieu.

70   The skin is an ideal proving ground for the development of new technologies that

71 measure microbial gene expression in mixed communities. The abundance of bacterial

72 species can vary widely at different skin sites but importantly, abundance is a poor

73 prognosticator of functional importance (10–12). At most skin sites host cells contribute

74 over 90% of the total extracted nucleic acids, making mWGS and metaTx extremely

75 inefficient, particularly for characterizing lower abundance organisms (13).

76 *Staphylococcus (S.)* spp. are keystone contributors to cutaneous immunity, barrier

77 integrity, and microbial community homeostasis, including antagonism with skin pathogen

78 *S. aureus* (14–17). Our previous metaTx work with skin swabs, however, yielded <5% of

79 reads corresponding to microbial transcripts, highlighting the challenges of studying these

80 communities using common methods (unpublished).

81   Targeting microbial profiling to specific genes or genomes enables more cost-

82 effective and comprehensive evaluation of microbial composition and activity. Array-

83 based (18–20) and capture-based (21–24) strategies have been used to assess bacterial

84   and viral abundance, but those methods are either difficult to customize, expensive,

85   laborious, or all three. The availability of low-cost oligonucleotide pools from multiple

86   vendors offers assay development options based on highly multiplexed custom-designed

87   probes, including the ability to iteratively optimize pools based on performance and facile

88   incorporation of new content.

89   We previously developed MA-GenTA (Microbial Abundances from Genome

90   Tagged Analysis) as a quantitative and cost-effective method for species-level microbial

91   profiling (25). We used highly multiplexed (>16,000) oligonucleotide probes to derive

92   relative abundance data for microbes from the mouse gut. MA-GenTA and mWGS data

93   demonstrated excellent correlation down to 0.01% relative abundance (25) and enabled

94   inference of gene pathway composition at a cost only modestly higher than 16S rRNA

95   amplicon sequencing. Here we build on this experience by extending the methods to
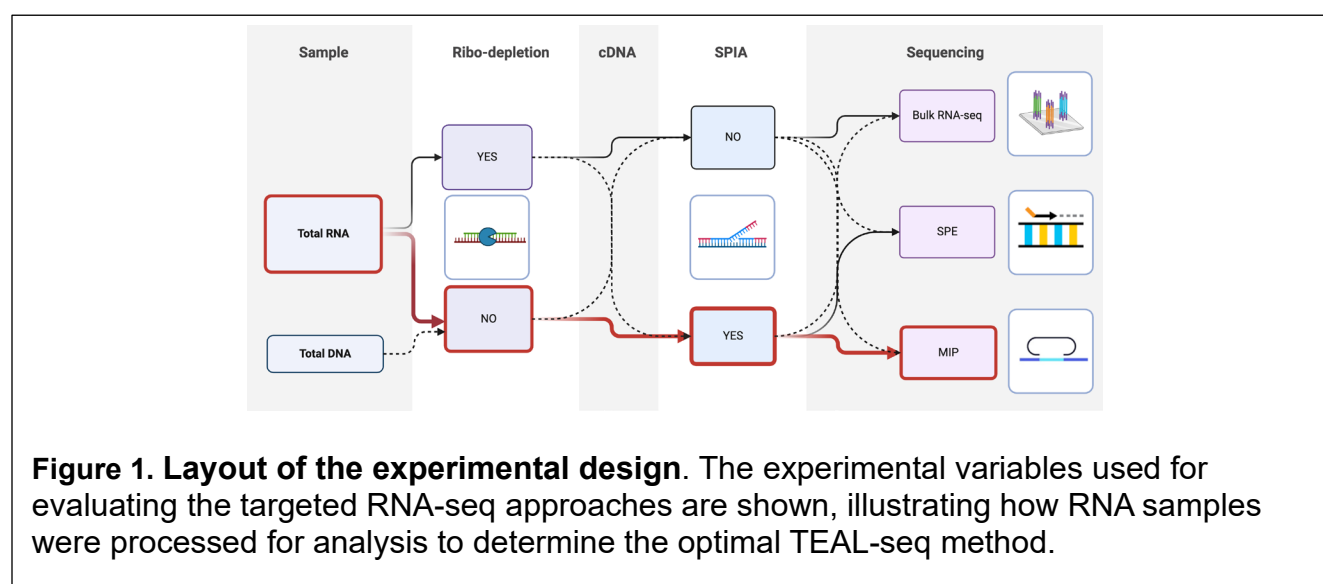
96   targeted gene expression profiling.

97

98

99   **RESULTS**

100   **Experimental design**

101   We selected *Staphylococcus aureus* (*Sa*) and *S. epidermidis* (*Se*) for developing targeted

102   transcriptome profiling methods as they are keystone species in the skin microbiome,

103   important for human health and disease, and well-characterized genomically and

104   functionally (26–28). Two targeting strategies were pursued: a single-primer extension

105   (SPE) method based on MA-GenTA (25) and a molecular inversion probe (MIP) method,

106   previously published for high-plex SNP genotyping assays (29). SPE uses a single 40-

107 base primer that anneals to target sequences to prime DNA synthesis on a partially

108 constructed Illumina sequencing library. MIP involves hybridization of the probe to the

109 target, followed by extension across the gap region, ligation, and amplification of closed

110 circles using primers with Illumina-compatible tails. Both methods target a cDNA or

111 genomic DNA template.

112 We established experimental conditions to evaluate each targeted sequencing approach.

113 Initial tests used RNA and genomic DNA isolated from pure cultures of *Se, Sa*, mixtures

114 of the two, and controls that should not contain target sequences. We compared technical

115 reproducibility, sensitivity to RNA input amount, specificity of targeting in mixed samples,

116 the impact of rRNA depletion, the effect of pre-amplification of cDNA prior to capture, and

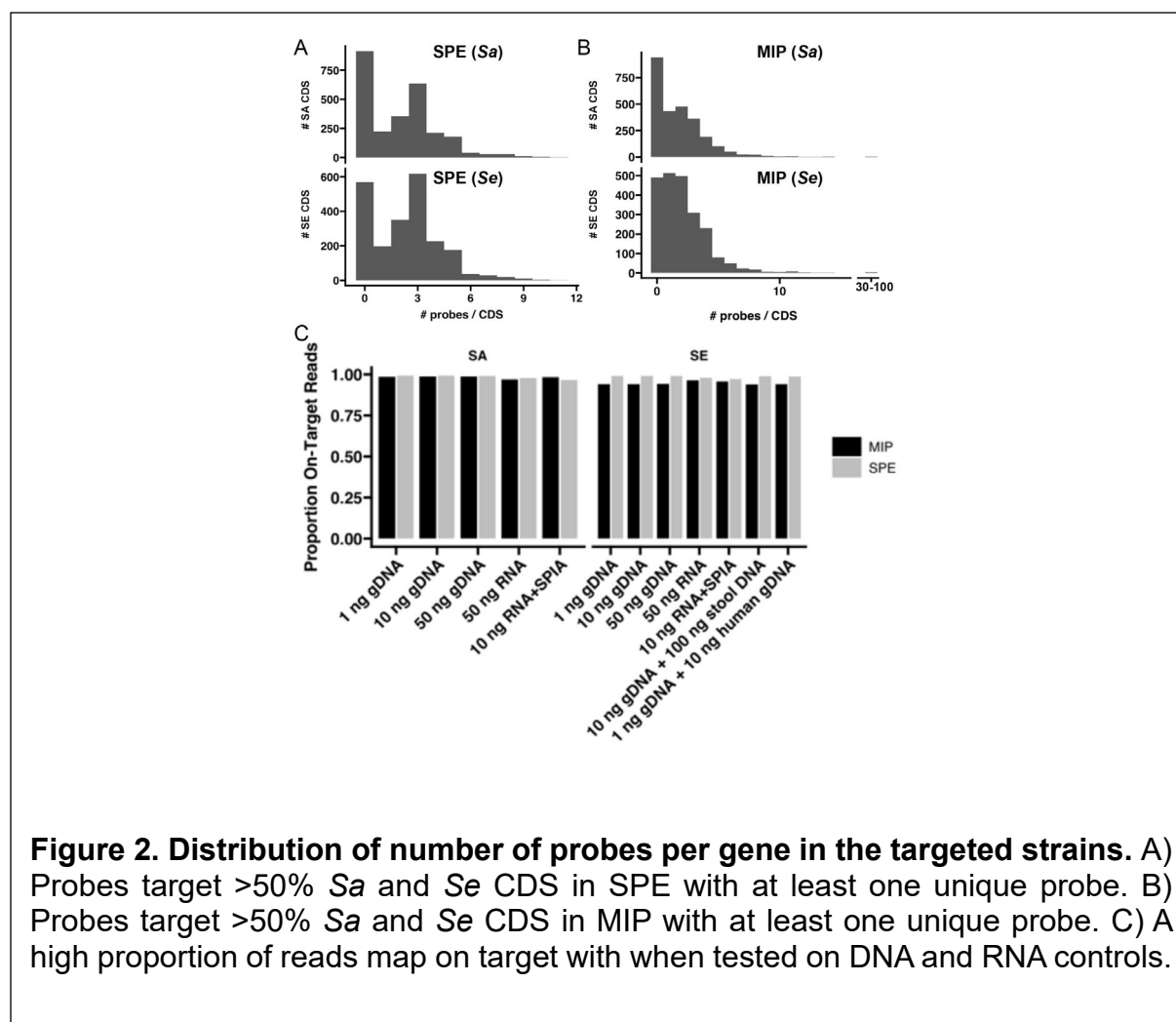117 comparison of targeted RNA-seq methods with standard bulk RNA-seq (**Figure 1**).

Figure 1. Layout of the experimental design. The experimental variables used for evaluating the targeted RNA-seq approaches are shown, illustrating how RNA samples were processed for analysis to determine the optimal TEAL-seq method.

118

Probe pools are >94% on-target and species-specific

120 Probe design for each species maximized within-species binding and limited cross-

121 species binding (see Methods). The SPE design encompassed 6,121 probes targeting

122 1,723 *Sa* coding sequences (CDS) and 5,879 probes targeting 1,670 *Se* CDS (**Figure**

123 **2A**). The MIP design encompassed 4,960 probes targeting 1,694 *Sa* CDS and 4,863

124 probes targeting 1,748 *Se* CDS (**Figure 2B**). A combined pool of *Se* and *Sa* probes was

125 synthesized for each targeting method.

126 Given the low biomass of the microbiome at many body sites, we evaluated the use of

127 single-primer isothermal amplification (SPIA) (30) to increase the amount of template

128 available for targeting and library construction. A theoretical advantage of targeted

129 sequencing is that it makes depletion of rRNA unnecessary; we therefore also tested the

130 effect of rRNA depletion on both bulk and targeted RNA-seq.

131 For both SPE and MIP, 94-99% of reads map to the targeted regions across all

132 experimental conditions, indicating a high efficiency of targeting (**Figure 2C**,

133 Supplemental Table 1). Notably, no ribosomal (r)RNA reads were detected, even in

134 samples processed without rRNA depletion. Single-species experiments showed that

135 probes were highly specific to the targeted genome when full length reads were aligned

136 to the reference genome. In addition, few reads were obtained from *E. coli* RNA, human

137 genomic DNA, and mouse microbiome RNA (cDNA) (data not shown), indicating that

138 cross-hybridization is minimal, even in the presence of complex microbial nucleic acids.

Figure 2. Distribution of number of probes per gene in the targeted strains. A) Probes target >50% *Sa* and *Se* CDS in SPE with at least one unique probe. B) Probes target >50% *Sa* and *Se* CDS in MIP with at least one unique probe. C) A high proportion of reads map on target with when tested on DNA and RNA controls.

Targeted transcriptomics is reproducible

We tested the reproducibility of inferred gene expression levels based on counts per million (CPM) per probe using MIP and SPE by comparing variation across technical replicates to that of bulk RNA-seq. We found technical replicates (libraries prepared from the same RNA sample) highly correlated for all sequencing approaches (**Figure 3A**). Subsequently, we determined that libraries from the same RNA sample with and without SPIA treatment were highly correlated for bulk RNA-seq and MIP, but less so for SPE (**Figure 3B**). Further, when SPIA was used, libraries prepared from standard amounts of

149   RNA ("Standard Input", 10 ng) and an order of magnitude less RNA ("Low Input", 1 ng)

150   were still correlated for both MIP and SPE (**Figure 3C**).

151
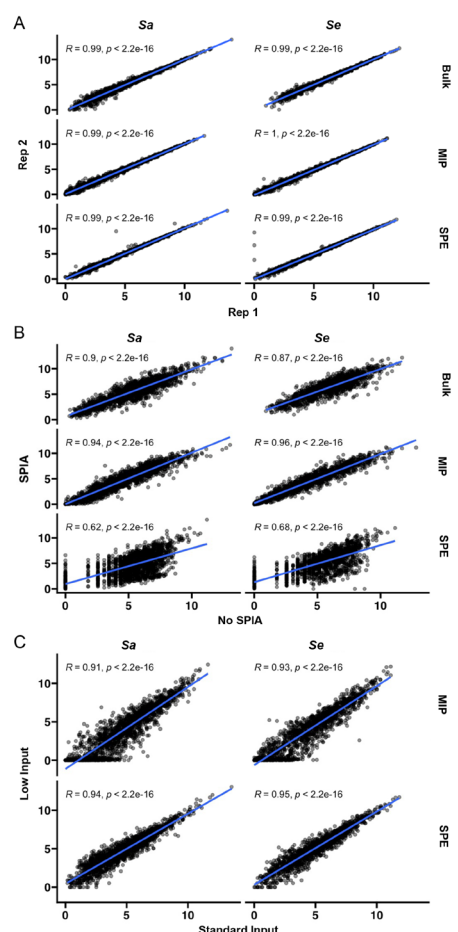
Figure 3. Evaluation of reproducibility using a 50:50 mixture of *Sa* and *Se* RNA. A) Technical replicates (n=2) are highly correlated for bulk RNA-seq, MIP and SPE. B) Samples with and without SPIA have good correlation for bulk RNA-seq and MIP, but less so for SPE. C) Samples with low (1 ng) and standard (10 ng) input total RNA amounts correlate well in MIP and SPE. Correlation coefficients (*R*) and p-values from Pearson tests.
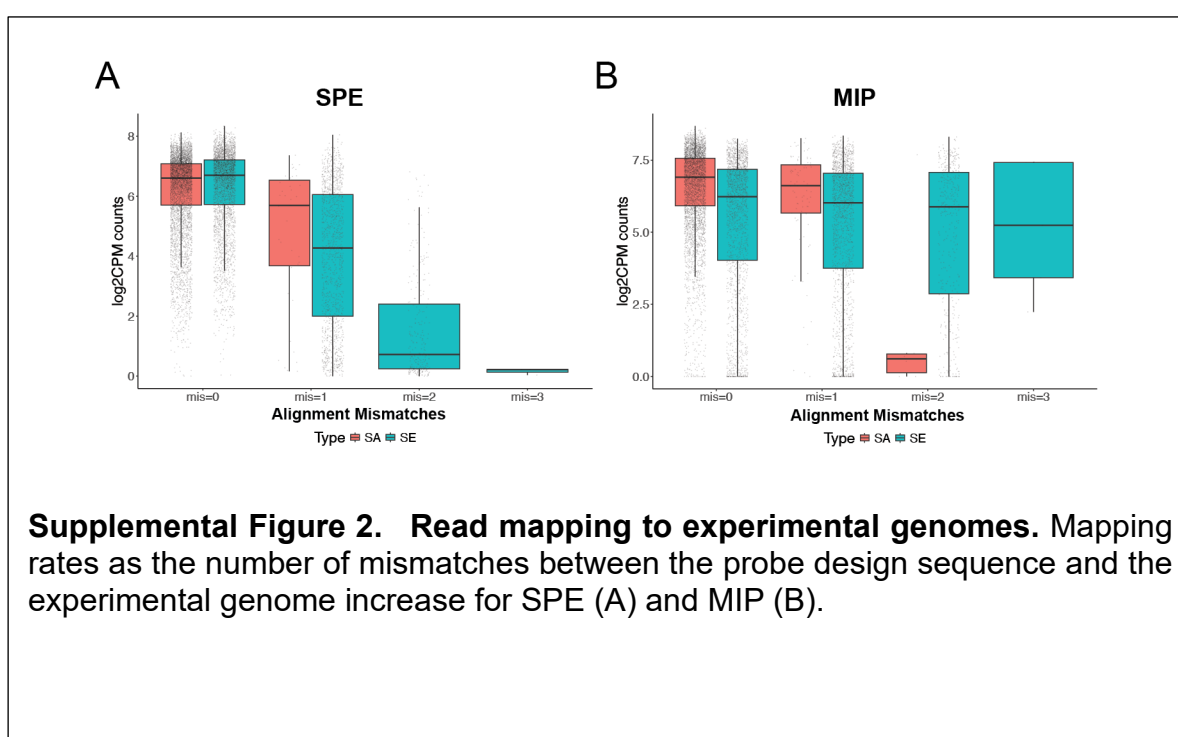
152

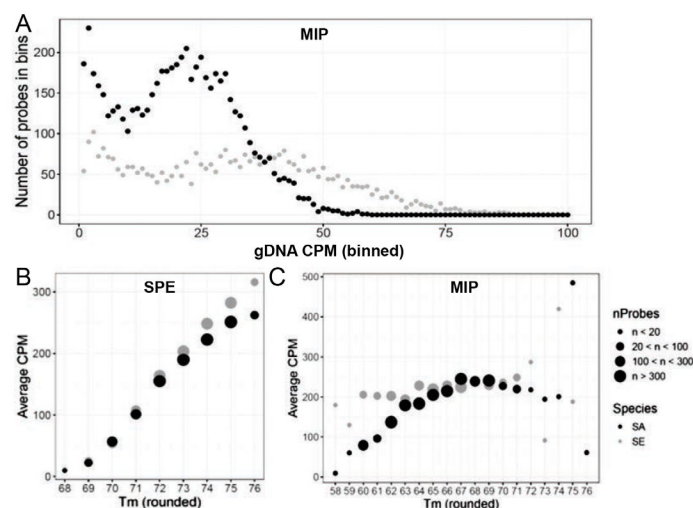153

154

Probe performance varies across CDS

Targeted sequencing of *Se* and *Sa* gDNA was used to evaluate the extent of variation in probe performance. We observed a wide range of CPM/probe, with *Se* probes having a much broader range than *Sa* probes in both the SPE and MIP datasets (**Figure 4A**). Since *Se* and *Sa* probes were synthesized as a single pool, there is no *a priori* reason why they should have such different performance between the two species. We considered whether there is a gradient of genomic copy number ranging from the origin to the terminus of replication due to differences in growth rate (31), however, we found that genome position was poorly correlated with CPM levels in both species (Supplemental Figure 1).
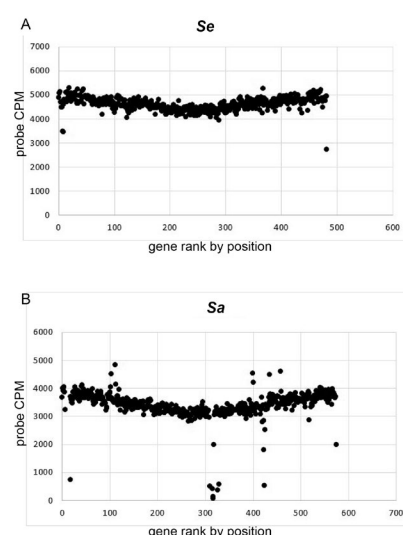
Supplementary Figure 1. CPM bias based on gene position relative to origin of replication. There is no substantial bias introduced by position relative to the origin of replication for *Se* Tu3298 (A) or *Sa* USA300 (B).

165   For SPE probes, there was a marked increase in average CPM with probe melting

166   temperature (T_M) (**Figure 4B**), whereas T_M had minimal correlation with CPM for MIP

167   probes (**Figure 4C**). We also considered the performance of probes that have

168   mismatches to the test genome compared to perfect-match probes. This is an important

169   consideration for profiling of microbial communities, because the exact sequence of the

170   target genomes will not often be known. SPE probes showed a substantial reduction in

171   read counts with increasing numbers of mismatches, whereas MIP showed little or no

172   effect of up to three mismatches (Supplemental Figure 2). Given the limitations of the

173   SPE approach, we focused our remaining analyses on the MIP datasets.

Supplemental Figure 2.   Read mapping to experimental genomes. Mapping rates as the number of mismatches between the probe design sequence and the experimental genome increase for SPE (A) and MIP (B).
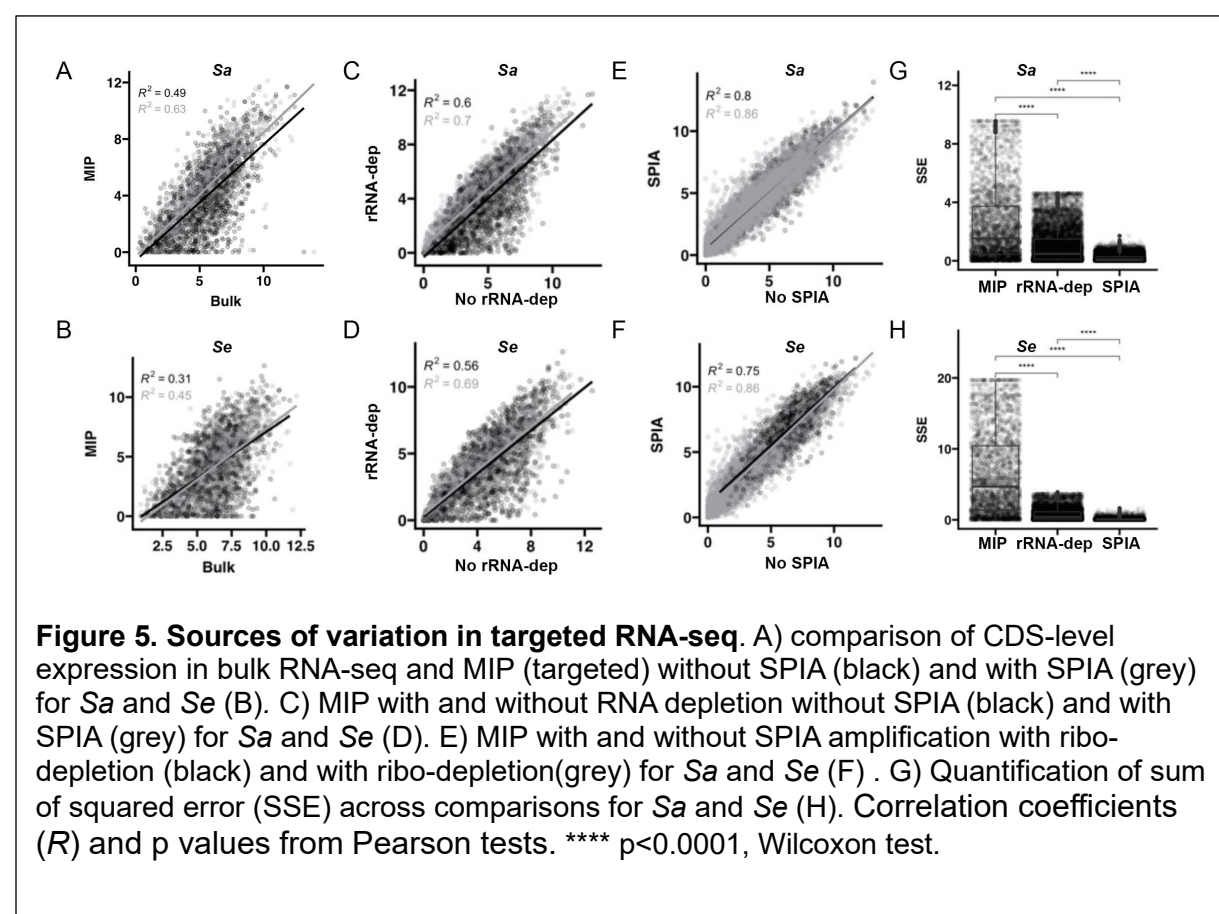
174

175

Figure 4. Probe performance varies and correlates with melting temperature in SPE data more so than MIP data. A) Sequencing of gDNA with SPE shows a range of CPMs for *Sa* (black) and *Se* (grey) probes. B) Probe CPM correlates with melting temperature in SPE data for both *Sa* (black) and *Se* (grey). C) CPM has minimal correlation with melting temperate in MIP data for both *Sa* (black) and *Se* (grey).

Targeting strategies are compatible with signal amplification

The inability to perform oligo-dT priming for cDNA synthesis from microbial transcripts means that rRNA must be depleted as part of library construction for bulk RNA-seq. However, targeting offers the possibility of using total RNA as input, thus reducing processing steps and any potential bias introduced by rRNA depletion. We tested the fidelity of targeting (MIP) with and without ribodepletion and compared the results to bulk RNA-seq. There is only slightly lower correlation in CPM between samples sequenced by bulk RNA-seq and MIP (mean R^2=0.47, **Figure 5A** for *Sa*, **Figure 5B** *Se*) than there are between MIP samples with and without ribo-depletion (mean R^2 = 0.64, **Figure 5C** for *Sa*, **Figure 5D** *Se*), indicating that rRNA depletion had a substantial impact on many non-ribosomal transcripts.

188   To assess SPIA's influence on resulting mRNA measurements, we quantified mRNA

189   abundance as inferred from bulk RNA-seq versus MIP with and without SPIA. Bulk RNA-

190   seq and MIP samples with and without SPIA correlated highly (mean R^2 = 0.83, **Figure**

191   **5E** for *Sa*, **Figure 5F** *Se*), more highly than samples with and without rRNA depletion.

192   Further, CDS-level CPM values across bulk and targeted RNA-seq were less variable

193   when SPIA amplification was used (**Figure 5A**, **B** grey) than when SPIA was not used

194   (**Figure 5A**, **B** black). In summary, since SPIA amplification introduces significantly less

195   sum squared error (SSE) than ribo-depletion does (p<0.0001 by Wilcoxen test) for both

196   species (**Figure 5G and 5H**), we determined MIP with SPIA to be the optimal workflow

197   for TEAL-seq (as outlined in **Figure 1**).

198

Figure 5. Sources of variation in targeted RNA-seq. A) comparison of CDS-level expression in bulk RNA-seq and MIP (targeted) without SPIA (black) and with SPIA (grey) for *Sa* and *Se* (B). C) MIP with and without RNA depletion without SPIA (black) and with SPIA (grey) for *Sa* and *Se* (D). E) MIP with and without SPIA amplification with ribo-depletion (black) and with ribo-depletion(grey) for *Sa* and *Se* (F) . G) Quantification of sum of squared error (SSE) across comparisons for *Sa* and *Se* (H). Correlation coefficients (*R*) and p values from Pearson tests. **** p<0.0001, Wilcoxon test.

199

TEAL-seq captures gene expression changes across growth conditions
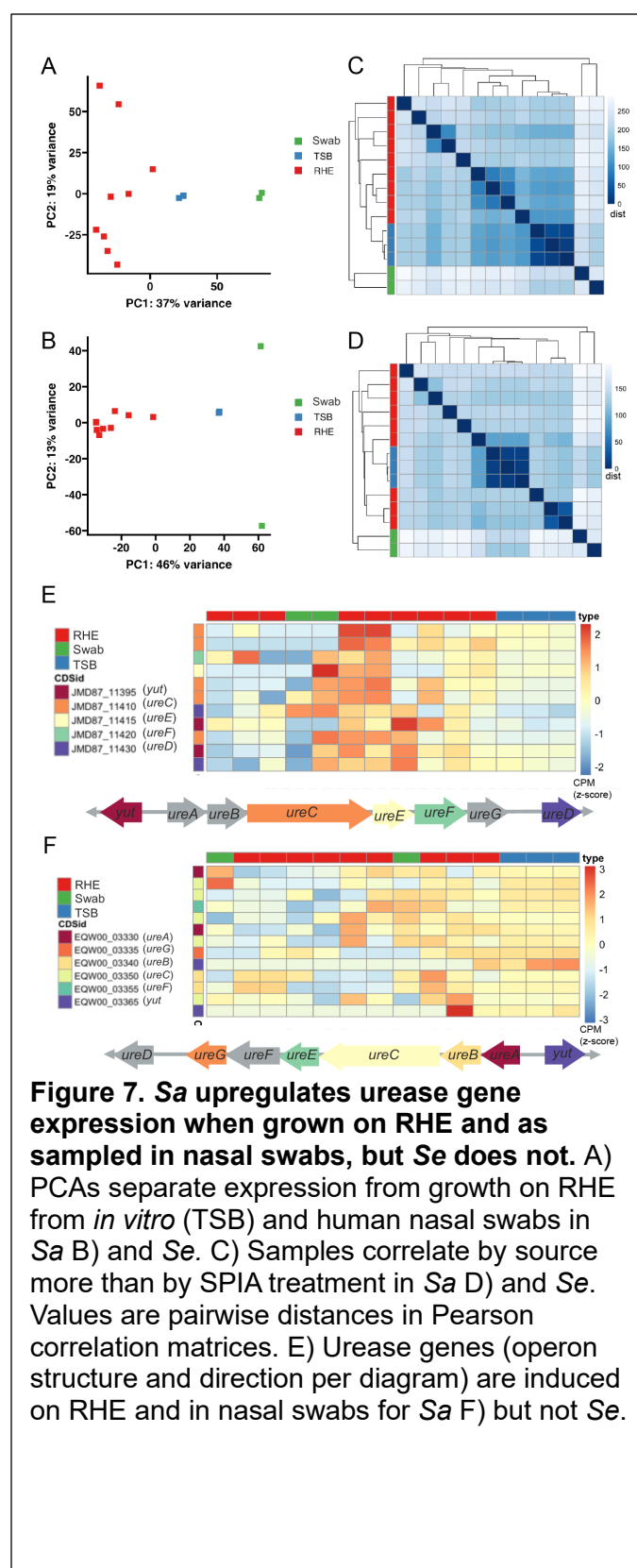
201   A major goal of RNA-seq analysis is to identify differentially expressed genes

202   (DEGs) across conditions of interest to infer underlying biology and nominate targets for

203   downstream functional studies. To test the ability of TEAL-seq to capture DEGs under

204   physiologically relevant conditions, we measured gene expression from *Sa* and *Se*

205   cultures grown in TSB (pH 7), acidic TSB (pH 4.8), and TSB with urea (4.5%w/v) to model

206   acid stress commonly encountered by staphylococci *in vivo*. Growth conditions were

207   delineated by the first and second PCs which together explained 97% of variance for *Sa*

208   (**Figure 6A**) and 94% of variance for *Se* (**Figure 6B**). Further, genes with the highest fold-

209   change in response to acid were consistent with known acid responses, in particular

210   urease genes in *Sa* (**Figure 6C**) and genes previously identified as upregulated by acid

211   in *Se* (**Figure 6D**). Clustering of *Sa* samples shows that the expression patterns are driven

212   first by condition (**Figure 6E**). For *Se*, condition also influences expression of these genes

213   of interest, although urea treatment had less of an effect than acid (**Figure 6F**). Our results

214   suggest that TEAL-seq performs predictably for the analysis of differential gene

215   expression and thus is a valuable approach for comparative microbial studies.
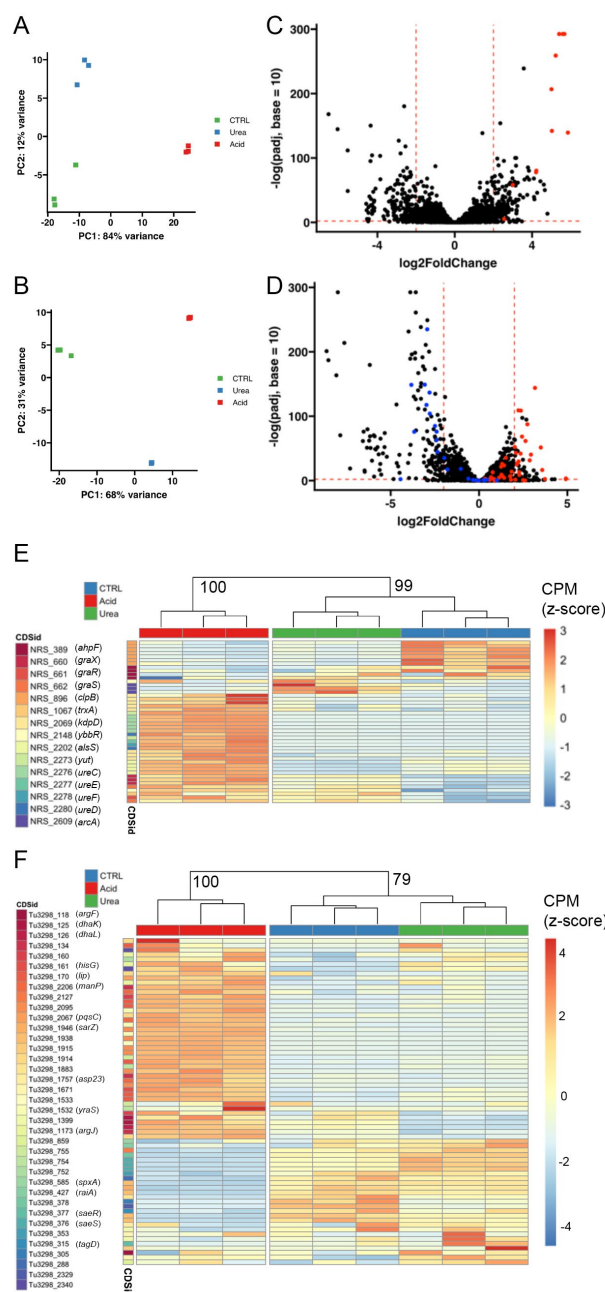
Figure 6. TEAL-seq detects urease and acid response genes as differentially expressed under acid stress. A) Principal components of targeted gene expression separate samples by growth condition for *Sa* and *Se* (B). C) Differential expression between control (TSB) and acid treatment identifies urease genes for *Sa (*red) and the previously published acid response for *Se* (red and blue) (D). E) Urease and acid response probes across conditions are consistent across CDSs for *Sa* and *Se* (F). Each column represents biological replicates (n=3).

DEG analysis from model and human samples

To model the technical challenge of low amounts of material extractable from skin microbiomes, we applied TEAL-seq to RNA extracted from reconstructed human epidermis (RHE) colonized with clinical skin microbiome isolates as well as RNA extracted directly from human nasal swabs. *Sa* and *Se* transcripts were readily detected from both samples. Transcript levels derived from these sources were substantially different from those grown in TSB (PC1, separation by sample type accounted for 37% of variance for *Sa* (**Figure 7A**) and 46% of variance for *Se* (**Figure 7B**). Notably, samples from independent RHE and swabs were more variable than replicates of growth in TSB, as expected for complex models and *ex vivo* samples. Further, there was high correlation in expression profiles for samples grown in TSB or between sequencing

Figure 7. *Sa* upregulates urease gene expression when grown on RHE and as sampled in nasal swabs, but *Se* does not. A) PCAs separate expression from growth on RHE from *in vitro* (TSB) and human nasal swabs in *Sa* B) and *Se.* C) Samples correlate by source more than by SPIA treatment in *Sa* D) and *Se.* Values are pairwise distances in Pearson correlation matrices. E) Urease genes (operon structure and direction per diagram) are induced on RHE and in nasal swabs for *Sa* F) but not *Se.*

239 replicates of the same sample for both *Sa* (**Figure 7C**) and *Se* (**Figure 7D**). Thus, this

240 variation is likely biological.

241 We used TEAL-seq to compare expression patterns of urease genes from bacteria

242 grown in TSB, on RHE, or isolated from swab samples. A subset of both nasal swabs and

243 RHE samples showed marked increase in the expression of urease genes relative to TSB

244 for *Sa* (**Figure 7E**). Interestingly, another subset of swab and RHE samples had lower

245 expression of urease genes compared to TSB for *Sa*. *Se* samples from RHE and swabs

246 also displayed variable urease gene expression compared to those grown in TSB,

247 although the trend was less pronounced than in *Sa* samples (**Figure 7F**). To further

248 validate TEAL-seq as a faithful measure of mRNA in complex samples, we attempted to

249 compare the expression of urease genes by TEAL-seq to that detectable by bulk RNA

250 sequencing on the same RHE sample. Notably, not all samples had detectable microbial

251 mRNA using bulk RNA sequencing; colonized RHE samples sequenced to depths of 8.5

252 – 42.1 million (m=30.4 million) contained only 0.06 – 1.5% (m=0.3%) reads that mapped

253 to *Sa* or *Se* mRNA and thus failed to provide sufficient reads to measure genome-wide

254 expression (Supplemental Table 3). These results demonstrate that TEAL-seq can be

255 used to measure gene expression of *in vivo* relevant pathways in complex samples in a

256 manner more sensitive than the current state-of-the-art method.

257

258 **Discussion**

259

260 We present an approach for targeted gene expression analysis optimized for samples

261 with low biomass—TEAL-seq—that yields a high proportion of reads on target, even in

262 the presence of total RNA input and complex background RNA or DNA. We tested two

263 oligo-based methods for targeting, MIP and SPE, and determined MIP to be more robust

264 than SPE. Both methods were capable of inference of mRNA levels relative to bulk RNA-

265 seq, but expression levels inferred from MIP had less probe-to-probe variation. Further,

266 we showed that cDNA amplification using SPIA reduced variation between replicates and

267 had less impact on inferred expression level than rRNA depletion. Therefore, we propose

268 the use of MIP with SPIA for measuring bacterial transcript levels in low biomass samples.

269

270 We showed that probes designed based on genes of laboratory strains captured gene

271 expression information from other common experimental strains and clinical isolates of

272 the same species. This is in part due to short probe length, ensuring coverage of

273 conserved genes across strains, as well as the design, which enabled integration of data

274 across multiple probes per CDS. Cross-strain capture is an important feature of this

275 method as the application of TEAL-seq to measure gene expression of real-world

276 samples will require probes to be designed from existing genomic or metagenomic data.

277 Our benchmarking of results from probes designed for *Sa* strain M2872 and *Se* strain

278 ATCC14990 and applied to *Sa* strain USA300 and *Se* strain Tu3298 as well as clinical

279 isolates of both species demonstrate the flexibility of TEAL-seq.

280

281 We used TEAL-seq to assess changes in gene expression of *Sa* and *Se* grown in TSB at

282 pH 7 and pH 4.8 and identified highly concordant changes in expression compared to

283 previous studies. We confirmed that *Sa* responds to acid stress with urease genes and

284 *Se* does not. This is significant as urea is a human skin-produced metabolite that has

285   important roles in barrier defense (32), and differences in *Sa* and *Se* responsiveness to

286   acidic conditions may explain differences in invasiveness (33).

287

288   MetaTx has had limited usage in studying low biomass microbiomes. However, TEAL-

289   seq successfully captured predictable expression profiles from multiple complex sample

290   types, including those isolated clinically. New tools are needed for identification of

291   population-level expression of virulence pathways or immunogenic metabolites at species

292   resolution, as these will contextualize the function of the microbiome as a collective and

293   refine our understanding of the contribution of individual species to the community (34–

294   37). Additional studies applying TEAL-seq in the context of complex natural microbial

295   communities will be needed to further demonstrate the potential of this method in

296   experimentally relevant systems. However, the robust performance of TEAL-seq in

297   capturing biologically significant transcriptional patterns from skin microbes grown under

298   skin-relevant stress and reconstructed human epidermis make it a promising tool for

299   future applications.

300

301   **Methods**

302   **Bacterial samples and RNA and DNA isolation.** Single colonies of *Staphylococcus*

303   *aureus* USA300 and *Staphylococcus epidermidis* Tu3298 were inoculated into 2ml Tryptic

304   Soy broth (TSB) (BD cat:211825) and grown overnight at 37°C and 200 rpm agitation.

305   Overnight cultures were sub-cultured into fresh 2 ml TSB medium at OD_{600} ~0.05 and

306   allowed to grow to OD_{600} ~0.5-0.6. For stress conditions, as in (38), overnight cultures

307   were grown as described above and then sub-cultured into mildly acidic TSB (pH 4.8) or

308 TSB+ 4.5% w/v urea and allowed to grow to OD_{600} ~0.5-0.6. For RNA extraction, to 1 ml

309 of each mid-log phase culture, 2 ml of RNAprotect® Bacteria Reagent (cat # 76506,

310 Qiagen Inc.) was added, and cells were centrifuged at 8000 x g for 5 minutes at 4°C. Cell

311 pellets were then resuspended in RLT buffer (RNeasy Mini kit, (cat # 74104, Qiagen Inc.)

312 and transferred to 2 ml safe-lock tubes containing 50 mg acid-washed glass beads (150–

313 600 μm diameter). Cells were homogenized using a TissueLyser for 4 minutes at

314 30Hz/sec and further processed according to the kit protocol. For DNA extraction, cells

315 from 1 ml of culture were pelleted by centrifugation at 16000 x g for 2 minutes and DNA

316 extracted using GenElute™ Bacterial Genomic DNA Kit (cat# NA2110, Sigma-Aldrich).

317 For mixed cultures, overnight cultures of individual strains were sub-cultured as described

318 above and allowed to grow to an OD_{600} ~ 0.5-0.6 in TSB at 37°C with 200 rpm agitation.

319 Equal numbers of cells from each strain were then inoculated into fresh 5 ml TSB at an

320 OD_{600} ~0.1-0.2 and grown at 37°C with agitation at 200rpm for 2-3 hours until the OD_{600}

321 reached ~0.5.

322 Mouse microbiome DNA was prepared as previously described (39). RNA was extracted

323 from bacteria grown in association with a reconstructed human epidermis (RHE) skin

324 organoid model (MatTek EpiDerm) with the Qiagen RNeasy kit (cat # 74104, Qiagen Inc.)

325 and *Escherichia coli* RNA was purchased from ThermoFisher (#AM7940).

326 **Reconstructed human epidermis cell culture cultivation.** 9mm primary normal human

327 RHE tissue cultures were obtained from MatTek Corporation (EpiDerm, Gothenburg,

328 Sweden). All batches were grown using cells from MatTek Corporation's EpiDerm

329 standard donor, a healthy male. The RHE were cultured according to the manufacturer's

330 directions. Briefly, upon arrival, the RHE cultures were placed in 6-well plates with 1mL of

331 warmed antibiotic-free EpiDerm Maintenance Media (MatTek Corporation,) or the

332 equivalent EpiDerm Assay Media basally per well. The basal media was replaced daily,

333 and the RHE cultures were kept at 37°C with 5% CO_2.

334 For each microbial isolate, a single colony was grown overnight in sterile 1X TSB. 10^8

335 colony-forming units from each liquid culture were washed with ultrapure water (Fisher

336 Scientific, #AAJ71786AP, Hampton, NH) and resuspended to a final concentration of 10^7

337 colony-forming units in 120 µL.

338 The RHE cultures were then dosed with 120uL of microbial isolate or vehicle (ultrapure

339 water alone). Dosed RHE cultures were incubated for 1 hour at 37°C, then inoculum was

340 aspirated to restore the air liquid interface. RHE were incubated with remaining bacteria

341 for 18 hours prior to harvest.

342 At harvest, 200 µL of PBS (MatTek Corporation) was added to the apical surface of each

343 RHE culture, pipette mixed, removed, and plated for CFU enumeration. 140 µL of RLT

344 buffer + 1% beta-mercaptoethanol was added to each RHE culture for RNA preservation.

345 The RLT buffer-tissue solution was frozen at -80°C until RNA extraction. *S. epidermidis*

346 Tü3298-GFP colonized RHE were visualized under blue light.

347

348 **Target gene selection and probe design.** Pangenome datasets for *Sa* and *Se* were

349 assembled using Roary (40) based on the available complete genome sequences of each

350 species: 51 complete *Sa* and 86 complete *Se* genomes downloaded from GenBank. CDS

351 present in at least 50/86 *Se* and 30/51 *Sa* genomes were submitted for probe design.

352 Candidate probe sets were designed using standard informatics workflows at Tecan

353 Genomics for SPE probes and Molecular Loop for MIP probes. We selected probes for

354 inclusion in the custom pool by subsampling based on perfect BLAST matches to at least

355 50 *Se* or 30 *Sa* genomes and exclusion of probes with alignments having fewer than five

356 mismatches to the alternative species or to a database of 10 *S. capitis* and 15 *S. hominis*

357 genomes. Additional filtering criteria removed probes at the extremes of melting

358 temperature and sequence composition (requiring no runs of >6 of a single base) and

359 established minimum probe spacing of at least 100 bases. All probes targeting both

360 genomes were synthesized and used in a single pool. Distributions of the number of

361 probes per CDS is provided in Figure 2. Probe sequences are in Supplemental Table 2.

362 Purified nucleic acids were split across several workflows as shown in Figure 1.

363 **Bulk RNA-seq library preparation and sequencing:** The NEBNext Ultra II Directional

364 Library Prep protocol (NEB #E7760S) was used with NEB's bacterial rRNA depletion kit

365 (NEB # E7850S) to make bulk RNA-seq libraries for whole transcriptome sequencing.

366 100 ng RNA input was used and adaptors were diluted 25-fold before ligating to the cDNA.

367 The same set of RNA samples was used in a hybrid workflow combining pre-amplification

368 with the NEBNext Ultra II Library Prep. 10 ng of total RNA was subjected to rRNA

369 depletion and then cDNA synthesis and amplification using the Single Primer Isothermal

370 Amplification (SPIA) method with the Crescendo cDNA synthesis for qPCR kit (Tecan

371 Genomics, Inc.). For these samples, RNA was not fragmented prior to cDNA synthesis.

372 After SPIA amplification, cDNA was fragmented using components of the Allegro Targeted

373 Genotyping V2 kit (Tecan Genomics, Inc.) before proceeding with end repair, adaptor

374 ligation (25-fold diluted adaptors) and PCR-enrichment using the NEBNext Ultra II

375 Directional Library Prep kit components. *E. coli* RNA was included as a positive control in

376 both the workflows. Index PCR was performed using the NEBNext Multiplex Oligos for

377   Illumina (Dual Index Primers Set 1) (NEB #E7600S). The final library purification was

378   carried out using 0.78x Ampure XP beads to remove the adaptor peak. Libraries were

379   assessed on the TapeStation and Qubit dsDNA high sensitivity assay before normalizing

380   to 4 nM.

381   **cDNA synthesis and SPIA amplification.** First and second strand cDNA synthesis were

382   performed using the Crescendo cDNA synthesis for qPCR protocol (Tecan Genomics,

383   Inc.). cDNA samples were then divided, with one portion subjected to SPIA amplification

384   per the Crescendo protocol. cDNA was purified with AMPure XP beads at 1.8x and eluted

385   in 33 ul of 1x TE. The resulting cDNA samples were quantified using NanoDrop following

386   the kit protocol and Qubit dsDNA high sensitivity assays. The samples that were amplified

387   with SPIA yielded >1 ug of cDNA and the non-SPIA samples were 6-8 ng.

388   **Targeted RNA-seq library preparation and sequencing using single-primer**

389   **extension probes.** SPIA-amplified cDNA samples were diluted 20-fold and 10-20 ng of

390   cDNA was used as input for the Allegro Targeted Genotyping V2 protocol (Tecan

391   Genomics, Inc.) with the custom probe pool. Non-amplified cDNA was not diluted and ~1

392   ng was used as input for the Allegro protocol. Genomic DNA was used at 10-50 ng input.

393   Each sample was enzymatically fragmented, followed by ligation of barcoded adaptors.

394   Barcoded samples were then purified, pooled, and placed in an overnight hybridization

395   reaction mixture with the probe pool. The following day (>12h) the DNA polymerase

396   enzyme was added to the reaction for extension at 72°C for 10 min. Post-enrichment

397   purification was done with 0.8x AMPure XP beads. A qPCR step was used to determine

398   the number of cycles for library amplification (11 cycles). The final library pool was bead

399   purified with 0.8x AMPure XP beads. A low template control (LTC) with 50 pg of *S.*

400 *epidermidis* RNA input and a mouse stool metagenome DNA sample were used as

401 sensitivity and specificity controls respectively. The library pool was adjusted to 32 nM

402 and 20 ul of the pool was used for sequencing (20 fmole/sample). The final sequencing

403 pool also comprised libraries created using the NEBNext Ultra II Library prep protocol and

404 Illumina DNA Prep (M) kit that were normalized to 4 nM and 0.2 nM respectively to

405 represent 20 fmoles for the NEBNext samples and 1 fmole for the Illumina DNA Prep

406 libraries. The exception was the *E. coli* RNA positive control (100 ng input) that was

407 pooled at a lower amount to represent 5 fmoles in the sequencing pool. The library pool

408 was loaded on NovaSeq 6000 SP flowcell 2x100bp run. The custom read 1 (R1) primer

409 provided in the Allegro kit was spiked into the Illumina read 1 primer position at 0.3 uM

410 and a phiX174 control library was spiked-in at 1% as recommended.

411 **Targeted RNA-seq library preparation and sequencing using molecular inversion**

412 **probes.** Targeted RNA libraries were prepared with the Low Input DNA Target Capture kit

413 (Molecular Loop Biosciences, Inc.) using the cDNA generated with the Crescendo kit

414 (Tecan Genomics, Inc.), including both the SPIA and non-SPIA pre-amplified samples, as

415 well as the gDNA samples. The sample input was not normalized (50 pg – 50 ng) to

416 evaluate its effect on probe hybridization efficiencies and consequently the number of

417 sequencing reads generated from each of these samples. Among the 32 samples, only

418 the human blood high molecular weight genomic DNA sample was pre-treated by heating

419 it to 95°C for 5 min for denaturation. The protocol was followed using the kit user guide

420 and the custom probe pool was hybridized for 18 hours (recommended ≥16 hours). After

421 hybridization, each library was amplified by PCR for 20 cycles, pooled at equal volume

422 (10 ul each), and the library pool was purified with 0.8x Agencourt RNAXP Clean beads

423 and eluted in 25ul of 1xTE. The library pool was quantified with the Qubit high sensitivity

424 dsDNA assay and on the TapeStation D1000 high-sensitivity ScreenTape, which resulted

425 in a single peak at ~415 bp. The library pool was loaded on a NovaSeq 6000 SP flowcell

426 2x100bp run. The custom workflow was selected for the sequencing run using custom

427 read1, read2 and sequencing primers provided by Molecular Loop. The phiX174 control

428 library was spiked-in at 1%.

429 **Illumina DNA Prep Libraries:** Libraries of *Se* and *Sa* gDNA were prepared using the

430 Illumina DNA Prep, Tagmentation kit (Illumina, 20018705) and indexed using IDT for

431 Illumina DNA/RNA UD Indexes (Illumina, 20042666). Reads were assembled using CLC

432 Genomics Workbench and the resulting assembly was compared to the reference

433 sequences for the *Se* and *Sa* genomes, confirming that the cultured cells represented the

434 expected genomes.

435 **Genome sequence alignment.** Reads were processed with trimmomatic 0.39 (41) to

436 remove low-quality reads. For bulk RNA and SPE reads, the adapters were removed from

437 the fastq reads using Cutadapt (42) version 4.4s, using the bait sequence

438 'AGATCGGAAGAG'. For MIP reads, seqtk v1.4-r122 was used

439 (https://github.com/lh3/seqtk) to remove the five bases from the 5′-end of the sequences

440 as *seqtk trimfq -b 5 <fastq>*. Reads were aligned to reference genomes *Sa* m2872

441 (GCF_017329165.1) and *Se* ATCC 14990 (GCF_006094375.1) as well as *Sa* USA300

442 (GCF_002993865.1) *Se* Tu3298 (https://github.com/ohlab/S.epi-CRISPRi-and-RNA-

443 Seq) using BWA-MEM (43) version 0.7.17-r1188 as *bwa mem -t 8 -k 19 -w 100.* The

444 mapped reads of each of the *Se* and *Sa* species were extracted with bamtools v2.5.2 (44)

445 with parameter *-isMapped 'true'* and subsequently filtered with MAPQ >30 and edit

446   distance NM <4. High-quality sequences were sorted and converted into .bam and .bed

447   files for downstream processing by SAMtools (45) v0.7.17-r1188. For bulk RNA-seq, the

448   read count aligned to each CDS region was computed using featureCounts v1.6.4(46)

449   (from Subread1.6.4) with -t CDS or -t rRNA, as appropriate, and summarized as counts

450   per CDS per million reads per kilobase CDS length (RPKM).

451

452   **Probe sequence alignment.** We also used BWA-MEM to align the SPE and MIP reads

453   to a database of the probe sequences. For MIP, we used the seqinr R package (47) to

454   extract the binding locations consisting of the first and the last 40bps of the probe

455   sequences and generated new fasta files with the 'first 40bps' and 'last 40bps' probe data.

456   We aligned the reads of each sample separately to the probe sequences of each species.

457   The mapped reads were extracted as before with bamtools v2.5.2 and subsequently

458   filtered with MAPQ >30 and NM <1 (perfect match) in SPE and MAPQ >30 and NM <3 in

459   MIP. The high-quality sequences were sorted and converted into bam and bed files for

460   downstream processing by SAMtools v0.7.17-r1188. The raw counts were generated with

461   *bedtools coverage* for each sample from the probe sequences (.bed) and cleaned read

462   (.bam) files by bedtools v2.31.0 (48). The raw counts were CPM normalized as

463   $1M \times r_i^{j,s}/R^{j,s}$, where $r_i^{j,s}$ denotes the number of reads mapped to location i in sample j

464   and species $s, s \in \{Se, Sa\}$ and $R^{j,s}$ is the total number of mapped reads in sample j and

465   species s. Probe count tables are provided as **Supplemental Tables 6-10**.

466

467   **Differential Expression Analysis and Correlation Analyses**

468   Differential expression analysis across stress conditions was performed in Rv4.4.1 (49)

469   using DESeq2 (50) using a two factor design matrix accounting for stress treatment (TSB,

470   acid, or urea) and SPIA status including interaction terms. In all comparisons, differentially

471   expressed genes were defined as those with fold-change magnitude > 4 and adjusted p-

472   value < 0.01. Correlation matrices were calculated in R as Euclidean distances using the

473   stats package (49) and pairwise correlations were assessed via Pearson correlation

474   coefficients (R values) of best fit linear models to two-dimensional data using ggpubr (51).

475

476   **Data availability**

477   All data used in these analyses are available at the NCBI Gene Expression Omnibus

478   under accession GSE279187.

479

480   **Acknowledgements**

481   We thank the Genome Technologies Scientific Service for performing sequence data

482   collection and Mitch Kostich for helpful discussions regarding data analysis.

483

484   **Author Contributions**

485   GD and MA performed data analysis and wrote the manuscript. SC wrote the manuscript.

486   PS performed the targeted sequencing experiments. IB and EM contributed to data

487   analysis. EA performed the RHE experiments. JO and MA conceived the study.

488

489   1.   Huang J, Jiang Z, Wang Y, Fan X, Cai J, Yao X, Liu L, Huang J, He J, Xie C, Wu Q, Cao Y,

490   Leung EL. 2020. Modulation of gut microbiota to overcome resistance to immune

491   checkpoint blockade in cancer immunotherapy. Curr Opin Pharmacol, 2020/07/04 ed.

492   54:1–10.

493   2.   van den Berg FF, van Dalen D, Hyoju SK, van Santvoort HC, Besselink MG, Wiersinga

494   WJ, Zaborina O, Boermeester MA, Alverdy J. 2021. Western-type diet influences

495   mortality from necrotising pancreatitis and demonstrates a central role for butyrate.

496   Gut, 20200901st ed. 70:915–927.

497   3.   Czaran T, Hoekstra RF. 2009. Microbial communication, cooperation and cheating:

498   quorum sensing drives the evolution of cooperation in bacteria. PLoS One, 2009/08/18

499   ed. 4:e6655.

500   4.   Brenner K, Karig DK, Weiss R, Arnold FH. 2007. Engineered bidirectional

501   communication mediates a consensus in a microbial biofilm consortium. Proc Natl

502   Acad Sci U A, 2007/10/26 ed. 104:17300–4.

503   5.   Franzosa EA, McIver LJ, Rahnavard G, Thompson LR, Schirmer M, Weingart G, Lipson

504   KS, Knight R, Caporaso JG, Segata N, Huttenhower C. 2018. Species-level functional

505   profiling of metagenomes and metatranscriptomes. Nat Methods, 2018/11/01 ed.

506   15:962–968.

507   6.   Peters BA, Wilson M, Moran U, Pavlick A, Izsak A, Wechter T, Weber JS, Osman I, Ahn J.

508   2019. Relating the gut metagenome and metatranscriptome to immunotherapy

509   responses in melanoma patients. Genome Med, 2019/10/11 ed. 11:61.

510   7.   Korry BJ, Cabral DJ, Belenky P. 2020. Metatranscriptomics Reveals Antibiotic-Induced

511   Resistance Gene Expression in the Murine Gut Microbiota. Front Microbiol 11:322.

512   8.   Kang D, Shi B, Erfe MC, Craft N, Li H. 2015. Vitamin B12 modulates the transcriptome

513   of the skin microbiota in acne pathogenesis. Sci Transl Med 7:293ra103.

514   9.   Abu-Ali GS, Mehta RS, Lloyd-Price J, Mallick H, Branck T, Ivey KL, Drew DA, DuLong C,

515   Rimm E, Izard J, Chan AT, Huttenhower C. 2018. Metatranscriptome of human faecal

516   microbial communities in a cohort of adult men. Nat Microbiol, 20180115th ed.

517   3:356–366.

518   10. Jousset A, Bienhold C, Chatzinotas A, Gallien L, Gobet A, Kurm V, Kusel K, Rillig MC,

519   Rivett DW, Salles JF, van der Heijden MG, Youssef NH, Zhang X, Wei Z, Hol WH. 2017.

520   Where less may be more: how the rare biosphere pulls ecosystems strings. ISME J,

521   2017/01/11 ed. 11:853–862.

522   11. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR,

523   Tap J, Bruls T, Batto JM, Bertalan M, Borruel N, Casellas F, Fernandez L, Gautier L,

524   Hansen T, Hattori M, Hayashi T, Kleerebezem M, Kurokawa K, Leclerc M, Levenez F,

525   Manichanh C, Nielsen HB, Nielsen T, Pons N, Poulain J, Qin J, Sicheritz-Ponten T, Tims

526   S, Torrents D, Ugarte E, Zoetendal EG, Wang J, Guarner F, Pedersen O, de Vos WM,

527   Brunak S, Dore J, Meta HITC, Antolin M, Artiguenave F, Blottiere HM, Almeida M,

528   Brechot C, Cara C, Chervaux C, Cultrone A, Delorme C, Denariaz G, Dervyn R,

529   Foerstner KU, Friss C, van de Guchte M, Guedon E, Haimet F, Huber W, van Hylckama-

530   Vlieg J, Jamet A, Juste C, Kaci G, Knol J, Lakhdari O, Layec S, Le Roux K, Maguin E,

531   Merieux A, Melo Minardi R, M'Rini C, Muller J, Oozeer R, Parkhill J, Renault P, Rescigno

532   M, Sanchez N, Sunagawa S, Torrejon A, Turner K, Vandemeulebrouck G, Varela E,

533   Winogradsky Y, Zeller G, Weissenbach J, Ehrlich SD, Bork P. 2011. Enterotypes of the

534   human gut microbiome. Nature, 2011/04/22 ed. 473:174–80.

535   12. Claussen JC, Skieceviciene J, Wang J, Rausch P, Karlsen TH, Lieb W, Baines JF, Franke

536   A, Hutt MT. 2017. Boolean analysis reveals systematic interactions among low-

537   abundance species in the human gut microbiome. PLoS Comput Biol, 2017/06/24 ed.

538   13:e1005361.

539   13. Human Microbiome Project C. 2012. Structure, function and diversity of the healthy

540   human microbiome. Nature, 2012/06/16 ed. 486:207–14.

541   14. Iwase T, Uehara Y, Shinji H, Tajima A, Seo H, Takada K, Agata T, Mizunoe Y. 2010.

542   Staphylococcus epidermidis Esp inhibits Staphylococcus aureus biofilm formation

543   and nasal colonization. Nature, 2010/05/21 ed. 465:346–9.

544   15. Uberoi A, Bartow-McKenney C, Zheng Q, Flowers L, Campbell A, Knight SAB, Chan N,

545   Wei M, Lovins V, Bugayev J, Horwinski J, Bradley C, Meyer J, Crumrine D, Sutter CH,

546   Elias P, Mauldin E, Sutter TR, Grice EA. 2021. Commensal microbiota regulates skin

547   barrier function and repair via signaling through the aryl hydrocarbon receptor. Cell

548   Host Microbe, 20210701st ed. 29:1235-1248 e8.

549   16. Byrd AL, Deming C, Cassidy SKB, Harrison OJ, Ng WI, Conlan S, Program NCS, Belkaid

550   Y, Segre JA, Kong HH. 2017. Staphylococcus aureus and Staphylococcus epidermidis

551   strain diversity underlying pediatric atopic dermatitis. Sci Transl Med, 2017/07/07 ed.

552   9.

553   17. Zhou W, Spoto M, Hardy R, Guan C, Fleming E, Larson PJ, Brown JS, Oh J. 2020. Host-

554   Specific Evolutionary and Transmission Dynamics Shape the Functional

555   Diversification of Staphylococcus epidermidis in Human Skin. Cell 180:454-470.e18.

556   18. Bernstein JA, Khodursky AB, Lin PH, Lin-Chao S, Cohen SN. 2002. Global analysis of

557   mRNA decay and abundance in Escherichia coli at single-gene resolution using two-

558   color fluorescent DNA microarrays. Proc Natl Acad Sci U A, 2002/07/18 ed. 99:9697–

559   702.

560   19. Wu L, Liu X, Schadt CW, Zhou J. 2006. Microarray-based analysis of subnanogram

561   quantities of microbial community DNAs by using whole-community genome

562   amplification. Appl Env Microbiol, 2006/07/06 ed. 72:4931–41.

563   20. Gardner SN, Jaing CJ, McLoughlin KS, Slezak TR. 2010. A microbial detection array

564   (MDA) for viral and bacterial detection. BMC Genomics, 2010/11/27 ed. 11:668.

565   21. Samorodnitsky E, Datta J, Jewell BM, Hagopian R, Miya J, Wing MR, Damodaran S,

566   Lippus JM, Reeser JW, Bhatt D, Timmers CD, Roychowdhury S. 2015. Comparison of

567   custom capture for targeted next-generation DNA sequencing. J Mol Diagn,

568   2014/12/22 ed. 17:64–75.

569   22. Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T,

570   Giannoukos G, Fisher S, Russ C, Gabriel S, Jaffe DB, Lander ES, Nusbaum C. 2009.

571   Solution hybrid selection with ultra-long oligonucleotides for massively parallel

572   targeted sequencing. Nat Biotechnol, 2009/02/03 ed. 27:182–9.

573   23. Depledge DP, Palser AL, Watson SJ, Lai IY, Gray ER, Grant P, Kanda RK, Leproust E,

574   Kellam P, Breuer J. 2011. Specific capture and whole-genome sequencing of viruses

575   from clinical samples. PLoS One, 2011/11/30 ed. 6:e27805.

576   24. Nimmo C, Shaw LP, Doyle R, Williams R, Brien K, Burgess C, Breuer J, Balloux F, Pym

577   AS. 2019. Whole genome sequencing Mycobacterium tuberculosis directly from

578   sputum identifies more genetic diversity than sequencing from culture. BMC

579   Genomics, 2019/05/22 ed. 20:389.

580   25. Benjamino J, Leopold B, Phillips D, Adams MD. 2021. Genome-Based Targeted

581   Sequencing as a Reproducible Microbial Community Profiling Assay. mSphere

582   6:10.1128/msphere.01325-20.

583   26. Oh J, Byrd AL, Park M, Kong HH, Segre JA. 2016. Temporal Stability of the Human Skin

584   Microbiome. Cell 165:854–866.

585   27. Grice EA, Segre JA. 2011. The skin microbiome. Nat Rev Microbiol 9:244–253.

586   28. Agostinetto G, Bozzi D, Porro D, Casiraghi M, Labra M, Bruno A. 2021. SKIOME Project:

587   a curated collection of skin microbiome datasets enriched with study-related

588   metadata. bioRxiv https://doi.org/10.1101/2021.08.17.456635.

589   29. Hardenbol P, Banér J, Jain M, Nilsson M, Namsaraev EA, Karlin-Neumann GA, Fakhrai-

590   Rad H, Ronaghi M, Willis TD, Landegren U, Davis RW. 2003. Multiplexed genotyping

591   with sequence-tagged molecular inversion probes. Nat Biotechnol 21:673–678.

592   30. Myrmel M, Oma V, Khatri M, Hansen HH, Stokstad M, Berg M, Blomström A-L. 2017.

593   Single primer isothermal amplification (SPIA) combined with next generation

594   sequencing provides complete bovine coronavirus genome coverage and higher

595   sequence depth compared to sequence-independent single primer amplification

596   (SISPA). PLOS ONE 12:e0187780.

597   31. Korem T, Zeevi D, Suez J, Weinberger A, Avnit-Sagi T, Pompan-Lotan M, Matot E, Jona

598   G, Harmelin A, Cohen N, Sirota-Madi A, Thaiss CA, Pevsner-Fischer M, Sorek R, Xavier

599   R, Elinav E, Segal E. 2015. Growth dynamics of gut microbiota in health and disease

600   inferred from single metagenomic samples. Science 349:1101–1106.

601   32. Grether-Beck S, Felsner I, Brenden H, Kohne Z, Majora M, Marini A, Jaenicke T,

602   Rodriguez-Martin M, Trullas C, Hupe M, Elias PM, Krutmann J. 2012. Urea Uptake

603   Enhances Barrier Function and Antimicrobial Defense in Humans by Regulating

604   Epidermal Gene Expression. J Invest Dermatol 132:1561–1572.

605   33. Zhou C, Bhinderwala F, Lehman MK, Thomas VC, Chaudhari SS, Yamada KJ, Foster

606   KW, Powers R, Kielian T, Fey PD. 2019. Urease is an essential component of the acid

607   response network of Staphylococcus aureus and is required for a persistent murine

608   kidney infection. PLoS Pathog 15:e1007538.

609   34. Mager LF, Burkhard R, Pett N, Cooke NCA, Brown K, Ramay H, Paik S, Stagg J, Groves

610   RA, Gallo M, Lewis IA, Geuking MB, McCoy KD. 2020. Microbiome-derived inosine

611   modulates response to checkpoint inhibitor immunotherapy. Science, 2020/08/15 ed.

612   369:1481–1489.

613   35. Patnode ML, Beller ZW, Han ND, Cheng J, Peters SL, Terrapon N, Henrissat B, Le Gall S,

614   Saulnier L, Hayashi DK, Meynier A, Vinoy S, Giannone RJ, Hettich RL, Gordon JI. 2019.

615   Interspecies Competition Impacts Targeted Manipulation of Human Gut Bacteria by

616   Fiber-Derived Glycans. Cell, 2019/09/21 ed. 179:59-73 e13.

617   36. Fischer CN, Trautman EP, Crawford JM, Stabb EV, Handelsman J, Broderick NA. 2017.

618   Metabolite exchange between microbiome members produces compounds that

619   influence Drosophila behavior. Elife, 2017/01/10 ed. 6.

620   37. Kim SG, Becattini S, Moody TU, Shliaha PV, Littmann ER, Seok R, Gjonbalaj M, Eaton V,

621   Fontana E, Amoretti L, Wright R, Caballero S, Wang ZX, Jung HJ, Morjaria SM, Leiner

622   IM, Qin W, Ramos R, Cross JR, Narushima S, Honda K, Peled JU, Hendrickson RC, Taur

623   Y, van den Brink MRM, Pamer EG. 2019. Microbiota-derived lantibiotic restores

624   resistance against vancomycin-resistant Enterococcus. Nature, 2019/08/23 ed.

625   572:665–669.

626   38. Spoto M, Riera Puma JP, Fleming E, Guan C, Ondouah Nzutchi Y, Kim D, Oh J. 2022.

627   Large-Scale CRISPRi and Transcriptomics of *Staphylococcus epidermidis* Identify

628   Genetic Factors Implicated in Lifestyle Versatility. mBio 0:e02632-22.

629   39. Long LL, Svenson KL, Mourino AJ, Michaud M, Fahey JR, Waterman L, Vandegrift KL,

630   Adams MD. 2021. Shared and distinctive features of the gut microbiome of C57BL/6

631   mice from different vendors and production sites, and in response to a new vivarium.

632   Lab Anim 50:185–195.

633   40. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D,

634   Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis.

635   Bioinforma Oxf Engl 31:3691–3693.

636   41. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina

637   sequence data. Bioinformatics 30:2114–2120.

638   42. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput

639   sequencing reads. 1. EMBnet.journal 17:10–12.

640   43. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with

641   BWA-MEM. ArXiv 1303.

642   44.   Barnett D, Garrison E, Quinlan A, Strömberg M, Marth G. 2011. BamTools: a C++ API

643   and toolkit for analyzing and managing BAM files. Bioinformatics 27.

644   45.   Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G,

645   Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics

646   25:2078–2079.

647   46.   Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program

648   for assigning sequence reads to genomic features. Bioinforma Oxf Engl 30:923–930.

649   47.   Gouy M, Milleret F, Mugnier C, Jacobzone M, Gautier C. 1984. ACNUC: a nucleic acid

650   sequence data base and analysis system. Nucleic Acids Res 12:121–127.

651   48.   Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic

652   features. Bioinforma Oxf Engl 26:841–842.

653   49.   R Core Team. 2024. R: A Language and Environment for Statistical Computing. R

654   Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

655   50.   Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and

656   dispersion for RNA-seq data with DESeq2. Genome Biol 15:550.

657   51.   Kassambara A. 2016. ggpubr: "ggplot2" Based Publication Ready Plots.

658

659