Baker IDI Research Online
http://library.bakeridi.edu.au

This is the postprint version of the work. It is the manuscript that was accepted by the journal following peer review. It does not include the publisher's layout and pagination.

Ziemann M, Kamboj A, Hove RM, Loveridge S, El-Osta A, Bhave M. Analysis of the barley leaf transcriptome under salinity stress using mRNA-Seq. Acta Physiol Plant 2013;35(6):1915-24.

http://hdl.handle.net/11187/1551

Copyright © Springer Verlag. This file is for personal use. Further distribution is not permitted.
The final publication is available at link.springer.com

1 **Analysis of the barley leaf transcriptome under salinity stress using mRNA-Seq**

2

3

4

5

6

7 Mark Ziemann[1], Atul Kamboj[2], Runyararo M. Hove[2], Shanon Loveridge[1], Assam El-Osta[1],

8 Mrinal Bhave[2*]

9

10

11 [1]Baker IDI Heart and Diabetes Institute, Melbourne, VIC 3004, Australia

12 [2]Environment and Biotechnology Centre, Faculty of Life and Social Sciences,

13 Swinburne University of Technology, PO Box 218, Hawthorn, VIC 3122, Australia.

14

15

16

17 *Corresponding author:

18 Phone 61-3-9214-5759

19 Facsimile 61-3-9819-0834

20 Email mbhave@swin.edu.au

21 **Abstract**

22 Salinity is a threat to the crops worldwide, and together with drought, it is predicted to be a
23 serious constraint to food security. However, understanding the impact of this stressor on
24 plants is a major challenge due to the involvement of numerous genes and regulatory
25 pathways. While transcriptomic analyses of barley (*Hordeum vulgare* L.) salt stress have
26 been reported with microarrays, there are no reports as yet of the use of mRNA-Seq. We
27 demonstrate the utility of mRNA-Seq by analysing cDNA libraries derived from acutely salt-
28 stressed and unstressed leaf material of *H. vulgare* cv. Hindmarsh. The data yielded >50
29 million sequence tags which aligned to 26,944 sequences in the Unigene reference database.
30 To gain maximum information, we performed *de novo* assembly of unaligned reads and
31 discovered >3,800 contigs, termed novel tentative consensus sequences (NTCs), which are
32 either new, or significant improvements on current databases. Differential gene expression
33 screening found 48 significantly up-regulated and 62 significantly down-regulated transcripts.
34 The work provides comprehensive insights into genome-wide effects of salinity and is a new
35 resource for study of gene regulation in barley and wheat. Further, the bioinformatics
36 workflow may be applicable to other non-model plants to establish their transcriptomes and
37 identify unique sequences.

38

39

40

41 Key words:
42 Salinity, barley, gene expression, mRNA-Seq

43

44

Introduction

Salt and drought stresses are the two most important environmental stresses which limit plant growth and development. Over 100 countries in the world were identified to be affected by salinity, covering about 350 million hectares in 1989 (Rengasamy 2006), and the scale of the problem seems to be increasing at an alarming rate, with >800 M ha (>6% of total land area) affected by 2000 (Munns and Tester 2008). Salinity, together with drought, has far-reaching implications for food security, economic sustainability and the irreplaceable biodiversity of any affected area, and the challenges are expected to be exacerbated by the projected impact of climate change. The effects of water-insufficiency stresses have been studied extensively, and in summary, they limit water and micronutrient uptake due to reverse osmotic effects, and lead to closure of stomata, decline in carbon metabolism, stunted growth, ion/salt toxicity and reduced yield (Langridge et al. 2006; Munns and Tester 2008).

For plants to survive under such conditions, they must be able to sense and respond rapidly. Molecular studies on various plants including *Arabidopsis thaliana* show that these events involve complex networks of gene regulation (Bartels and Sunkar 2005; Langridge et al. 2006; Munns and Tester 2008), including intracellular signalling pathways such as those mediated by plant hormones such as abscisic acid (Ma et al. 2009) and ethylene (Xu et al. 2007) and effected through specific transcription factors (Urano et al. 2010), and expression of diverse functional genes for osmo-regulation/cell protection/acclimation, e.g., *Nax* (Munns 2005), aquaporins (Tyerman et al. 2002; reviewed in Forrest and Bhave 2007), dehydrins (Close 1996), redox enzymes (Selote and Khanna-Chopra 2006) and chaperones (Meiri and Breiman 2009). While the perception of salt and drought may share a common mechanism (Shinozaki and Yamaguchi-Shinozaki 1997), each stress may also have some unique effects.

For the continued improvement of crops in the face of future environmental and socio-economic challenges, our understanding of crop responses to drought and salinity will need to grow. Central to these is a comprehensive understanding of the roles of individual genes, their transcripts including alternative splice forms, their protein products, as well as the 'sum' of all pathways that plants use to manage abiotic stresses, often in a plant-specific manner. Despite being staple foods around the world, the elucidation of the complete genome sequence of wheat has been hindered by the complexity of its genome, while a draft barley

77 whole genome sequence has only recently been described (International Barley Genome
78 Sequencing Consortium et al. 2012). Even in the absence of a whole genome sequence,
79 transcript profiling has provided important data in recent years for the cataloguing of their
80 genes. These reports include the barley GeneChip (Walia et al. 2006) that identified several
81 stress responsive genes; profiling of drought tolerant versus susceptible wheat lines that led to
82 identification of altered responses of several genes and a new transcription factor
83 (Mohammadi et al. 2007); wheat arrays that led to several hundred genes related to abiotic
84 stress response (Kawaura et al. 2008); and wheat GeneChip (Schreiber et al. 2009) that
85 produced transcript data highly comparable to the barley gene chip (Druka et al. 2006),
86 supporting the functional closeness of the two species.

87

88 The next-generation mRNA-Seq, a high throughput cDNA sequencing technology, is a
89 powerful method for rapid characterisation of transcript sequences and gene expression levels
90 in biological samples. It is being applied widely in human genetics and medicine, but is still
91 an emerging technology for plants. The use of high-throughput sequencers such as the
92 Roche-454, Solid and Illumina systems has considerable potential to bring high resolution
93 transcriptome maps to non-model species such as barley. Marioni et al. (2008) critically
94 evaluated gene expression profiling by RNA-Seq by the Illumina platform to that by
95 Affymetrix arrays from the same RNA samples, and concluded that RNA-Seq was not only
96 comparable in elucidating differentially expressed genes, but also had added capabilities of
97 detecting transcripts with low level expression, identifying sequence variants and new
98 transcripts. Transcriptome analysis from short-read Illumina sequencing is now beginning to
99 be carried out for crop species, e.g., rice (Mizuno et al. 2010) and soybean (Severin et al.
100 2010), which have the advantage of reference whole genome data, and also species such as
101 chickpea (Garg et al. 2011) without such information. By the FAO (2005) classification of
102 salinity tolerance, both corn and soybean are moderately tolerant, wheat is tolerant, while
103 barley is classified as 'highly tolerant'; hence it may display important genetic attributes
104 under salt challenge. The cultivar Hindmarsh was chosen for transcriptome analysis here
105 because it is the most widely cultivated barley variety in Australia (GRDC 2008;
106 http://www.grdc.com.au/uploads/documents/GRDC_ImpAss_BarleyBreeding1.pdf; p17) and
107 is particularly suited to regions of South-Eastern Australia with lower rainfall (Modra Seeds
108 Fact Sheet). In this paper, we compare the transcriptomes of the leaf of barley, a major cereal
109 crop and a close relative of wheat, in acute salt stressed versus control conditions, and show
110 the utility of mRNA-Seq for qualitative and quantitative analyses of gene expression profiles.

111 This analysis aims to identify genes which may confer resistance to acute salinity stress and
112 may thereby be candidates for future crop improvement where soil salinity is posing an
113 increasing problem.

114

115 **Materials and methods**

116 **Plant material**

117 *Hordeum vulgare* L. cv. Hindmarsh seedling were grown in trays filled with potting mix
118 consisting of vermiculite:perlite (2:1), in a Thermoline plant growth cabinet set at 12 h of
119 light per day, 72% humidity and 20°C temperature for 14 days. The 12 h time point
120 represents a period of acute stress where differential genes are likely to be maximally altered
121 in expression as suggested by previous reports in Arabidopsis (Seki et al. 2002). Salt stress
122 was applied to five plants by supplying 150 mM NaCl in Hoagland's solution (Hoagland and
123 Arnon 1950) for 12 h, while five others remained unstressed (controls). Leaves of each plant
124 were harvested individually, snap-frozen in liquid nitrogen and stored at -80°C.

125

126 **RNA Isolation**

127 Snap-frozen leaf material from individual plants was crushed in a microcentrifuge tube using
128 a sterilized metal rod to a fine powder. RNA was extracted using TRIsure reagent (Bioline
129 Australia). After phenol-chloroform extraction, the RNA was precipitated, the pellet washed,
130 air-dried and dissolved in DEPC-treated water. It was then treated with 10U of RQ1 RNase-
131 free DNaseI (Promega Australia) in the presence of 2U of RNase inhibitor (Bioline Australia)
132 for 30 minutes at 37°C. RNA was recovered by LiCl precipitation (Ambion Technical
133 Bulletin #160) and dissolved in 20 μL DEPC-treated water. The integrity of RNAs was
134 assessed with capillary electrophoresis on a MultiNA system (Shimadzu Corporation, Japan).

135

136 **Next Generation mRNA Sequencing**

137 RNA from two salt-stressed plants was pooled in equal quantities for the mRNA-Seq library
138 preparation, as reported for mRNA-Seq (Mizuno et al. 2010) and other methods (Ando and
139 Grumet 2010), to minimise any biological variations in transcriptomes. RNA from two
140 control plants was pooled likewise. Libraries were produced as per the Illumina mRNA-Seq
141 library preparation protocol (September 2009, CA, USA). The main steps included mRNA

142 enrichment on oligo(dT) beads, reverse transcription (using random primer), synthesis of
143 second strand, end repair, 3' adenylation, sequencing adapter ligation and PCR amplification.
144 A V4 kit was used for cluster generation with a DNA concentration of 8 pM on the Illumina
145 Cluster Station and the flow cell was loaded onto Illumina Genome Analyzer IIx for
146 sequencing-by-synthesis for 76 cycles (V4 reagents). Image analysis with RTA v1.8 was
147 performed for base-calling and sequence file generation.

148

149 **Bioinformatics methods**

150 Datasets were filtered for spurious reads using the Fastx Artifacts Filter and poor quality
151 bases were removed from the 3' end with the Fastq Quality Trimmer using a threshold of Q30
152 (http://hannonlab.cshl.edu/fastx_toolkit/). Initially, *Hordeum vulgare* Unigene transcript
153 sequences downloaded from the NCBI database (http://www.ncbi.nlm.nih.gov/unigene; last
154 accessed November 2011) were used as reference sequences. This database consisted of
155 26,941 transcripts including those annotated as 'complete CDS' and 'partial CDS'. The 76
156 nucleotide (nt) mRNA reads were aligned using Burrows-Wheeler Aligner (BWA) (Li and
157 Durbin 2009) using default settings which allowed up to 4 mismatches in the 76b reads.

158

159 Unaligned reads were extracted with SAMtools software (Li et al. 2009) and underwent *de*
160 *novo* assembly using the 'Assembly By Short Sequence' (ABySS) software package
161 (Simpson et al. 2009) to elucidate any previously unidentified transcripts. ABySS was run at
162 a range of k-mer lengths from 27 to 63, using a coverage threshold of 3x. To increase the
163 contiguity of the assembly, these discrete assemblies were concatenated and re-assembled
164 using a range of k-mer lengths. Any contigs less than 100 bp were discarded and remaining
165 contigs were named tentative consensus sequences (TCs). To determine whether these TCs
166 represented novel sequences, they were BlastN searched to the above Barley Unigene
167 collection as well as to a rice cDNA database (http://rice.plantbiology.msu.edu/). Novel TCs
168 were expected to find a relatively strong hit to the rice database and a poor match in the
169 barley TA database. After trial and error of various ratios (that turned out too stringent or too
170 non-selective; data not shown), we implemented a rice/barley blast bit score ratio threshold of
171 ≥2 and discarded TCs with a score <2, leaving a set of novel TCs (NTCs), which were
172 subsequently appended to the Unigene reference. A work-flow diagram is given in Fig. 1.

173

174 To perform differential gene expression (DGE) analysis, the Q30 quality trimmed reads were
175 aligned with BWA to the updated transcriptome database. Ambiguously aligning reads were
176 not filtered, as this would have biased against contigs which have been extended in length by
177 the assembly process. Counts for each transcript were extracted with SAMtools (idxstats
178 feature) and these were subject to DGE analysis using the DESeq software package (Anders
179 and Huber 2010), using the conservative "blind" method to estimate variance despite lack of
180 replicates. Transcripts with false discovery rate (Benjamini-Hochberg procedure) adjusted p-
181 values < 0.05 were considered significantly differentially expressed.
182
183 Gene ontology (GO) analysis was performed for up- and down-regulated sets of 200 genes
184 (selected based on adjusted p-value rank) by first mapping each barley gene to its closest
185 BlastN match in the rice cDNA database as above, using an e-value threshold of <0.1. The
186 rice locus name sets were then analysed with the agriGO Singular Enrichment Analysis tool
187 (Du et al. 2010), using the suggested rice whole transcriptome background. Significance of
188 the gene set enrichment was evaluated with Fisher test using Yekutieli FDR adjustment, with
189 a significance threshold set at 0.05.
190

191 **Validation of NGS findings by semi-quantitative reverse transcriptase PCR**

192 The transcripts analysed by semi-quantitative reverse transcriptase PCR (sqRT-PCR) were
193 those of Hv.469, Hv.10251, Hv.8888, Hv.8276, Hv.22598, Hv.20929, Hv.30571, Hv.25954,
194 Hv. 22828, Hv.808, Hv.23281, with actin and α-tubulin as housekeeping controls (Suprunova
195 et al. 2004). Primers were designed using Netprimer
196 (http://www.premierbiosoft.com/netprimer/index.html), with the following criteria: length 15-
197 25 bases; GC content ~50%, minimal secondary structures, and comparable annealing
198 temperatures (~60°C) of the primers of a pair, to amplify products of 60-372 bp (Table S1).
199 Three salt-stressed and three control plants were used to extract total RNA individually and
200 cDNAs were synthesised from each using Bioscript reverse transcriptase (Bioline, Australia).
201 Purified RNA (1 µg) was mixed with 1 µL of oligo(dT)18 primer (0.5µg/µL) in 12 µL,
202 incubated at 70 °C for 5 minutes and chilled on ice. 10 U (0.25 µL) of RNase inhibitor,
203 40mM (1 µL) dNTP, 4 µL of 1X Reaction buffer and 50 U (0.25 µL) of Bioscript (all from
204 Bioline) were added to a final volume of 20µL, and incubated for 1 h at 37 °C. The reaction
205 was terminated at 70 °C for 10 min and the cDNAs stored at -20 °C. Absence of genomic

DNA in cDNAs was confirmed by PCR using intron-spanning primers for actin and α-tubulin (Suprunova et al. 2004) and comparisons to gDNA amplifications. The 25 µL PCR mixes contained 200 ng cDNA, 12.5 µL of 2X Biomix (Bioline; contains *Taq* polymerase, dNTPs) and 0.5 µL (0.1 µg/µL) each of the forward and reverse primers. For sqRT-PCR of each gene, four identical PCR tubes were prepared for each cDNA, to amplify for 20, 25, 30 and 35 cycles for genes showing relatively high expression in mRNA-Seq, or 25, 30, 35 and 40 cycles for those showing relatively low expression, with typical annealing temperatures of 60°C. Aliquots (5 µL) of each reaction were electrophoresed and the intensity of bands recorded using Chemidoc XRS Documentation Station and Quantity One software (Bio-Rad). Differential expression (fold change) was calculated using actin and α-tubulin (Suprunova et al. 2004) as housekeeping controls that exhibit relatively constant expression.

Results

Next Generation Sequencing

The mRNA-Seq libraries from the control and salt-stressed cDNA were each loaded on one lane of Illumina Genome Analyser IIx, and yielded over 50 million sequence tags (Table 1). The reads were curated with artefact filtering and quality trimming and then aligned to the current barley Unigene database (see Methods) using the BWA program under default settings, which allowed up to 4 mismatches in 76 nt reads. From this dataset it was found that out of the 26,944 present in the Unigene reference database, 21,336 transcripts were detected in the control and 21,574 detected in the salt-stressed sample (1 read or more).

Identification of new transcripts

To discover previously unrecognised transcripts, the remaining 16,434,520 unaligned reads underwent a two-step *de novo* assembly in ABySS (Fig. 2). A range of k-mer lengths was utilised for phase 1, which generated 5,723,131 overlapping contigs from 18 assemblies (k-mer range 27, 29, etc., up to 63). These contigs then underwent phase 2 assembly, with the average contig length and N50 length improving dramatically and the number reducing to <50,000. We selected the k55 assembly for downstream analyses, which yielded 39,707 contigs ≥100 bp in length. These had an average length of 343.9 bp and N50 length of 518 bp (Table 2). The longest contig (TC21595; 13,710 bp) putatively encodes an auxin transport

237 protein of the 'BIG-like' family, based on homology to a *Brachypodium* cDNA sequence. To
238 hone in on potentially novel sequences, a rice/barley blastn ratio of ≥2 was implemented,
239 which removed 90.3% of contigs, resulting in 3,828 potentially novel tentative consensus
240 sequences (NTCs). Of the top 1,000 transcripts ranked by expression (DESeq normalised
241 expression across both control and salt datasets), 61 were NTCs, demonstrating that some of
242 these are highly expressed in the barley leaf. These often had a close but incomplete blast hit
243 in the Unigene database, compared to a longer but less-similar blast hit from rice. For
244 instance, 45,336 reads aligned to NTC41084, its closest barley Unigene match being to
245 Genbank accession BF620510.2 with an alignment length of 572 bp, while the rice match
246 spanned 1,588 bp.
247
248 **Analysis of differential gene expression**
249 Quality trimmed reads from control and salt-stressed samples were aligned to the appended
250 reference sequence, which saw the total unaligned reads decrease from 16,434,520 to
251 13,505,967 (Table 3). Using DESeq to scan for differential gene expression between control
252 and acute salinity stress in leaves using a negative binomial model, we found 110 genes to be
253 significantly de-regulated (FDR adjusted p-val < 0.05). From these, 48 transcripts showed
254 increases and 62 showed decreases (Fig. 3). The top 20 differentially expressed (up- and
255 down-regulated) transcripts ranked by p-value from the barley Unigene and NTC sets are
256 shown in Table 4. The list of up-regulated genes includes a number of genes (or homologs
257 thereof) which have been shown previously to mediate osmotic/drought/salinity stress
258 tolerance, such as cellulose synthase-like protein, lipoxygenase 2.1, protein phosphatase 2C,
259 late embryogenesis abundant, calcium/calmodulin dependent protein kinase, as well as those
260 encoding membrane bound proteins such as a peptide transporter, two plasma membrane
261 ATPases and a novel wall-associated receptor kinase. Down-regulated transcripts include
262 those in the Jumonji, Pumilio RNA binding and MYB transcription factor classes, as also
263 several transcripts of unknown function. The full sequence data set is available at SRA in
264 Genbank (accession number SRA062960) and the differential gene expression spread sheet is
265 attached as Table S3.
266
267 Gene ontology analysis was performed with sets of 200 differentially regulated barley genes
268 (ranked by p-value). As few GO analysis tools exist as yet for barley, we mapped each barley

269 sequence to its best blast hit in rice, and compared these lists to the rice transcriptome-wide
270 background using agriGO (Du et al. 2010). The set of genes which were down-regulated
271 (only 120 loci had GO annotation) did not show any significant enrichment (p-val <0.05), but
272 the set of 200 up-regulated genes (only 104 loci had any GO annotation) was significantly
273 enriched for genes annotated with 'response to abiotic stimulus' (Table S2). The genes in
274 significant GO terms in the up-regulated list were all linked to response to osmotic stress,
275 desiccation or water limitation. Combining the up- and down regulated lists showed that the
276 GO terms 'response to chemical stimulus' and 'response to abiotic stimulus' were over-
277 represented.

278

279 **Validation of transcript profiles by sqRT-PCR**

280 SqRT-PCR of eleven randomly selected sequences from the mRNA-Seq data resulted in
281 successful amplification of the bands of expected sizes (Fig. S2; Table S1), and quantitation
282 of the band intensities in relation to house-keeping controls largely supported the direction of
283 change of expression as detected by mRNA-Seq for all transcripts (Fig. S1). Exceptions to
284 this occurred when the fold-changes detected were slight, and the deviations were within the
285 observed variation of the 'housekeeping' genes alpha-tubulin (-1.39) and actin (+1.17) in
286 mRNA-Seq datasets. In cases of extreme fold-changes, sqRT-PCR recorded smaller fold
287 changes in comparison to those determined with mRNA-Seq due to known limitations (e.g.,
288 saturation and detection range limits) of ethidium bromide staining of gels and quantitation.

289

290 **Discussion**

291 Transcriptome profiling by mRNA-Seq is fast becoming an attractive method as it facilitates
292 rapid generation of large datasets for transcript identification and quantification, even in the
293 absence of a reference genome. In this work, using just 2 lanes of an Illumina Genome
294 Analyser flow cell, over 50 million 75 nt reads were generated, amounting to 3.56 Gbp after
295 quality trimming. In comparison, Genbank contained 525,999 capillary-sequenced barley
296 ESTs, amounting to 272.6 Mbp (November 2011). It thus appears that mRNA-Seq has a
297 profound potential for plant biology, as also indicated by recent studies on crop species such
298 as rice (Mizuno et al. 2010) and soybean (Severin et al. 2010) with reference genomes, and
299 chickpea (Garg et al. 2011) which assumed no a-priori sequence information.

300

301 In this work, we aimed at addressing two main points; to determine barley genes which are
302 differentially expressed under acute salt stress, and to discover previously un-identified
303 transcripts in the barley leaf. To this end, we show that the two-phase assembly method
304 significantly improved the contiguity over the standard assembly, with k55 N50 increasing
305 from 223 bp to 518 bp. Furthermore, the use of stringent blast ratio filtering enabled the
306 discovery of 3,828 NTCs, some of which are new and others are more complete in size.
307 While many of the differentially regulated genes identified here agree with previous work
308 using microarrays (Ueda et al. 2004), differentially expressed NTCs found here represent
309 transcripts which could not have been detected using microarray technology. One such NTC
310 is a 2.0 kb transcript encoding a wall-associated receptor kinase-like 22, which has only a 450
311 bp blast hit in the current barley Unigene database but a 1.2 kbp match to rice and
312 Brachypodium homologues. A related gene in Arabidopsis, WAKL4, is responsive to Na+ as
313 well as other cations such as K^+, Ni^{2+} and Zn^{2+} (Hou et al. 2005).
314
315 An early-responsive to dehydration stress (ERD4) homolog, known as late embryogenesis
316 abundant (LEA), is strongly up-regulated by acute salinity in this dataset and has been
317 investigated in maize (Liu et al. 2009), wherein this gene is not only induced upon salinity
318 stress, but its over-expression in Arabidopsis leads to enhanced tolerance to drought and
319 salinity. In barley, this gene is shown to confer tolerance to osmotic stresses (Xu et al. 1996).
320 Other strongly upregulated candidates for future functional work could include the
321 chloroplast localised lipoxygenase 2.1, a plasma membrane bound ATPase, an ODORANT1
322 homologue, as well as a protein phosphatase 2C. A highly expressed aquaporin was among
323 the most decreased in expression (ranked 121 highest in control, down to 3,016 in salt),
324 indicating water transport processes within the leaf could be strongly reduced under acute salt
325 stress.
326
327 Comparison of this mRNA-Seq dataset to previously described array experiments (Walia et
328 al. 2007) yielded a moderate correlation of fold change, with 5,334 of the 15,000 highest
329 expressed transcripts showing contradictory fold changes (data not shown). Many factors
330 could explain this disparity, including the different lines of barley used, different regimens of
331 salt stress and different analysis chemistries. Investigation of 11 transcripts with sqRT-PCR
332 shows general agreement between the two techniques (Fig. S1). While the direction of

333 expression changes was largely consistent between both methods, the magnitude of fold
334 change found with sqRT-PCR was generally smaller than that of mRNA-Seq (Table S1).
335

336 The data demonstrate that mRNA-Seq is an excellent high-throughput methodology for gene
337 expression, which will be crucial to revealing the scale of variations in barley germ-plasm
338 and accurate mapping of quantitative trait loci. As next-generation sequencing technologies
339 and associated bioinformatics methods continue to improve, these will become more
340 commonplace in plant biology will result in a comprehensive high quality annotation of the
341 barley and wheat genomes. Until then, this study provides a valuable dataset containing
342 thousands of novel transcripts and a snapshot of differential expression due to acute salt
343 stress. The outcomes serve as a useful reference for future hypothesis-driven studies in
344 barley and the closely related and most important cereal, wheat. Reverse genetic studies of
345 these salinity responsive genes could uncover genes which contribute to salinity tolerance.
346

347 **Author Contributions Statement**

348 Mark Ziemann performed mRNA-Seq and bioinformatics, generated figures and co-wrote the
349 manuscript. Atul Kamboj and Runyararo M. Hove prepared plant material, undertook qPCR,
350 generated figures and edited the manuscript. Shanon Loveridge assisted with expert
351 bioinformatics analysis. Assam El-Osta co-wrote and edited the manuscript. Mrinal Bhave
352 undertook experimental planning and co-wrote and edited the manuscript.
353

354 **References**

355
356 Ando K, Grumet R (2010) Transcriptional profiling of rapidly growing cucumber fruit by
357 454-pyrosequencing analysis. J Amer Soc Hort Sci 135:291-302
358

359 Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome
360 Biol 11: R106
361

362 Bartels D, Sunkar R (2005) Drought and salt tolerance in plants. Crit Rev Plant Sci 24:23-58
363

364 Close TJ (1996) Dehydrins: emergence of a biochemical role of a family of plant dehydration
365 proteins. Physiol Plant 97:795-803
366

367 Druka A, Muehlbauer G, Druka I, Caldo R, Baumann U, Rostoks N, Schreiber A, Wise R,
368 Close T, Kleinhofs A, Graner A, Schulman A, Langridge P, Sato K, Hayes P, McNicol J,
369 Marshall D, Waugh R (2006) An atlas of gene expression from seed to seed through
370 barley development. Funct Integr Genom 6:202-211
371

372 Du Z, Zhou X, Ling Y, Zhang Z, Su Z (2010) agriGO: a GO analysis toolkit for the
373 agricultural community. Nucleic Acids Res 38 (Suppl 2):w64-w70
374
375 Food and Agriculture Organisation of the United Nations (2005) Management of irrigation-
376 induced salt-affected soils ftp://ftp.fao.org/agl/agll/docs/salinity_brochure_eng.pdf
377
378 Forrest KL, Bhave M (2007) Major intrinsic proteins (MIPs) in plants: a complex gene
379 family with major impacts on plant phenotype. Funct Integr Genom 7:263-289
380
381 Garg R, Patel RK, Tyagi AK, Jain M (2011) De novo assembly of chickpea transcriptome
382 using short reads for gene discovery and marker identification. DNA Res 18:53-63
383
384 Grains Research & Development Corporation (2008) GRDC Impact Assessment Report
385 Series: An Economic Analysis of GRDC's Investment in Barley Breeding
386 http://www.grdc.com.au/uploads/documents/GRDC_ImpAss_BarleyBreeding1.pdf
387
388 Hoagland DR, Arnon DI (1950) The water culture method for growing plants without soil.
389 University of California Agric. Exp station, Berkley Circular 347
390
391 Hou X, Tong H, Selby J, Dewitt J, Peng X, He ZH (2005) Involvement of a cell wall-
392 associated kinase, WAKL4, in Arabidopsis mineral responses. Plant Physiol 139:1704-
393 1716
394
395 International Barley Genome Sequencing Consortium, Mayer KF, Waugh R, Brown JW,
396 Schulman A, Langridge P, Platzer M, Fincher GB, Muehlbauer GJ, Sato K, Close TJ,
397 Wise RP, Stein N (2012) A physical, genetic and functional sequence assembly of the
398 barley genome. Nature 29:711-716
399
400 Kawaura K, Mochida K, Ogihara Y. (2008) Genome-wide analysis for identification of salt-
401 responsive genes in common wheat. Funct Integr Genomics 8:277-286
402
403 Langridge P, Paltridge N, Fincher G (2006) Functional genomics of abiotic stress tolerance in
404 cereals. Briefings Funct Genom Proteom 4:343-354
405
406 Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler
407 transform. Bioinforma 25:1754-1760
408
409 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin
410 R; 1000 Genome Project Data Processing Subgroup (2009) The Sequence
411 Alignment/Map format and SAMtools. Bioinforma 25:2078-2079
412
413 Liu YH, Li HY, Shi YS, Song YC, Wang TY, Li Y (2009) A maize early responsive to
414 dehydration gene, *ZmERD4,* provides enhanced drought and salt tolerance in
415 Arabidopsis. Plant Mol Biol Rep 27:542-548
416
417 Ma Q, Dai X, Xu Y, Guo J, Liu Y, Chen N, Xiao J, Zhang D, Xu Z, Zhang X, Chong K
418 (2009) Enhanced tolerance to chilling stress in OsMYB3R-2 transgenic rice is mediated

419 by alteration in cell cycle and ectopic expression of stress genes. Plant Physiol 150:244-
420 256
421
422 Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008) RNAseq: An assessment of
423 technical reproducibility and comparison with gene expression arrays. Genom Res
424 18:1509-1517
425
426 Meiri D, Breiman A (2009) Arabidopsis ROF1 (FKBP62) modulates thermotolerance by
427 interacting with HSP90.1 and affecting the accumulation of HsfA2-regulated sHSPs.
428 Plant J 59:387-399
429
430 Mizuno H, Kawahara Y, Sakai H, Kanamori H, Wakimoto H, Yamagata H, Oono Y, Wu J,
431 Ikawa H, Itoh T, Matsumoto T (2010) Massive parallel sequencing of mRNA in
432 identification of unannotated salinity stress-inducible transcripts in rice (*Oryza sativa* L).
433 BMC Genom 11:683-695
434
435 Modra Seeds Fact Sheet, February 2011
436 http://www.modraseeds.com.au/pdf/Hindmarsh_Factsheet_Feb_2011.pdf
437
438 Mohammadi M, Kav NNV, Deyholos MK (2007) Transcriptional profiling of hexaploid
439 wheat (*Triticum aestivum* L.) roots identifies novel, dehydration-responsive genes. Plant
440 Cell Environ 30:630-645
441
442 Munns R (2005) Genes and salt tolerance: bringing them together. New Phytol 167:645-663
443
444 Munns R, Tester M (2008) Mechanisms of salinity tolerance. Ann Rev Plant Biol 59:651-681
445
446 Rengasamy P (2006) World salinization with emphasis on Australia. J Expt Bot 57:1017-
447 1023
448
449 Schreiber AW, Sutton T, Caldo RA, Kalashyan E, Lovell B, Mayo G, Muehlbauer GJ, Druka
450 A, Waugh R, Wise RP, Langridge P, Baumann U (2009) Comparative transcriptomics in
451 the Triticeae. BMC Genom 10:285-301
452
453 Seki M, Narusaka M, Ishida J, Nanjo T, Fujita M, Oono Y, Kamiya A, Nakajima M, Enju A,
454 Sakurai T, Satou M, Akiyama K, Taji T, Yamaguchi-Shinozaki K, Carninci P, Kawai J,
455 Hayashizaki Y, Shinozaki K. (2002) Monitoring the expression profiles of 7000
456 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length
457 cDNA microarray. Plant J 31:279-292
458
459 Selote DS, Khanna-Chopra R (2006) Drought-acclimation confers oxidative stress tolerance
460 by inducing coordinated antioxidant defense at cellular and subcellular level in leaves of
461 wheat seedlings. Physiol Plant 127:494-506
462
463 Severin AJ, Woody JL, Bolon YT, Joseph B, Diers BW, Farmer AD, Muehlbauer GJ, Nelson
464 RT, Grant D, Specht JE, Graham MA, Cannon SB, May GD, Vance CP, Shoemaker RC

465 (2010) RNA-Seq atlas of *Glycine max*: a guide to the soybean transcriptome. BMC Plant
466 Biol 10:160-175
467
468 Shinozaki K, Yamaguchi-Shinozaki K (1997) Gene expression and signal transduction in
469 water-stress response. Plant Physiol 115:327-334
470
471 Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I (2009) ABySS: a parallel
472 assembler for short read sequence data. Genome Res 19:1117-1123
473
474 Suprunova T, Krugman T, Fahima T, Chen G, Shams I, Korol A, Nevo E (2004) Differential
475 expression of dehydrin genes in wild barley, *Hordeum spontaneum*, associated with
476 resistance to water deficit. Plant Cell Environ 27:1297-1308
477
478 Tyerman SD, Niemietz CM, Bramley H (2002) Plant aquaporins: multifunctional water and
479 solute channels with expanding roles. Plant Cell Environ 25:173-194
480
481 Urano K, Kurihara Y, Seki M, Shinozaki K (2010) 'Omics' analyses of regulatory networks
482 in plant abiotic stress responses. Curr Opin Plant Biol 13:132-138
483
484 Ueda A, Kathiresan A, Inada M, Narita Y, Nakamura T, Shi W, Takabe T, Bennett J (2004)
485 Osmotic stress in barley regulates expression of a different set of genes than salt stress
486 does. J Exp Bot 55:2213-2218
487
488 Walia H, Wilson C, Wahid A, Condamine P, Cui X, Close TJ (2006) Expression analysis of
489 barley (*Hordeum vulgare* L.) during salinity stress. Funct Integr Genom 6:143-156
490
491 Walia H, Wilson C, Condamine P, Ismail AM, Xu J, Cui XP, Close TJ (2007) Array-based
492 genotyping and expression analysis of barley cv. Maythorpe and Golden Promise. BMC
493 Genom 8:87-100
494
495 Xu D, Duan X, Wang B, Hong B, Ho T, Wu R (1996) Expression of a Late Embryogenesis
496 Abundant Protein Gene, HVA1, from Barley Confers Tolerance to Water Deficit and Salt
497 Stress in Transgenic Rice. Plant Physiol 110:249-257
498
499 Xu ZS, Xia LQ, Chen M, Cheng XG, Zhang RY, Li LC, Zhao YX, Lu Y, Ni ZY, Liu L, Qiu
500 ZG, Ma YZ (2007) Isolation and molecular characterization of the *Triticum aestivum* L.
501 ethylene-responsive factor1 (TaERF1) that increases multiple stress tolerance. Plant Mol
502 Biol 65:719-732

503 **Figure Legends**

504

505 **Fig. 1** Study design.

506 mRNA-Seq is performed on material derived from control and salt-stressed barley leaves.
507 Reads unaligned to current databases are assembled to discover novel sequences. These
508 sequences are blasted to rice and barley databases to determine novelty. The final contig set is
509 analysed for differential gene expression.

510

511 **Fig. 2** Process of identifying novel sequences from mRNA-Seq data.

512 (a) The schematic consolidation of data by assembly of reads, merging of overlapping
513 contigs, filtering by Blast ratio. (b) The number of identified contigs ≥100bp in phase one
514 assembly using k-mer 27 to 63. (c) Phase 1- average contig length. (d) Phase 1 - N50 length.
515 (e) Phase 2 - number of contigs ≥100bp. (f) Phase 2- average contig length. (g) Phase 2- N50
516 length. (h) Length distribution for final assembly using k55, as a k-mer of 55 was selected
517 for further analysis. (i) Blast ratio filtering: the Blast bit score of the best hit in the rice
518 database (y-axis) is plotted against the bit score of the best hit in the barley database (x-axis),
519 with points in red denoting contigs passing filtering (NTCs) and those in black being
520 discarded.

521

522 **Fig. 3** Differential gene expression of NTCs and known transcripts.

523 A smear plot (a) showing the base mean expression versus the Log2 of fold change for NTCs
524 shown in red and known contigs shown in black. Large points indicate those considered
525 significantly differentially regulated by salt stress (adj p-value < 0.05). A distribution of p-
526 values for NTCs and known transcripts (b).

Table 1 mRNA-Seq data yields from Genome Analyzer IIx sequencing

	Control dataset	Salt stress dataset
Original read length (bp)	76	76
Original number of reads (million)	23.7	26.7
Original sequence yield (Gbp)	1.80	2.03
Sequence yield after artifact filtering (Gbp)	1.80	2.03
Number of reads after Q30 quality filtering (million)	23.4	26.1
Sequence yield after Q30 quality filtering (bp)	1.70	1.86
Number of reads aligning to Unigene DB (million)	16.6	16.4
% Reads aligned	70.9	63.0
Number of unaligned reads (million)	6.79	9.64
Unaligned sequence (Gbp)	0.496	0.698

Table 2 Results from the two-phase assembly using AbySS

Phase1 of Assembly	
Number of unmerged contigs from k27-k63 assembly	5,723,131
Number of unmerged contigs ≥100bp	954,420
Average length (bp)	235
N50 Length (bp)	256
Longest contig (bp)	12,314
Phase2 of Assembly (k-mer =55)	
Number of merged contigs	50,499
Number of merged contigs ≥100bp	39,707
Average length (bp)	344
N50 Length (bp)	518
Longest contig (bp)	13,710
Assembly size (bp)	13,696,077
Contigs with Rice/Barley Blast Ratio ≥2	3,828

Table 3 Alignment results post-assembly

Parameter	Control sample	Salt stress sample
Post-process read length (bp)	36	36
Post-process number of reads (million)	23.4	26.1
Number of reads aligning to Improved DB (million)	19.9	22.1
% Improvement on first alignment	20.1	34.4

Table 4 Top 20 up-regulated and down-regulated transcripts ranked by p-value

Accession	Reads Control	Reads Salt	Log2 Fold Change	P-value (FDR adjusted)	Nearest rice blast hit*	Rice_annotation
Hv.29838	20	827	5.3	1.12E-5	LOC_Os07g36750	CSLF3 - cellulose synthase-like family F; beta1,3;1,4 glucan synthase, expressed
Hv.5008	125	3701	4.9	7.76E-5	-	unclassified transcript
Hv.31363	4	224	5.8	2.82E-4	LOC_Os12g37260	lipoxygenase 2.1, chloroplast precursor, putative, expressed
Hv.8934	29	640	4.4	3.40E-4	LOC_Os04g40990	malate synthase, glyoxysomal, putative, expressed
Hv.17368	44	843	4.2	0.001	LOC_Os05g46040	protein phosphatase 2C, putative, expressed
Hv.2654	9	261	4.8	0.001	LOC_Os07g44060	haloacid dehalogenase-like hydrolase family protein, putative, expressed
Hv.29473	10	264	4.7	0.002	-	unclassified transcript
Hv.3400	2	137	6.1	0.002	LOC_Os01g12580	late embryogenesis abundant protein, putative, expressed
Hv.30848	3	149	5.6	0.002	LOC_Os03g48310	plasma membrane ATPase, putative, expressed
Hv.32578	4	164	5.3	0.002	LOC_Os04g02000	zinc finger family protein, putative, expressed
Hv.17120	40	603	3.9	0.003	LOC_Os10g41490	CAMK_CAMK_like.41 - CAMK includes calcium/calmodulin dependent protein kinases, expressed
Hv.15443	41	566	3.8	0.004	LOC_Os04g47700	expressed protein
NTC2548 2	58	754	3.7	0.004	LOC_Os02g42110	wall-associated receptor kinase-like 22 precursor, putative, expressed
Hv.32190	6	172	4.8	0.004	LOC_Os03g19600	retrotransposon protein, putative, unclassified, expressed
Hv.5085	11	229	4.4	0.005	LOC_Os07g05365	photosystem II 10 kDa polypeptide, chloroplast precursor, putative, expressed
Hv.30861	2	104	5.7	0.007	LOC_Os03g48310	plasma membrane ATPase, putative, expressed
Hv.10528	234	2825	3.6	0.008	LOC_Os09g35880	B-box zinc finger family protein, putative, expressed
Hv.5729	1	56	4.8	0.008	LOC_Os06g38294	peptide transporter PTR2, putative, expressed
NTC2618 5	3	230	4	0.009	LOC_Os09g25700	TsetseEP precursor, putative, expressed
Hv.12388	89	976	3.4	0.009	LOC_Os09g02180	expressed protein
Accession	Reads Control	Reads Salt	Log2 Fold Change	P-value (FDR adjusted)	Nearest rice blast hit*	Rice_annotation
Hv.16656	5229	25	-7.7	2.65E-9	LOC_Os12g31000	pumilio-family RNA binding repeat domain containing protein, expressed
Hv.2383	1413	10	-7.2	6.16E-9	LOC_Os10g25060	expressed protein
Hv.6975	798	5	-7.3	8.44E-9	LOC_Os04g47140	expressed protein
Hv.33010	382	2	-7.6	2.52E-7	LOC_Os09g31380	jmjC domain-containing protein 5, putative, expressed
Hv.13882	771	10	-6.3	2.52E-7	LOC_Os04g02880	expressed protein
Hv.10251	950	14	-6.1	2.52E-7	LOC_Os03g58300	indole-3-glycerol phosphate lyase, chloroplast precursor, putative, expressed
Hv.34103	230	1	-7.9	1.13E-5	LOC_Os03g08580	expressed protein
Hv.37409	1475	43	-5.1	1.13E-5	LOC_Os04g57880	heat shock protein DnaJ, putative, expressed
Hv.20312	1873	57	-5.1	1.73E-5	LOC_Os01g74020	MYB family transcription factor, putative, expressed
Hv.30597	712	22	-4.8	6.27E-5	LOC_Os01g05060	mitochondrial glycoprotein, putative, expressed
Hv.13356	234	3	-6.3	6.30E-5	LOC_Os04g49450	MYB family transcription factor, putative, expressed

Hv.19411	4919	160	-5	7.65E-5	LOC_Os06g19444	CCT/B-box zinc finger protein, putative, expressed
Hv.9005	1081	51	-4.4	2.31E-4	LOC_Os03g55280	semialdehyde dehydrogenase, NAD binding domain containing protein, putative, expressed
Hv.8557	391	14	-4.8	2.82E-4	LOC_Os03g16780	ankyrin repeat family protein, putative, expressed
Hv.20948	7350	295	-4.6	3.49E-4	LOC_Os05g37520	expressed protein
Hv.19979	934	50	-4.2	4.90E-4	LOC_Os07g42650	expressed protein
Hv.19759	1439	82	-4.2	0.001	LOC_Os02g40510	response regulator receiver domain containing protein, expressed
Hv.30983	2304	135	-4.1	0.001	LOC_Os03g63910	PPR repeat domain containing protein, putative, expressed
Hv.8625	214	5	-5	0.001	LOC_Os07g48050	peroxidase precursor, putative, expressed
Hv.21993	270	11	-4.6	0.002	LOC_Os12g43600	RNA recognition motif containing protein, expressed

*Annotations were mined from best Blastn hits in the Rice database using an e-value threshold of <0.1

Figure 1. Study design

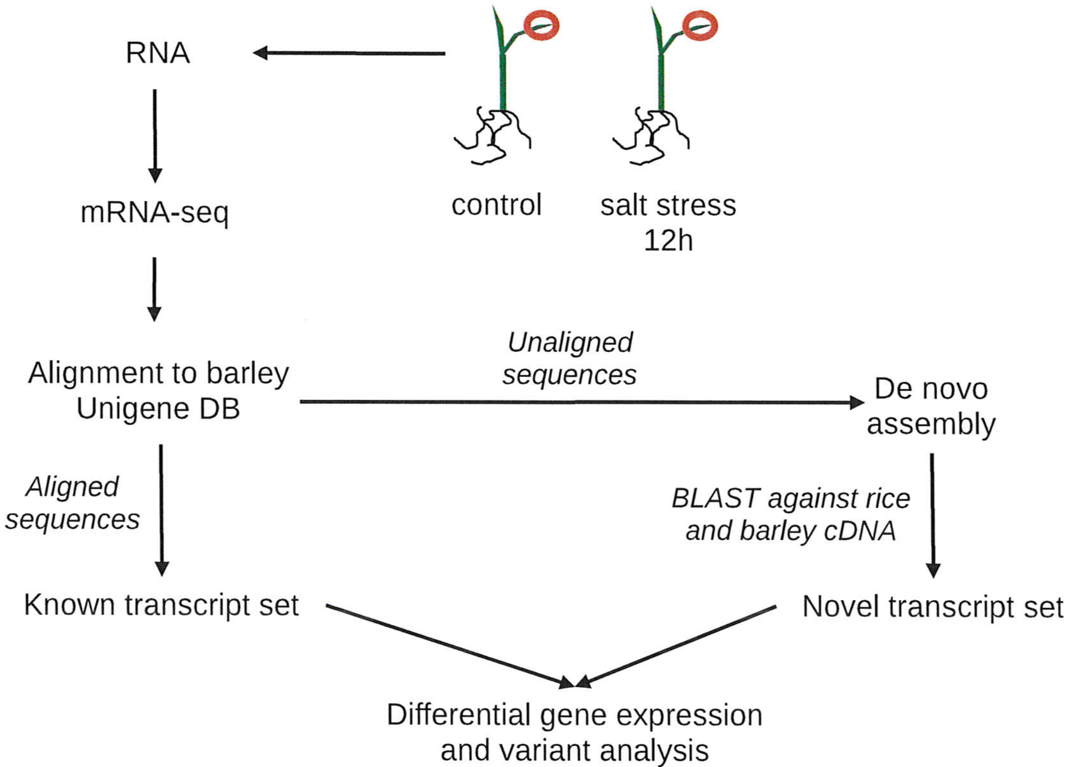

Figure 2. Process of identifying novel sequences from mRNA-Seq data.

Figure 2 (a)

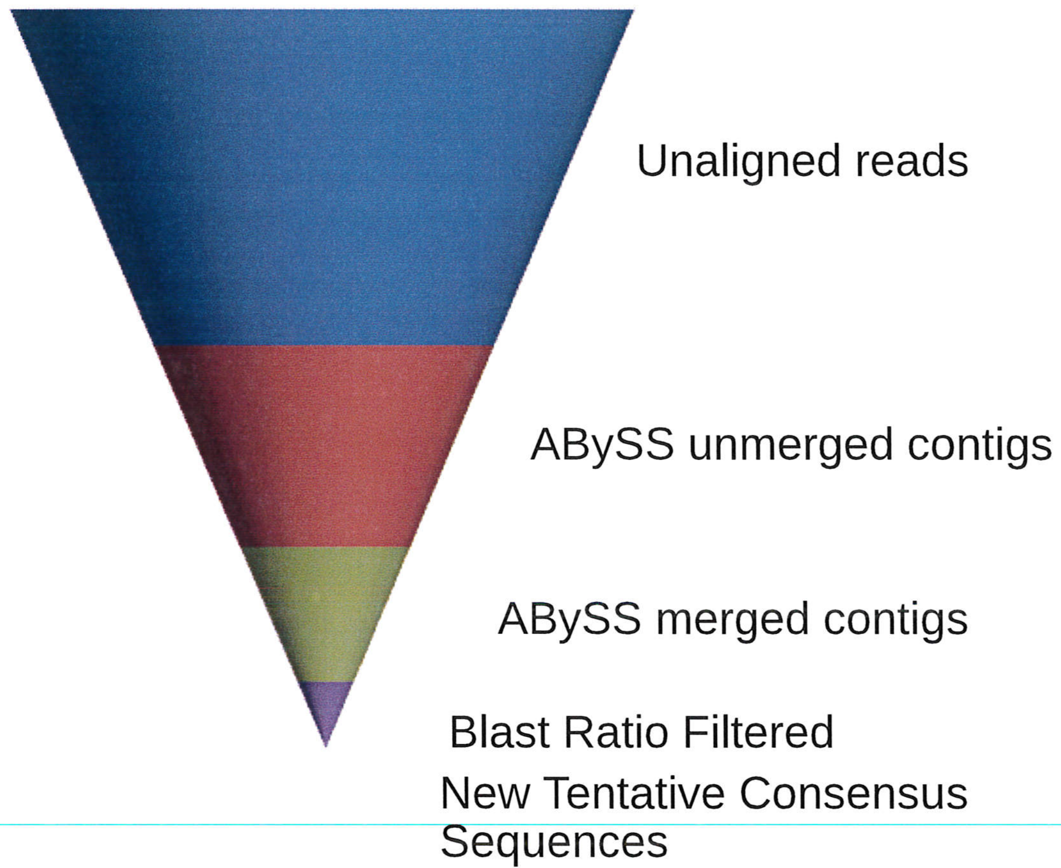

Figure 2(b)

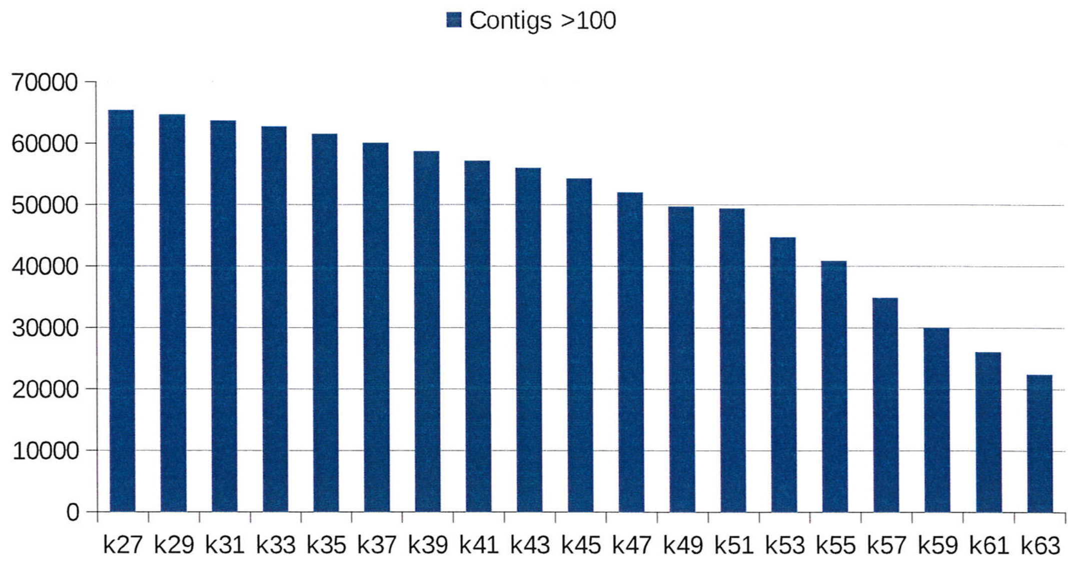

Figure 2(c)

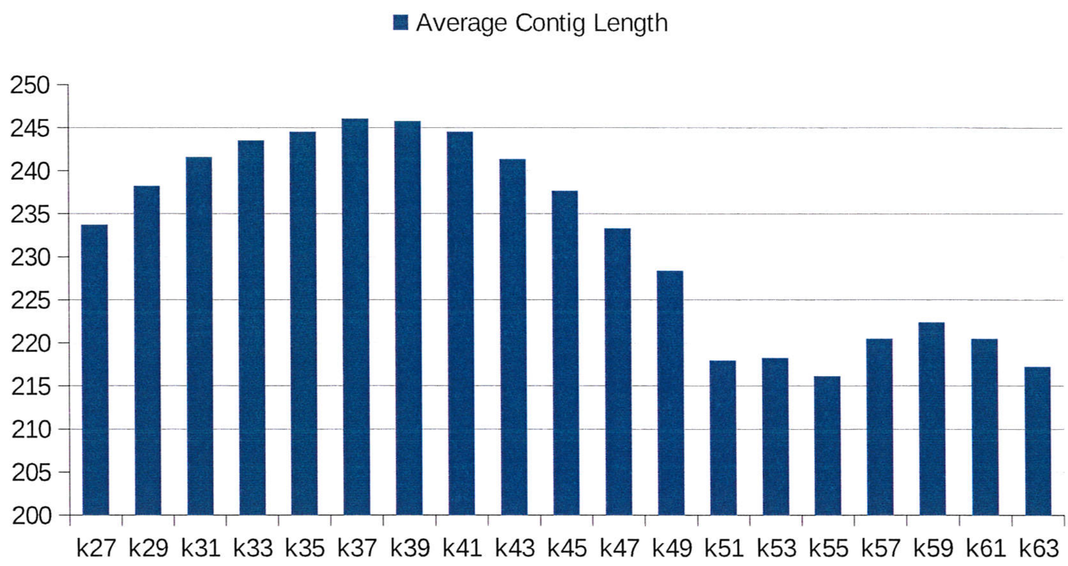

Figure 2(d)

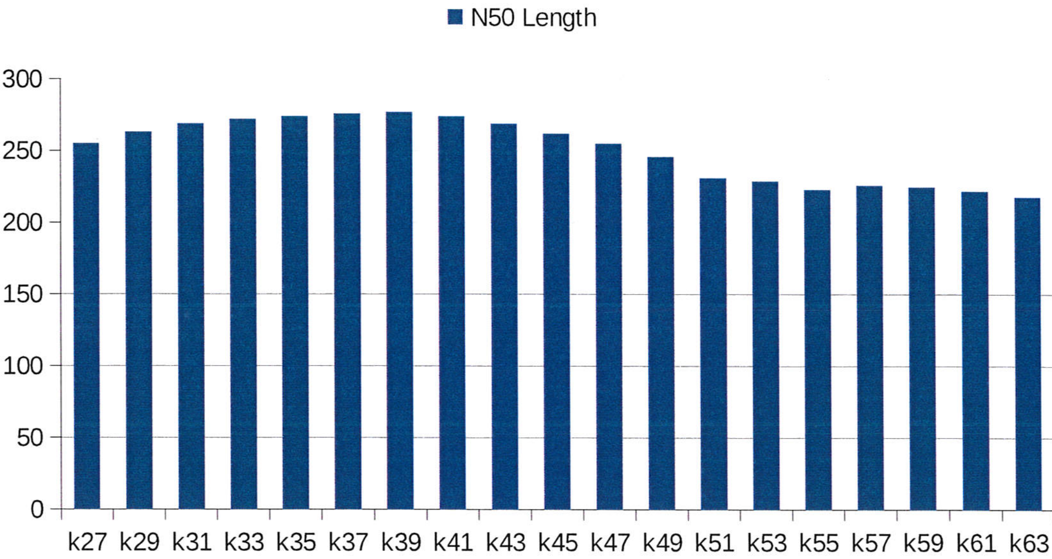

Figure 2(e)

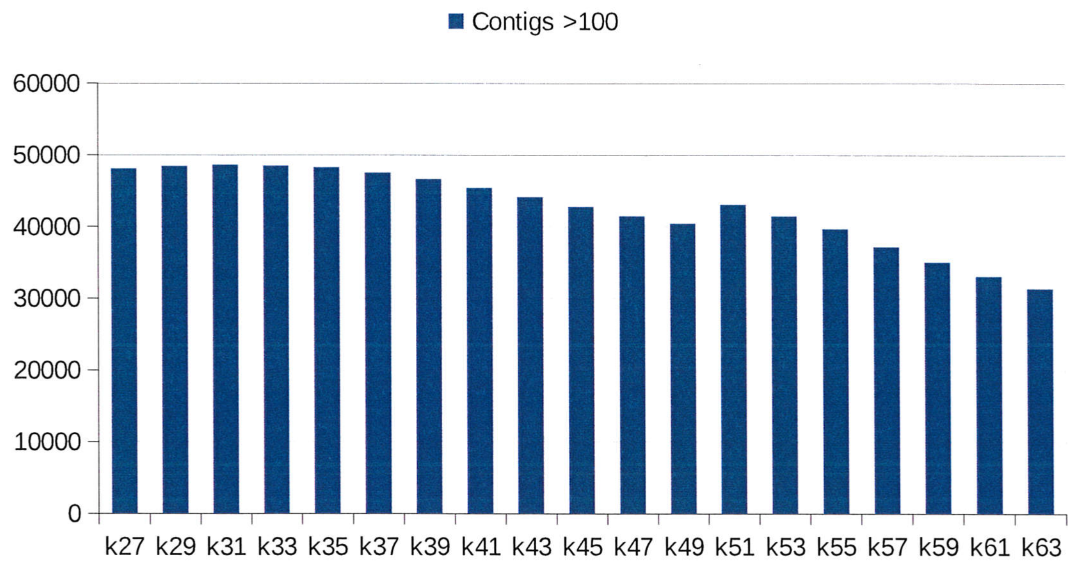

Figure 2(f)

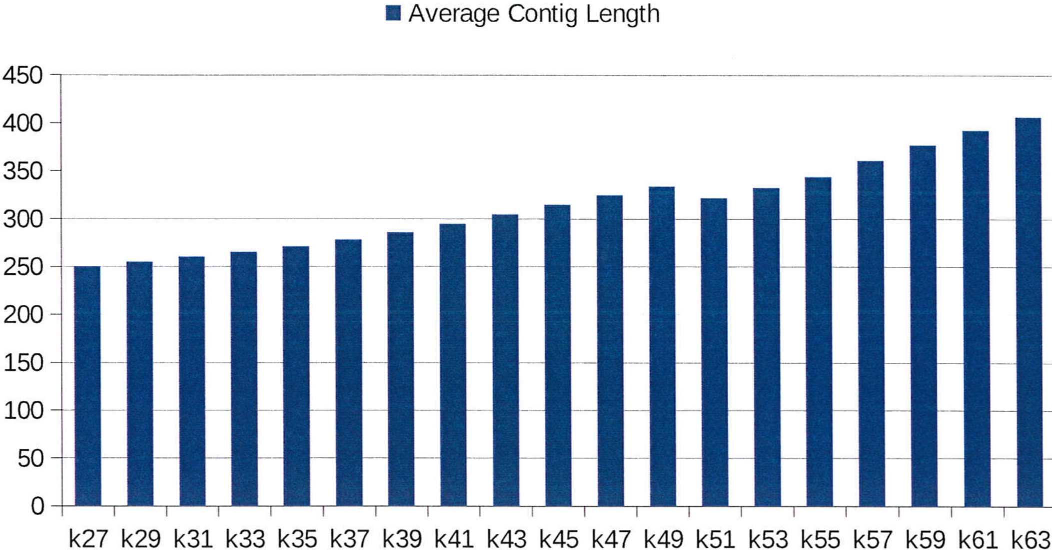

Figure 2(g)

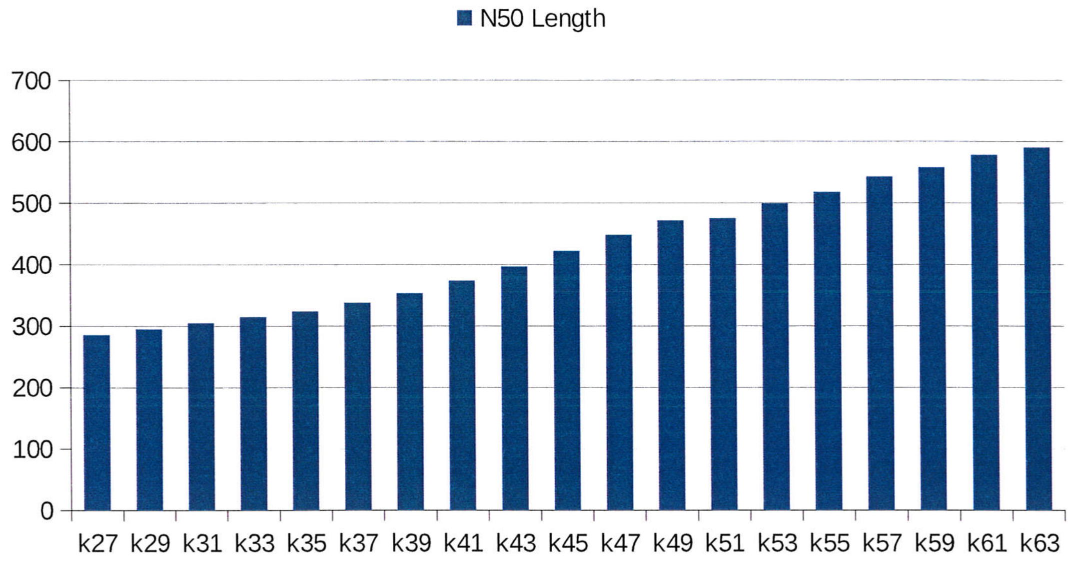

Figure 2(h)

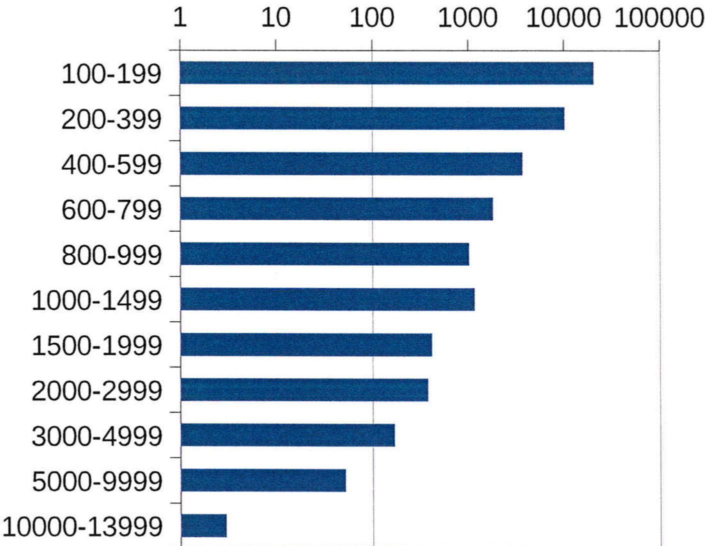

Figure 2(i)

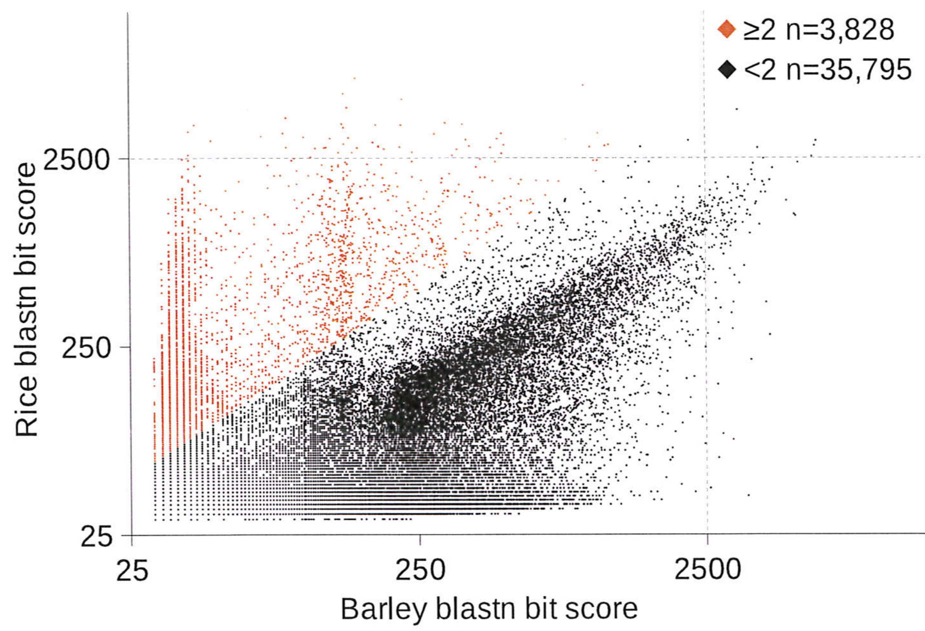

Figure 3. Differential gene expression of NTCs and known transcripts.
Figure 3(a) Comparison of control and salt gene expression

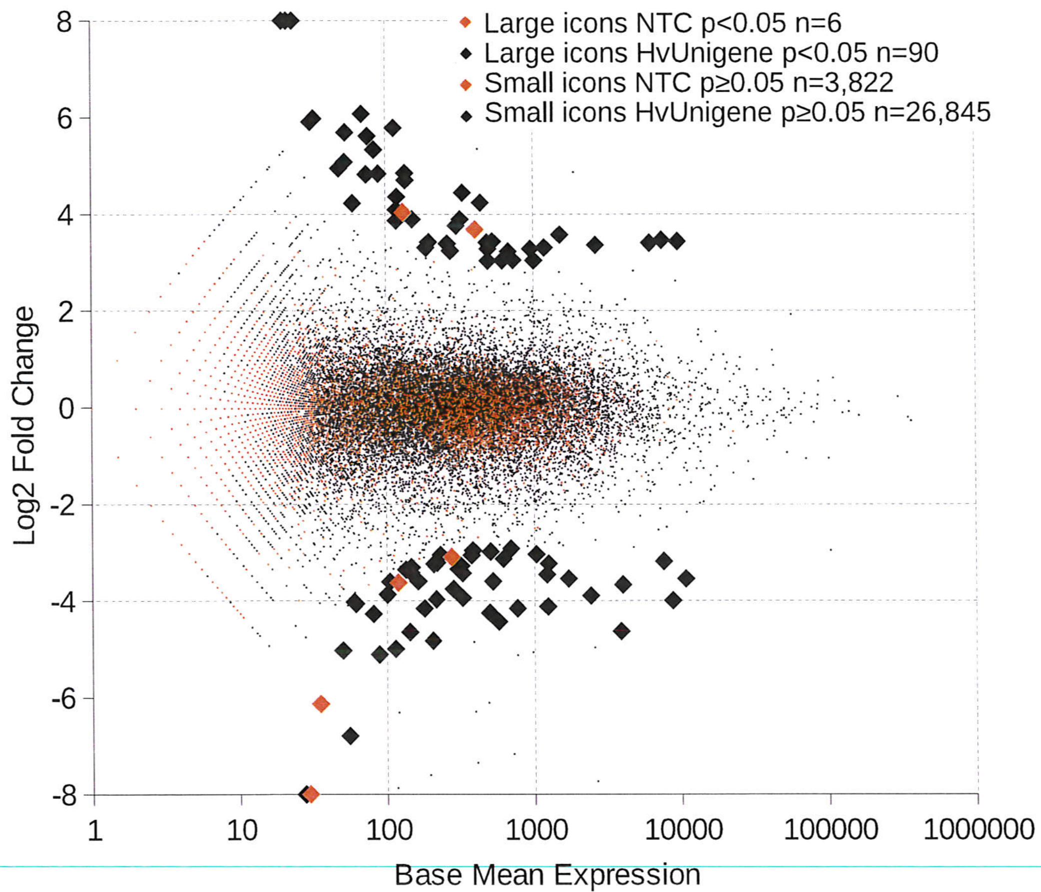

Figure 3(b) Distribution of p-values