High temporal resolution of gene expression dynamics in developing mouse embryonic stem cells

Brian S. Gloss[1,2], Bethany Signal[1,2], Seth W. Cheetham[3], Franziska Gruhl[4], Dominik Kaczorowski[1], Andrew C. Perkins[5], Marcel E. Dinger[1,2,*]

[1] Garvan Institute of Medical Research, Sydney, Australia

[2] St Vincents Clinical School, Faculty of Medicine, UNSW Australia, Sydney, Australia

[3] The Gurdon Institute and Department of Physiology, Development, and Neuroscience, University of Cambridge, Cambridge, United Kingdom

[4] Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland

[5] Mater–UQ Research Institute, The University of Queensland, Translational Research Institute, Brisbane, Australia

[*]Corresponding Author: Marcel E. Dinger; tel: +61 2 9355 5860; fax: +61 2 9295 8101, email: m.dinger@garvan.org.au

Abstract

Investigations of transcriptional responses during developmental transitions typically use time courses with intervals that are not commensurate with the timescales of known biological processes. Moreover, such experiments typically focus on protein-coding transcripts, ignoring the important impact of long noncoding RNAs. We evaluated coding and noncoding expression dynamics at high temporal resolution (6-hourly) in differentiating mouse embryonic stem cells and report the effects of increased temporal resolution on the characterization of the underlying molecular processes. We present a refined resolution of global transcriptional alterations, including regulatory network interactions, coding and noncoding gene expression changes as well as alternative splicing events, many of which cannot be resolved by existing coarse developmental time-courses. We describe novel short lived and cycling patterns of gene expression and temporally dissect ordered gene expression at bidirectional promoters and responses to transcription factors. These findings demonstrate the importance of temporal resolution for understanding gene interactions in mammalian systems.

Links to data

Data has been deposited into GEO: The Reviewer access link is:
http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=cnglummejbkltyj&acc=GSE75028

Introduction

Over the past decade, transcriptomic investigations into the of nature embryonic stem cell (ESC) differentiation have elucidated key biochemical features of stemness and differentiation. Increasingly, it has become apparent that understanding the dynamics and coordination of gene expression signatures over time during the key phases of differentiation is critical to adequate characterization of fundamental biological processes.

ESC differentiation in mouse is a highly complex cascade of gene expression changes that allow single pluripotent cells in culture to progress to an organoid resembling a pre-implantation blastocyst within only five days. The spontaneous differentiation of these cells in culture has provided key insights into the developmental processes underlying the generation of the primary germ cell layers(Martello and Smith 2014). Microarray and RNA sequencing have provided a means to characterize the molecular transitions in gene expression underlying ESC biology and more recently single cell transcriptomic studies have provided the first glimpses into the molecular history of these cells(Liu et al. 2014). However, it is clear that much more of the transcriptional landscape of ESC remains to be elucidated(Rosa and Brivanlou 2013).

Access to new technologies, such as massively parallel sequencing (MPS), has led to a dramatic increase in our knowledge of the mammalian transcriptome. Early genomic tiling array analysis indicated that most of the genome was transcribed into RNA(Bertone et al. 2004). MPS of the transcriptome validated this observation and revealed that the majority of the mammalian genome is pervasively transcribed as interlaced and overlapping RNAs(Djebali et al. 2012), many of which lack protein-coding potential(Guttman et al. 2009). The large number of long-noncoding transcripts (lncRNA) has become the focus of significant interest due to their exquisite cell type specific expression(Mercer et al. 2008), potent biological function (Bonasio and Shiekhattar 2014; Fatica and Bozzoni 2014), and rapid transactivation of cellular processes. However, in general, lncRNAs are lowly expressed and short lived(Clark et al. 2012), possibly because, unlike mRNAs that require translation, are able to exert their function directly. These qualities obfuscate their identification and characterization with traditional approaches that are tuned to the properties of mRNAs. Owing to the relative infancy of the field, the vast majority of noncoding transcripts are of unknown function(Quek et al. 2015). Additionally, the expression patterns of

these genes imply that their function is dependent on cellular context and likely regulatory(Bonasio and Shiekhattar 2014), thus the identification of these molecules and the context in which they act remains a research priority(Gloss 2015).

Various expression profiling studies, using both microarrays and RNA-seq(Bruce et al. 2007a; Cloonan et al. 2008; Dinger et al. 2008; Bergmann et al. 2015), have been used to explore the molecular changes occurring during ES cell development, typically at 24-hourly or more. This potentially has lead to incomplete gene expression relationships through the phenomenon of temporal aggregation bias whereby each time point is assumed to represent all the signaling changes occurring in that time window (Bay et al. 2004). In contrast to single cell based approaches- which provide insight into the state of individual cells - examinations of whole cell populations provides system-wide behavior and a practical means to explore gene expression dynamics across time. The combination of these techniques has recently shed light the molecular framework of cellular differentiation (Chu et al. 2016). Higher temporal resolution has also shown rapid induction (within two hours of retinoic acid stimulation) of lncRNAs associated with the HOX locus (De Kumar et al. 2015). Furthermore, high temporal resolution has provided valuable insights into transcriptional annotation and regulation in drosophila (Arbeitman et al. 2002; mod et al. 2010), Xenopus (Tan et al. 2013) and C.elegans (Boeck et al. 2016).

Here we show that additional temporal resolution of the global transcriptome in spontaneously differentiating mESC cells following LIF withdrawal enables the capture of the rapid and complex dynamic regulatory and noncoding changes occurring during ES development. We analyzed the transcriptome of differentiating mouse ESCs at six-hourly intervals over a five-day period, over which time the three primordial germ layers are specified. Using this fine-resolution temporal sampling approach, we identify significant transitions in the transcriptome and large-scale shifts in observable transcription factor activities that could not be observed at 24 hourly sampling periods. Moreover, we identify entirely novel coding and noncoding transcripts that are expressed only within specific sub-24-hour window. By leveraging the high sampling frequency of the data, we are able to both accurately recapitulate known regulatory cascades in ES development and predict and refine others. Finally, using correlative approaches, we can infer

77 functions for uncharacterized lncRNAs and predict the regulatory centers across the genome that
78 coordinate early development.

79 Results

80 The dynamic transcriptome of mESC differentiation at high temporal resolution

81 A median 42-million, paired-end 100-bp reads (Supplementary Figure S1A) were mapped from
82 stranded, poly-A derived cDNA libraries derived from biological duplicate, six-hourly time
83 courses of mESC differentiation over five days where key differentiation programs occur (0-120
84 hours, Figure 1A). Transcript-level expression data was generated as previously described
85 (Anders et al. 2013), then normalized for library size and transformed for data visualization and
86 differential gene expression analysis. Evaluation of 24 hourly time points indicated that our data
87 was comparable to previously published data in a similar model (Hirst et al. 2006)
88 (Supplementary Figure S1B). An interactive gene expression portal was created to visualise this
89 data (https://betsig.shinyapps.io/paper_plots).

90 To assess the reproducibility and provide confidence in the biological validity of the global
91 transcriptome trends, a principle components analysis (PCA) was performed on the 2,000 most
92 variable genes (Figure 1B). This analysis indicated that biological replicates clustered closely,
93 indicating that synchrony was retained, and that the major contributor to the determination of
94 variance was explained by time. Deconvolution of the dimensions yielded time-dependent
95 expression (in the first dimension) of genes enriched in focal adhesion/ ECM interactions KEGG
96 pathways. Interestingly, the second dimension deconvolution (PC2), in which undifferentiated
97 ESCs resemble the more differentiated embryoblast, yielded genes enriched in MAPK-signaling
98 and cancer pathways, implying that the process of differentiation involves a partial retention of a
99 cells capacity for self renewal. In the third component (PC3), in which the undifferentiated ES cell
100 is separate, the axon-guidance pathway was enriched. We then evaluated expression patterns of
101 genes associated with pluripotency, primitive streak formation and cell specialization (Figure
102 1C). We observed that, although the gene expression patterns were broadly consistent with
103 published studies (Supplementary Figure S1B), there were changes in expression on less than 24
104 hourly timeframes that could not be attributed to measurement biases (within the top 5% of

105 deviation from loess-smoothed expression values). To establish how prevalent sub-24 hour gene
106 expression changes were in in the transcriptome of developing ESC;s, we evaluated the extent to
107 which gene expression patterns observed 24 hourly were unable to capture gene expression
108 changes happening within that window (temporal aggregation bias (Bay et al. 2004)). We
109 observed that, compared to 24 hour time points, 417 more genes had counts data considered
110 sufficient for differential gene expression analysis; reflecting a substantial increase in detected
111 noncoding genes over protein coding (>12% vs. 2% respectively, chi-squared p value <0.001,
112 Supplementary Table S1, Supplementary Figure S1C). Furthermore, the additional time points
113 allowed the assembly of 58% more novel multiexonic intergenic, antisense and intronic
114 noncoding RNAs from the data - indicating that a substantial proportion of noncoding transcripts
115 are present on timescales much shorter than 24 hours. Finally, to ensure that the 6-hourly
116 measures represented distinct gene expression patterns to the 24-hourly measures, we observed
117 that no single 24-hourly measure was representative of the average expression over that day
118 (Mann-Whitney U p. adj. <1E-145) and that more than 1,000 genes displayed a more than 2-fold
119 difference mostly in the first 24 hours of differentiation (Supplementary Figure 1D-E). These
120 results indicate that enhanced temporal resolution reduces the phenomenon of temporal
121 aggregation bias and allows the observation of more distinct cell expression states than typical
122 time-courses..

123 An improved signaling cascade described by higher temporal resolution

124 Increased sampling frequency can provide a powerful insight into understanding of the
125 contribution of gene regulatory networks to cellular differentiation (mod et al. 2010). We utilized
126 the DREM v2 analysis tool (Schulz et al. 2012) to evaluate transcription factor (TF) target gene
127 expression patterns. Divergence of gene targets responsive to groups of TF at each time point,
128 either 24-hourly or 6-hourly (Figure 2A-B) was shown if the overall difference was significant at
129 p<0.001. Compared to 24-hourly, the observed complexity was significantly higher, especially in
130 the first 48 hours. We observed that significant changes in gene regulation occurred continuously
131 within the 24-hour windows. Most notably, first 24 hours following depart from pluripotency
132 resembles an ordered cascade of TF activity (Figure 2A, Supplementary Figure S2A) with large-
133 scale changes in TF activity at 12, 18 and 24 hours; of which little can be deduced measuring at

134 just 24 hours (Figure 2B, Supplementary Figure S2B). Focusing on the interplay between two key
135 transcription factors (OTX2 and POU5F1/OCT4(Yang et al. 2014), Figure 2A), we observed a
136 rapid rise in OTX2 activity in the first six hours and stable POU5F1 activity for the first 24 hours
137 (Red Box). OTX2 activity did not coincide with mRNA expression of the factor itself (Figure 2C),
138 although previous studies have observed increased in OTX2 protein expression within 3-hours of
139 differentiation(Yang et al. 2014), however periodic drops in POU5F1 mRNA expression appeared
140 to coincide with decreases in Pou5f1 target genes, we calculated the time taken for POU5F1
141 expression to result in changes in highly positively correlated (r>0.8) target genes using a cross-
142 correlation approach similar to (Li et al. 2002). We then evaluated how these "delays" enriched
143 for certain Reactome pathways (Figure 2D). We found rapid effects for targets enriched for "gene
144 expression"- and a delayed effect on "cell cycle" pathways compared to a null distribution
145 produced by 500 random "target" selections (grey). These were similarly observed in the DREM
146 GO-term enrichment tool for Pou5F1 targets decreasing in expression at 42 (early- Transcription
147 Factor Activity) and 54 hours (late- Epithelial Proliferation; Figure 2A, Blue Box &
148 Supplementary Figure S2C) and associated with the decrease in Pou5F1 expression (Figure 2C,
149 Blue Box). Importantly, Pou5F1 mRNA and protein expression are temporally correlated (Yang et
150 al. 2014). This result implies that TF-target genes may be activated in an ordered- time
151 dependent fashion. To explore this more broadly, we evaluated other TF-target gene temporal
152 dynamics for other TFs that exhibited strong positive or negative correlations between the TF
153 and their target genes. We found evidence of highly structured TF-target expression patterns in
154 time for negatively correlated Pou5f1 and Suz12 targets, as well as positively correlated Nanog,
155 Myc, Sox2 and Suz12 targets (Supplementary Figure S3).
156 These observations of precise temporal ordering of transcriptional events emphasize the
157 importance of factoring time delays into understanding gene regulatory networks (Chen et al.
158 2014) and highlight the capacity of increased temporal resolution to directly identify –rather
159 than inference in most cross-correlation approaches -valuable new knowledge of regulator-
160 target gene interactions.

161 ## Increased temporal resolutions identifies genes with previously uncharacterized
162 expression patterns (Short-lived (slRNA) & Cycling (cycRNA))

163 Having established that the increased temporal resolution markedly improves the molecular
164 framework for evaluating the contribution of gene expression to ES differentiation, we next
165 sought to identify gene expression signatures previously unable to be resolved using lower
166 temporal resolution. For each 24-hour period, we identified genes that were differentially
167 expressed between 0 and 6, 12 and 18 hours but not between any 24-hourly measures
168 (Supplementary Figure S2D). We identified 1,135 genes with significant changes in gene
169 expression that were unchanged between any 24-hourly comparison (adjusted $p<0.0001$). Of
170 these, 354 were differentially expressed for more than half of the corresponding 24-hour
171 window, mostly in the first and last 24-hour periods. These genes were described as short-lived
172 RNAs (slRNAs). slRNA expression patterns over the first 24 hours of differentiation were found
173 to be positively correlated with the same time window of retinoic acid directed differentiation
174 (De Kumar et al. 2015) (Supplementary Figure S2E) implying that these genes may form part of
175 the early response to differentiation signals. K-means clustering and KEGG pathway analysis of
176 the expression profiles of these genes (Figure 2E) revealed enrichment in genes associated with
177 the spliceosome ($p=0.02$) dramatically decreasing in expression over the first 24 hours before
178 returning slowly to baseline. To examine whether this impacted gene-splicing patterns, we
179 employed a differential exon (DEX) analysis between consecutive six-hourly time points and
180 counted the number of genes displaying DEX usage (Figure 2E). Consistent with previous studies,
181 the alternate splicing was most highly associated with cell differentiation(Salomonis et al. 2010)
182 (Figure 2E). Increased temporal resolution has elucidated that these changes happen very
183 rapidly (majority of changes in the first six hours), and that slRNAs may be involved in
184 suppressing the alternate splicing of genes and limiting transcriptional plasticity.
185 Some slRNAs appeared to have periodic expression profiles. We thus sought to uncover periodic
186 expression patterns genome-wide, by applying a fast-Fourier transformation to our data (see
187 Methods). Periodogram analysis was utilized to ascertain the dominant cycling period for each
188 gene. We found 137 genes, which we termed cycling RNAs (cycRNAs), sharing the same
189 dominant cycling period of less than 36 hours in both biological replicate experiments

190 (Supplementary Table S2). Supporting the efficacy of the approach, we found *Clock*, which
191 encodes a key regulator of circadian rhythm in mammals, to have a period of 24.2 hours. We
192 identified 20 genes that displayed characteristics of both slRNAs and cycRNAs (Supplementary
193 Figure S2F), including *Ewsr1* and *Clk1*, involved in gene splicing(Paronetto et al. 2011; Liu et al.
194 2013) as well as five uncharacterized lncRNAs. Given the highly specific expression patterns in
195 this context, we propose these genes may similarly have roles in maintaining or establishing
196 biological rhythms. Together these investigations show that the augmented temporal resolution
197 approach provides access to gain insights from regulatory pathways by identifying transitions in
198 expression that would otherwise have remained hidden.

199 Increased temporal resolution gives insight into local gene regulation in the 200 genome

201 Evaluating gene transcription at high temporal resolution in a highly dynamic process such as ES
202 development, we anticipated that it might be feasible to dissect structural gene regulation within
203 a given locus. To explore this possibility, we examined expression arising from transcripts that
204 are oriented head-to-head as so-called bidirectional pairs (Trinklein et al. 2004; Yang and
205 Elnitski 2014). Interestingly, we observed that the antisense transcript for Evx1 (Figure 1C)
206 displayed a previously unobserved (Dinger et al. 2008) increase in expression in the first 24
207 hours after departure from pluripotency that was reflected in its paired protein coding gene Evx1
208 (Supplementary Figure S4A), highlighting the increased power of frequent sampling over time. In
209 total, we identified 1,251 gene pairs with bidirectional transcriptional start sites (TSS) within
210 2,000 bp and evaluated correlation coefficients across the time course, distance between TSS and
211 median expression values. Consistent with other studies, we found expression correlation more
212 positive for bidirectional gene pairs than random transcript pairs(Trinklein et al. 2004)
213 (Supplementary Figure S4B). We were also able to show that the distance between TSS of highly
214 correlated bidirectional gene promoters is typically less than 500 bp (Figure 3A), consistent with
215 a common regulatory domain. Highly correlated or anti-correlated genes pairs displayed
216 differences in total gene expression, particularly with discordant gene biotypes (Figure 3B). We
217 found that protein coding gene pairs were more likely to be of similar expression levels and
218 positively correlated ($p<0.05$) than protein coding/noncoding pairs (Supplementary Figure S4C).

219 Applying a variant of the temporal offset analysis used to measure TF- gene target delays, we
220 calculated the time taken and defined the apparent driver gene type for peak correlation in
221 coding/noncoding bidirectional pairs (Supplementary Figure S4D, E). This did not reveal a
222 generalized bias in either time taken or particular "driving" gene type. However, this approach
223 shows that the lncRNA HOTAIRM1, required for activation of HOXA1(Zhang et al. 2009), appears
224 to have a six-hour delay between its expression changes and HoxA1. We present evidence of
225 other examples of lncRNA-led expression of protein coding genes in small numbers of
226 bidirectional pairs (Supplementary Figure S5).

227 To investigate whether the strong correlative potential between gene pairs could facilitate the
228 identification of regions of the genome that are coordinately regulated (Lercher et al. 2002), we
229 scanned across the genome for regions containing five or more contiguous genes that were
230 coexpressed (r>0.5). This revealed 59 regions with a mean size of 821 kb -each containing 5-14
231 genes (mean of 6) genes. The majority of these regions were each contained within a single
232 topological associated domain(Dixon et al. 2012) (Supplementary Figure S4F), increasing the
233 propensity for a common regulatory architecture. Evaluation of gene-expression patterns across
234 these clusters revealed evidence of high co-expression at both the inter- and intra-chromosomal
235 levels (Supplementary Figure S4G). We assembled a map of regions of the mouse genome
236 displaying high levels of clustered co-expression (Figure 3C) by comparing the expression
237 profiles of the regions. Two independent modules were identified with distinct decreasing
238 (green)- and increasing (blue) expression patterns with differentiation. Given the independent
239 location and expression patterns of these clusters, we suggest these regions may form core
240 expression-factories of cellular differentiation. In support of this notion, this analysis identified
241 the gene cluster -associated with the "increasing module"- containing the imprinting locus of
242 H19, IGF2, Tnn3 and Mrpl23(Kaffer et al. 2001) (Supplementary Figure S4H); previously shown
243 to be activated in concert during early stem cell differentiation(Poirier et al. 1991).

244 These investigations illustrate how analysis of high-resolution temporal transcriptomic data
245 provides an independent and convenient approach (relying only RNA-Seq) to guide the
246 partitioning of the genome into regulatory domains.

247 ## Increased temporal resolution refines the noncoding landscape of mESC
248 differentiation

249 Having shown that rapid changes in lncRNAs are a key feature of ES differentiation, and that co-
250 expression analysis is a powerful tool for understanding gene regulation with augmented
251 temporal resolution, we sought to unravel the roles that lncRNAs might play in ES differentiation.

252 Analysis of gene annotations yielded confident expression data for 588 lncRNA genes at six-
253 hourly resolution (520 for 24-hourly, Supplementary Table S1). Indeed, added temporal
254 resolution increased information of all noncoding transcript biotypes indicating that a
255 proportion of these genes were only present for a short duration in this system. Clustering
256 lncRNA expression patterns with time-dependent protein coding gene expression showed that
257 lncRNAs were enriched at lower expression levels and shared related expression profiles to
258 protein coding genes (Figure 4A). This relationship was further examined whereby K-means
259 clustering of these expression profiles compared to clustering of a similar number of time-
260 dependent protein coding genes (Figure 4B, Supplementary Figure S6A) revealed clusters of
261 lncRNA genes resembling gene expression patterns associated with stemness (cluster a)
262 primitive streak formation (cluster b) and WNT signaling (cluster c)(Dinger et al. 2008).
263 Determining the role that these lncRNAs play in these processes will be important in
264 understanding the molecular events underlying cell differentiation.

265 As lncRNAs often exert their function through guiding or assembling transcriptional machinery,
266 we sought to identify potential regulatory lncRNAs in this system. We selected 50 highly or
267 variably expressed lncRNAs (Figure 4A) and tested for evidence of gene regulatory behavior
268 across the transcriptome. Since lncRNAs typically exert their function as a transcript, we set a
269 maximum time offset of 18 hours to avoid secondary (altered protein level) effects and examined
270 patterns in the predicted gene targets of lncRNAs ($r>0.8$, $p<0.05$, divided by positive or negative
271 associations). Reactome pathway analysis revealed that 11 of these lncRNAs (including well
272 characterized lncRNAs, Supplementary Figure S6B&C) were potentially involved in regulating
273 networks of genes associated with key developmental processes ($p.adj<0.05$, Supplementary
274 Figure S6C). These analyses assigned target gene networks consistent with characterized lncRNA
275 biological functions for Malat1 (oncogenic(Li et al. 2009)), Neat1 & Rian (association with gene

276 repression(Guttman et al. 2011)) and Meg3 (tumour suppressor(Zhang et al. 2003)).
277 Interestingly, these data suggest that the pro-tumorigenic function of Malat1 may be mediated
278 through facilitating the increase of MAPK signaling molecules. Importantly, these data also
279 provide testable evidence for seven previously uncharacterized lncRNAs role in ES development
280 and describes a map of regulatory interactions driven by lncRNAs (Figure 4C) whereby lncRNAs
281 can affect gene expression across the genome. The identification of lncRNAs with a predicted
282 biological role is important for unraveling lncRNA function, providing candidate functional
283 lncRNA and providing a level of molecular detail that is currently lacking in many lncRNA studies.

284 Discussion

285 Transcriptional regulation of key biological events is a key feature in understanding the
286 complexity of cellular processes. Here we describe a detailed transcriptomic resource for
287 research in cellular development, a framework for unraveling this detail and identifying new
288 targets for analysis. We also present a comprehensively detailed survey of noncoding transcripts
289 throughout early stem cell development. We have identified many previously uncharacterized
290 noncoding RNAs with potentially pivotal roles in cellular differentiation. This will provide a
291 valuable tool for researchers unraveling the transcriptional complexity of cellular differentiation.

292 Increased interpretive power

293 The understanding of molecular events underlying the departure from pluripotency has been
294 determined by the extant knowledge of how biological functions are exerted – often measured at
295 24 hourly or greater intervals. We hypothesized that interpretations of this model were missing
296 detail in light of evidence indicating the unforeseen dynamics in RNA biology and regulation. By
297 probing this detail with finer time distinctions, we show that gene expression profiles of well-
298 characterized genes display significant variation of expression levels and that such variations are
299 manifest in a significantly more complex gene regulatory framework. This is consistent with a
300 reduction in temporal aggregation bias (Bay et al. 2004) and highlights early array-based
301 investigations in yeast demonstrating the importance of sufficient temporal resolution in
302 understanding gene expression patterns (Bar-Joseph et al. 2003). As such, much detail is likely
303 missing from other systems that involve a change in phenotype or cellular behavior. With large-

304 scale transcriptomic analyses becoming increasingly accessible, it is opportune to revisit other
305 well-studied transitions with the view of improving understanding and applicability of their
306 results rather than relying on presuppositions about gene expression patterns (Rosa et al. 2012).

307 Insights into short bursts of transcription

308 We have shown the benefit of frequent sampling over time in observing the transcription of
309 genes that are observable only within sub-24 hourly windows. This approach highlights the
310 importance of taking into account the presence of short-lived transcripts and shows that cells
311 express more of the transcriptome in a time-dependent fashion. To this end, we have identified
312 rapid changing and periodically expressed genes, which we term short-lived (slRNA) and cycling
313 (cycRNA), that were unobservable outside this framework. That many slRNAs exhibited changes
314 in expression over the first 24 hours of differentiation is consistent with rapid initial cellular
315 response to stimuli (Gasch et al. 2000; De Kumar et al. 2015). Indeed the it is likely that
316 significant gene expression changes- especially noncoding- occur on timeframes shorter than
317 those presented that may not be amenable to optimal timepoint prediction strategies (Rosa et al.
318 2012). By probing deeper into time-dependent gene transcription-possibly by interpolating
319 available datasets-(Bar-Joseph et al. 2003) it will be possible to uncover further complexity
320 underlying cellular plasticity and gene regulation. These observations reinforce the concept that
321 adequate temporal resolution is vital for describing biological transitions- for example in
322 dissecting primary from follow on effects in gene knockdown studies – and that end-point
323 analysis likely does not reflect the complex biology of phenotype changes.

324 Insight genome organization and regulation

325 Similarly, by using time to separate the order of gene transcription, we have been able to predict
326 local gene regulation across the genome. We have been able to observe concerted gene
327 expression (in *trans*) of hundreds of genes separated by large genome differences (in *cis*). Typical
328 studies of this nature involve correlative analysis requiring large samples sizes and resources
329 (Prieto et al. 2008). We have instead leveraged the time axis to achieve these as well as
330 discriminate driver from passenger molecular events. This has allowed the estimation of the time
331 delay for changes in expression of regulatory molecules to manifest in changes in their target

332 gene transcription and we have been able to unravel a potentially complex network of gene
333 profiles responding to lncRNA transcription. Finally, we have been able to use an integrated
334 biological system to draw strong associations in trans relationships with bidirectional promoters.
335 Typically these associations are observed by using thousands of gene expression profiles, yet
336 here we have been able to do so with only 42.

337 General experimental considerations

338 The design and interpretation of time course experiments has been of great interest over the past
339 decade (Bay et al. 2004; Rosa et al. 2012) and they have been used effectively to elucidate
340 transcriptome expression and regulation in many organisms (Arbeitman et al. 2002; mod et al.
341 2010; Tan et al. 2013; Boeck et al. 2016). Furthermore, improvements in sequencing
342 technologies are making the dynamics of larger and complex genomes more available to closer
343 inspection. By probing transcriptional complexity in mouse ES development, we have gained
344 insight into many areas of molecular inquiry. Using uniform dense sampling enables strong gene
345 expression relationships to be drawn whilst simultaneously facilitating the dissection of
346 expression ordering and kinetics. Importantly these data show that substantial changes in gene
347 expression cannot be inferred from coarse time-points and that the continuous representation of
348 gene expression data in many developmental time courses obscures detail. Therefore the
349 assumptions made when choosing time points for these kinds of studies (such as how long a
350 biologically significant event takes to occur) need to be re-evaluated; RNA and protein turnover
351 is extremely rapid (Schwanhausser et al. 2011) and transcriptional responses are extremely
352 rapid (Gasch et al. 2000; Bar-Joseph et al. 2003; De Kumar et al. 2015) and can be transient (Lee
353 et al. 2014). It is therefore likely that dense profiling will yield more insights into reprogramming
354 until some sort of maxima or real-time picture is reached, especially in the first 24 hours.
355 Furthermore, using temporal approaches to augment single cell transcriptome studies such as
356 dissecting cellular heterogeneity (Trapnell et al. 2010) and lncRNA expression patterns (Kim et
357 al. 2015) similar to the method employed in (Chu et al. 2016) may allow the temporal tracking of
358 single cell alterations over time.

359 Analysis of high-resolution temporal transcriptomic data reveals an unprecedented level of
360 regulatory complexity and presents a tantalizing opportunity to revisit and bring new insight

361 into other clinically or biotechnologically significant biological transitions. In designing these
362 experiments it is important to choose the approach to match the aim. For example, gene
363 knockdown experiments using siRNAs may benefit from early time point transcriptomes for
364 dissecting primary from secondary or tertiary effects. Uniform temporal sampling simplifies the
365 interpretation of temporal correlations in gene expression whereas focus on early responses will
366 necessarily require rapid initial time point selection with tracking samples. The frequency of
367 collection will necessarily depend on the duration of the response and practical and financial
368 considerations. The ability to observe transient gene expression increases incrementally with
369 each additional time point measured (Supplementary Figure S7) Finally, temporal experiments
370 should always be performed in duplicate at least to ensure uniformity of the biology underlying
371 the process in question.

372

373 # Methods

374 ## Sample Generation and Library Preparation

375 Biological duplicate, low passage number (P18) W9.5 ESCs were cultured and differentiated as
376 described previously(Bruce et al. 2007b; Dinger et al. 2008). Cultures were harvested every six
377 hours from the induction of differentiation to 120 hours post differentiation induction. Total RNA
378 from cultures was purified using Trizol (Life Technologies) and DNase treatment was performed
379 by RQ1 DNase (Promega) according to the manufacturer's instructions. RNA integrity was
380 measured on a Bioanalyzer RNA Nano chip (Agilent). RNA-Seq library preparation and
381 sequencing of Poly-A-NGS libraries generated from 500 ng total RNA using SureSelect Strand
382 Specific RNA Library Preparation Kit (Agilent) according to the manufacturer's instructions.
383 Paired-end libraries were sequenced to the first 100 bp on a HiSeq 2500 (Illumina) on High
384 Output Mode.

385 ## Quality control and read mapping

386 Library sequencing quality was determined using FastQC (Babraham Bioinformatics) and FastQ
387 Screen (Babraham Bioinformatics). Illumina adaptor sequence and low quality read trimming
388 (read pair removed if < 20 base pairs) was performed using Trim Galore! (Babraham
389 Bioinformatics: www.bioinformatics.babraham.ac.uk/). Tophat2 (Kim et al. 2013) was used to
390 align reads to the December 2011 release of the mouse reference genome (mm10) as outlined by
391 Anders et al.(Anders et al. 2013). Read counts data corresponding to GENCODE vM2 transcript
392 annotations were generated using HTSeq (Anders et al. 2014). *de novo* transcript assembly was
393 performed on each merged BAM file using Cufflinks' reference annotation based transcript
394 (RABT) assembly(Roberts and Pachter 2011), using the Gencode vM2 transcriptome(Harrow et
395 al. 2012) as a guide (options: -u -I 500000 -j 1.0 -F 0.005 --trim-3-dropoff-frac 0.05 –g
396 gencode.vM2.annotation.gtf --library-type fr-firststrand). Transcript assemblies were then
397 merged using Cuffmerge(Trapnell et al. 2010) using default parameters, and compared to the
398 Gencode vM2 reference transcriptome using Cuffcompare(Trapnell et al. 2010). Novel transcripts
399 with a Cuffcompare class code of j, i, o, u or x were filtered using three steps to find novel
400 lncRNAs. First, a Browser Extensible Data (BED) format file was generated using a python script

401 (https://gist.github.com/davidliwei/1155568) and any single exon transcripts were removed.
402 Second, the FASTA-formatted sequence for each transcript was obtained using
403 BEDTools(Quinlan and Hall 2010), the nucleotide (nt) length and open reading frame (ORF) size
404 found using Perl scripts, and those with a length less than 200 nt or a ORF size greater than 300
405 nt were removed. Lastly, transcript sequences were submitted to Coding Potential Calculator
406 (CPC)(Kong et al. 2007), and those with a coding potential of >0 were removed.
407

408 ## Bioinformatics

409 All analyses were performed in the R Statistical Environment(R Core Team 2014). Briefly, counts
410 data were background corrected and normalized for library size using edgeR(Robinson et al.
411 2010), then transformed using voom(Law et al. 2014) for differential expression analysis using
412 LIMMA(Smyth 2004). Transcription Factor (TF) activity was inferred from gene expression data
413 using DREM(Schulz et al. 2012) with a branching P-value of 0.001 based on curated TF-target
414 gene lists associated with mouse ESC differentiation from ChEA(Lachmann et al. 2010). TF-target
415 gene was calculated by maximal Pearson's correlation coefficient of >0.8 using a custom
416 autocorrelation analysis and verified with the "ccf" function in R. Gene differential exon (DEX)
417 usage was analyzed by DEXSeq(Anders et al. 2012) on vM2 gene annotations using default
418 settings and an adjusted p value cutoff of 0.001 for DEX between biological duplicates at each
419 consecutive time-point. Genome position analyses were performed using genomic
420 ranges(Lawrence et al. 2013) based on vM2 annotations imported with 'rtracklayer'(Lawrence et
421 al. 2009) and Pearson's correlation coefficient of gene expression Bidirectional genes were
422 defined as two genes with expression data on opposing strands with <2000 bp between the
423 transcriptional start sites (TSS). Co-expressed gene clusters were defined as >5 contiguous genes
424 with expression data displaying a Pearson's Correlation Coefficient of >0.5 with neighbouring
425 genes. Cluster co-expression data was visualized with corrplot(Wei 2013) and Cytoscape
426 (v3.1.0(Shannon et al. 2003)), location of related clusters was visualized by Circos(Krzywinski et
427 al. 2009). Gene expression periodicity was measured on 120 interpolated expression
428 values(Orlando et al. 2008) for each replicate time series using GeneCycle(Strimmer 2012),
429 candidate periodically expressed genes were identified as having the same calculated dominant

430 cycling frequency between biological replicates. Time-dependent expression signatures were
431 established using maSigPro(Nueda) with a replicate correlation coefficient cutoff of 0.8. Target
432 genes of potential regulatory (top 50 most highly and/or variably expressed) lncRNAs were
433 identified using the GeneReg package(Huang 2012) on 100 point-interpolated expression data
434 based on fitted expression values between duplicates and setting a maximum time delay of 18
435 hours and a global correlation coefficient of 0.9 and visualized using Cytoscape. Gene lists were
436 functionally annotated with KEGG and Reactome pathways (adjusted p value <0.05) using the
437 clusterProfiler and ReactomePA packages(Yu et al. 2012).

Acknowledgements:

BG is supported by Cancer Institute NSW Early Career Fellowship 13/ECF/1-45

The authors acknowledge Kenneth Sabir and Ruth Pidsley for reviewing the manuscript; The Garvan Foundation and the Peter Wills Bioinformatics Facility for providing facilities. BS acknowledges comments from Aaron Statham, Mark Pinese, Nenad Bartonicek, Jesper Maag and Quek Xiucheng.

Author Contributions:

BG wrote the manuscript, assisted study conception, performed the analyses and library preparations (assisted by DK). MD conceived the study and assisted writing the manuscript. BS performed some analyses (de-novo assembly and PC deconvolution), designed the web portal, assisted with figure composition and reviewed the manuscript. SC, FG and DK performed lab work and reviewed the manuscript. AP provided biological samples and facilities.

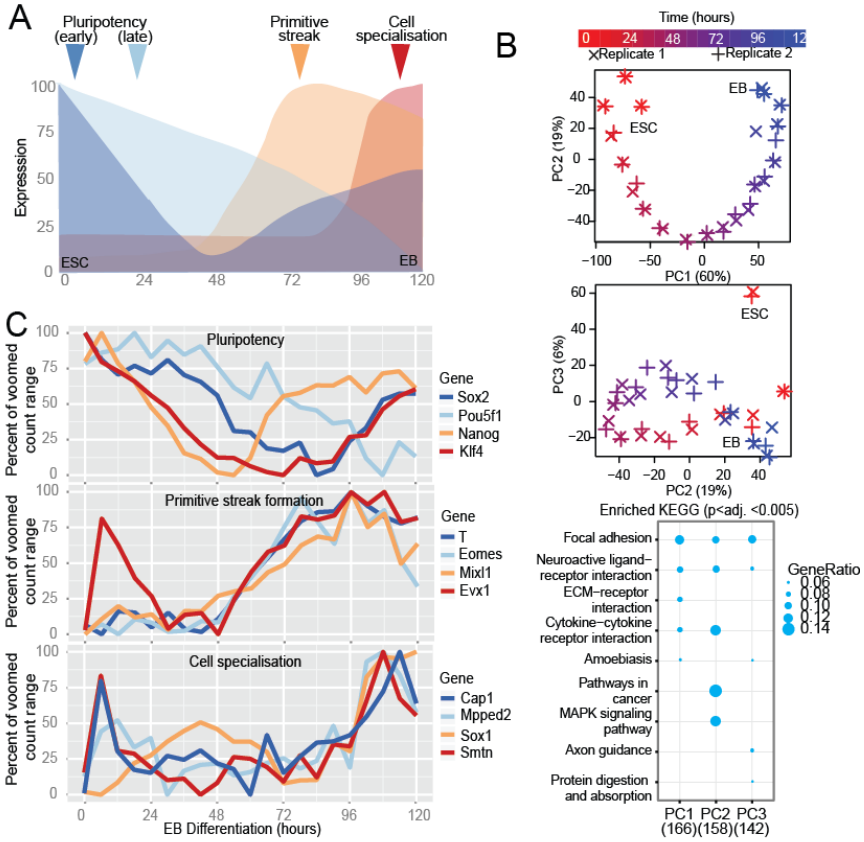
Disclosure and data access

Data has been deposited into GEO accession GSE75028

The authors declare no conflict of interest.

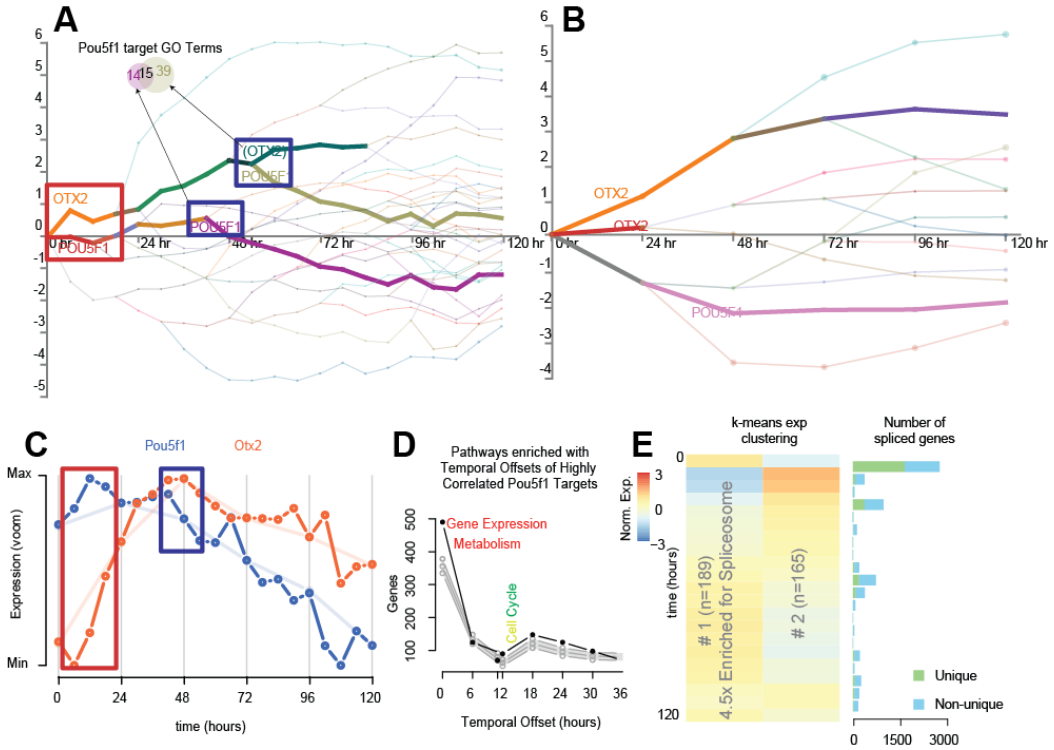
Correspondence and requests for materials should be addressed to m.dinger@garvan.org.au and b.gloss@garvan.org.au

Figures and Legends



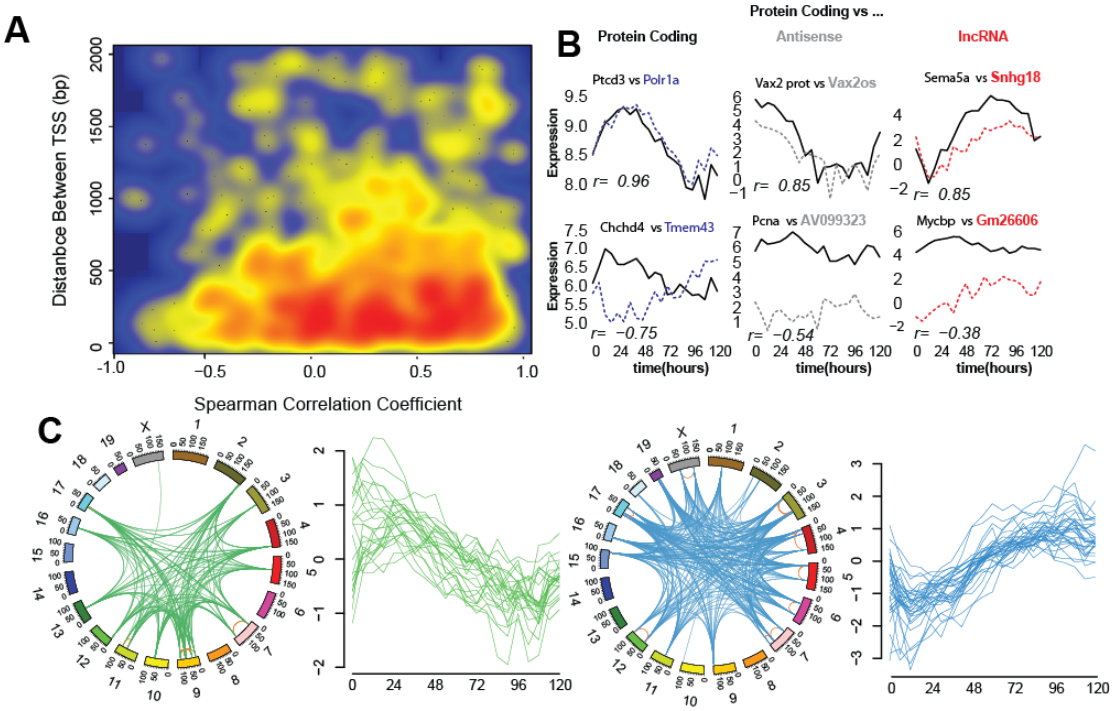
Gloss *et al.* Figure 1

Global and gene-specific evaluation of augmented temporal resolution in mES differentiation. (A) Schematic of mouse embryonic stem cell (ESC) differentiation into embryoid bodies (EB) over the time course evaluated here. (**B**)Analysis of the top three principle components (PCs) based on the 2,000 most variable genes from biological duplicate-6 hourly transcriptomes.**and** KEGG pathway enrichment for 500 genes contributing most to each of the top three PCs. (**C)** Expression profiles of genes associated with pluripotency, primitive streak formation and cell specialization .



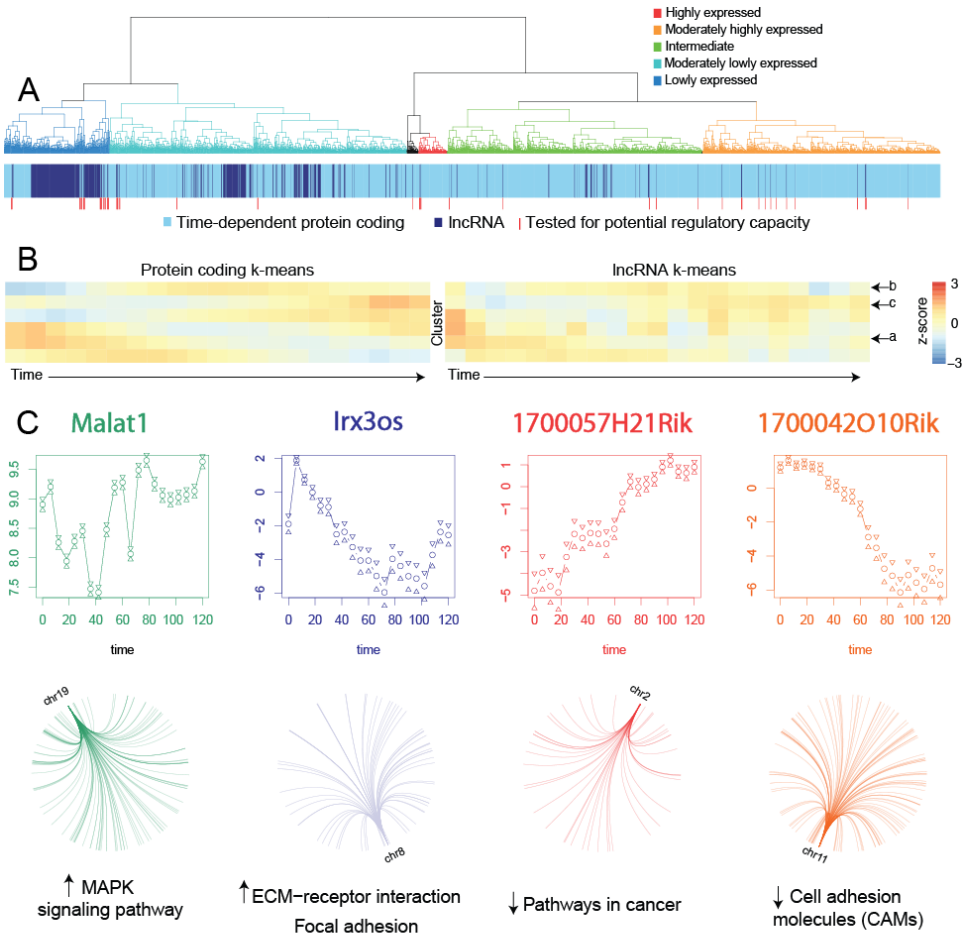
Gloss *et al.* Figure 2

Insights into regulatory and gene expression kinetics . (A & B) Observable regulatory network dynamics at 24- and 6-hourly measures with OTX2 and POU5F1 target containing profiles annotated and in bold, See Supplementary Figure S2 for full figure. Transcriptomes at 24- (top) and 6-hourly (bottom) were subjected to DREM analysis of mouse TF/target gene interactions. A p-value cutoff of 0.001 was applied to calculating divergent TF activity (splits). Relative circle sizes are proportional to the spread of gene expression levels corresponding to that point. Red and blue boxes pertain to branch points of interest **(C)** Expression of the key transcription factors Pou5F1 (Oct4) and OTX2. Red and blue boxes correspond to the time points highlighted in part A. **(D)** Distribution of the number of genes and the time delay required to meet a maximum correlation (>0.8) between gene targets of Pou5f1 and pPou5f1 itself compared to 95% quantiles of 500 random gene selections. **(E)** Two k-means clusters short-lived RNA (slRNA) genes displaying differential expression without changes at 24-hourly time points (adj. p<0.0001).



Gloss *et al.* Figure 3

Analysis of gene coexpression patterns using augmented temporal resolution. (**A)** Smoothed scatter plot showing the correlation coefficient across the time course vs. distance between transcriptional start sites (TSS) of bidirectional gene pairs. Blue indicates no gene pairs; yellow and red indicate increasing numbers of pairs sharing similar properties. (**B)** Expression patterns of example bidirectional genes of the same or different gene biotype. Spearman's correlation coefficient is reported for each pair. (**C)** Genomic location (circos) and expression pattern (line plot) of two independent co-expressed groups of 5 or more contiguous genes sharing correlated expression (r>0.5).

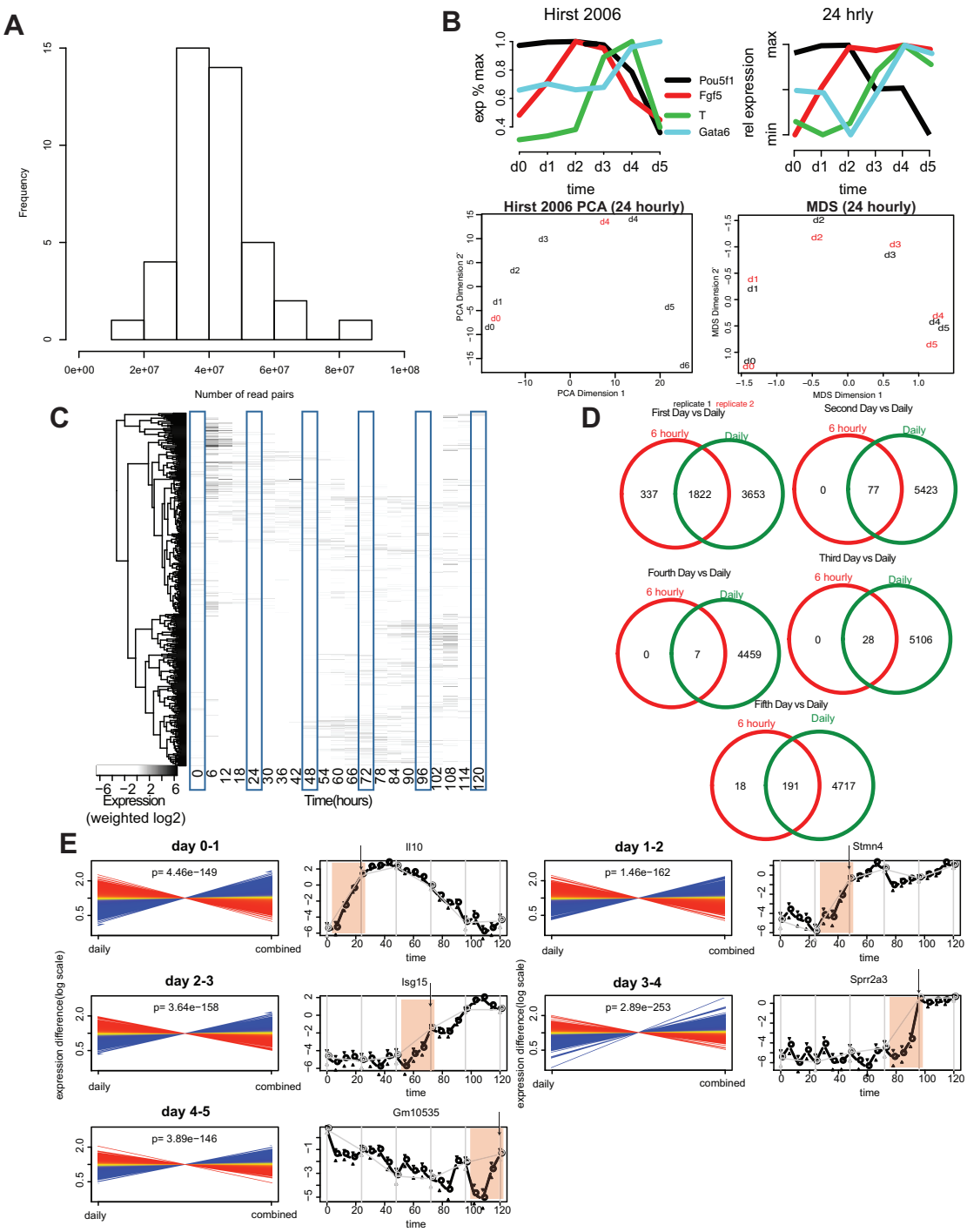


Gloss *et al.* Figure 4

Augmented temporal resolution of ncRNA expression in cellular differentiation. (A) Hierarchical clustering of lncRNAs (dark blue) with time-dependent protein coding genes (light blue) by their expression patterns over time. Dendrogram was manually colored to reflect gene expression levels of the top-level clusters. **(B)** K-means clustered expression profiles of protein coding genes compared to the same number of lncRNA gene expression clusters. Common profiles are marked with arrows. **(C)** Expression profiles of four lncRNAs predicted to have regulatory roles in ES development as well as the genome location & pathways enriched in their gene targets. Malat1 and IRX3os display a positive association with their targets, whereas 1700057H21Rik and 1700042O10Rik have a putative repressive impact.

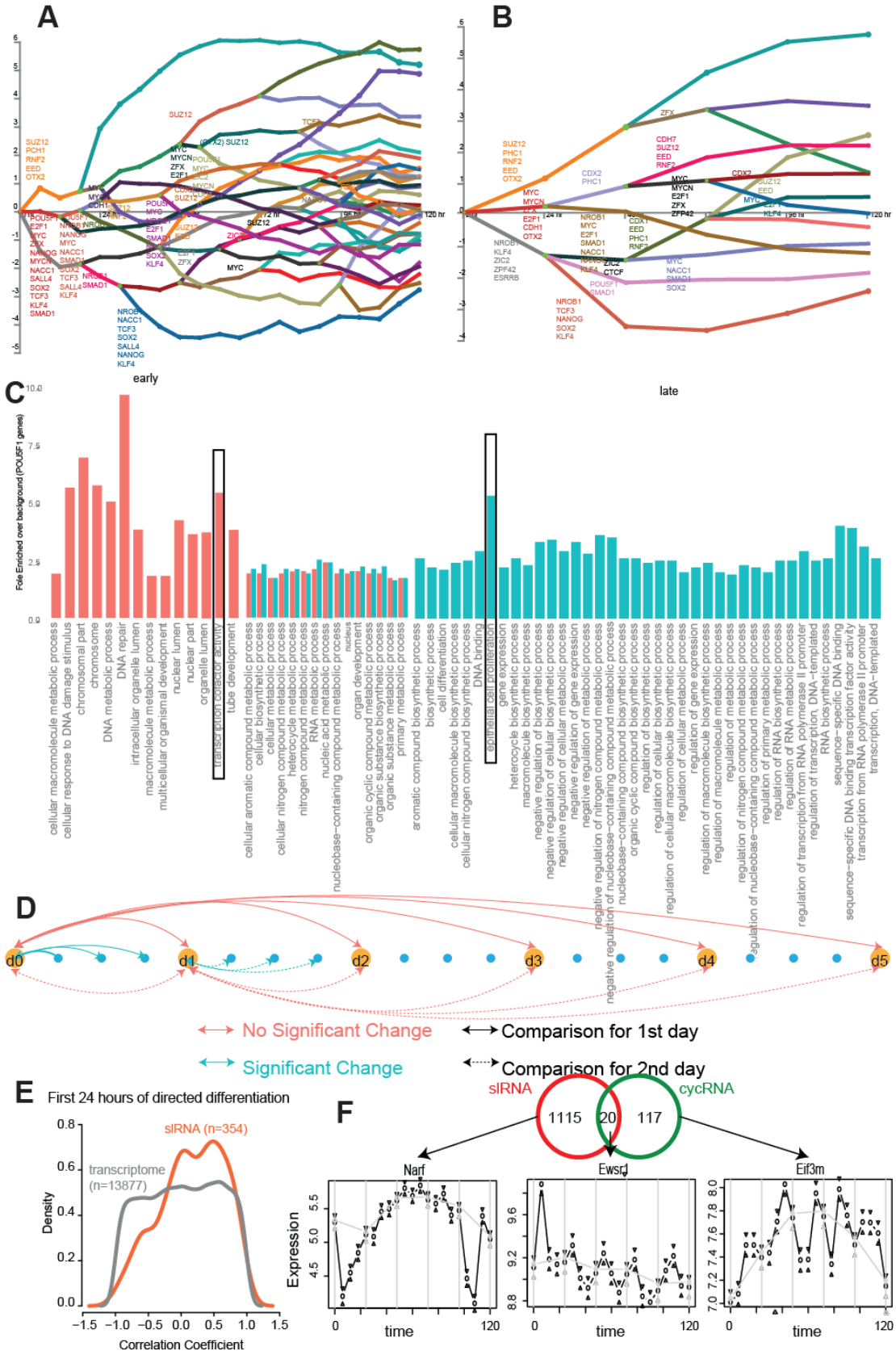
Supplementary Figures and Legends

Please see attached Excel Spreadsheet (Gloss_2015_SuppTab.xlsx) containing
Supplementary Table S1: Counts of Genes by biotype
Supplementary Table S2: Periodic Genes



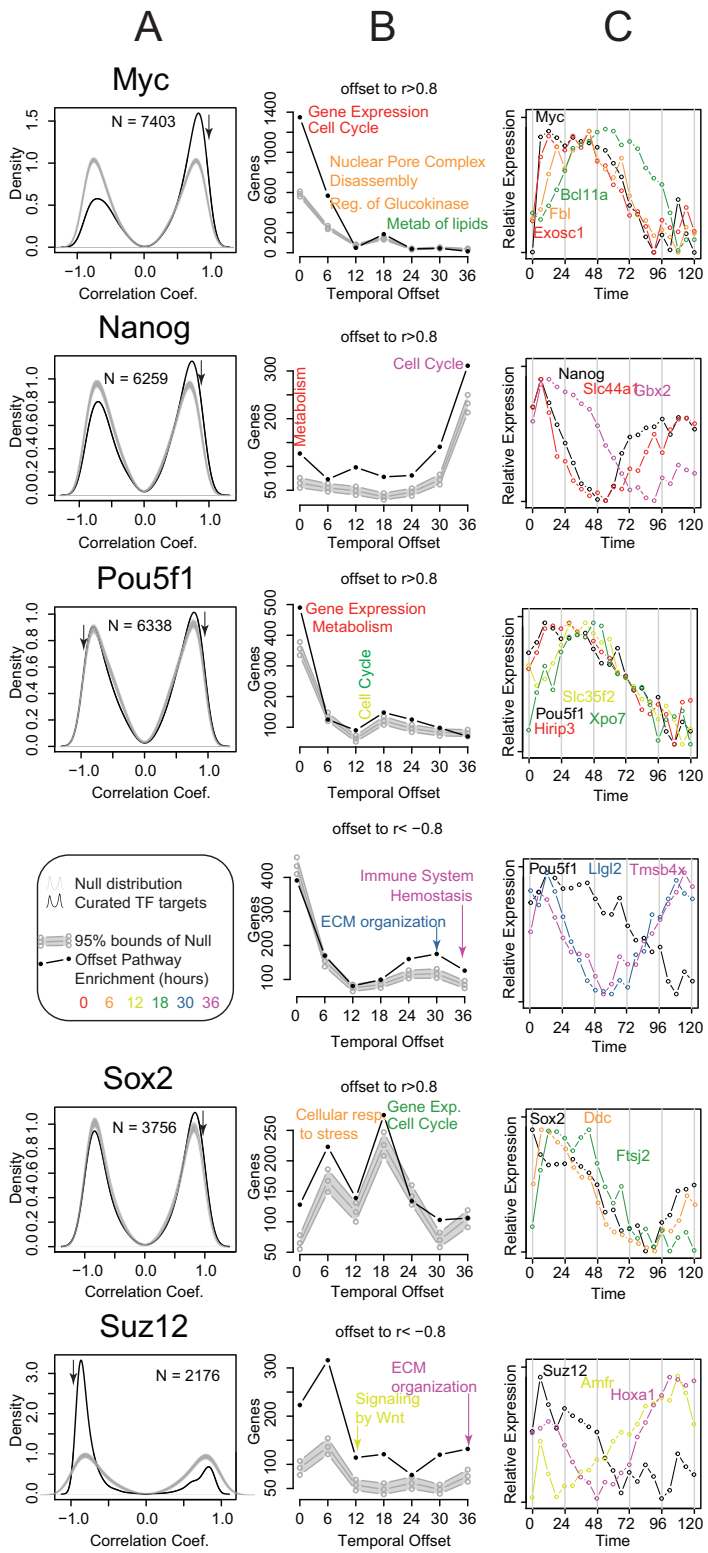
Gloss *et al.* Supplementary Figure S1

Global evaluation of high-resolution transcriptomic data. **(A)** Histogram of mapped read number distribution per sample (pooled from biological replicates). **(B)** Comparison of expression levels and principle components analysis measurd 24 hourly between this study and Hirst 2006 **(C)** Heatmap of expression levels for genes only expressed outside of 24 hourly timepoints, clustered by expression pattern. **(D)** Evidence of differential expression within one 24 hour period vs. any change across all 24hourly times (p <0.0001). **(E)** Comparing whether the 24 hourly measures "summarize" that 24 hour window by comparing mean expression for that window with the 24 hour time point



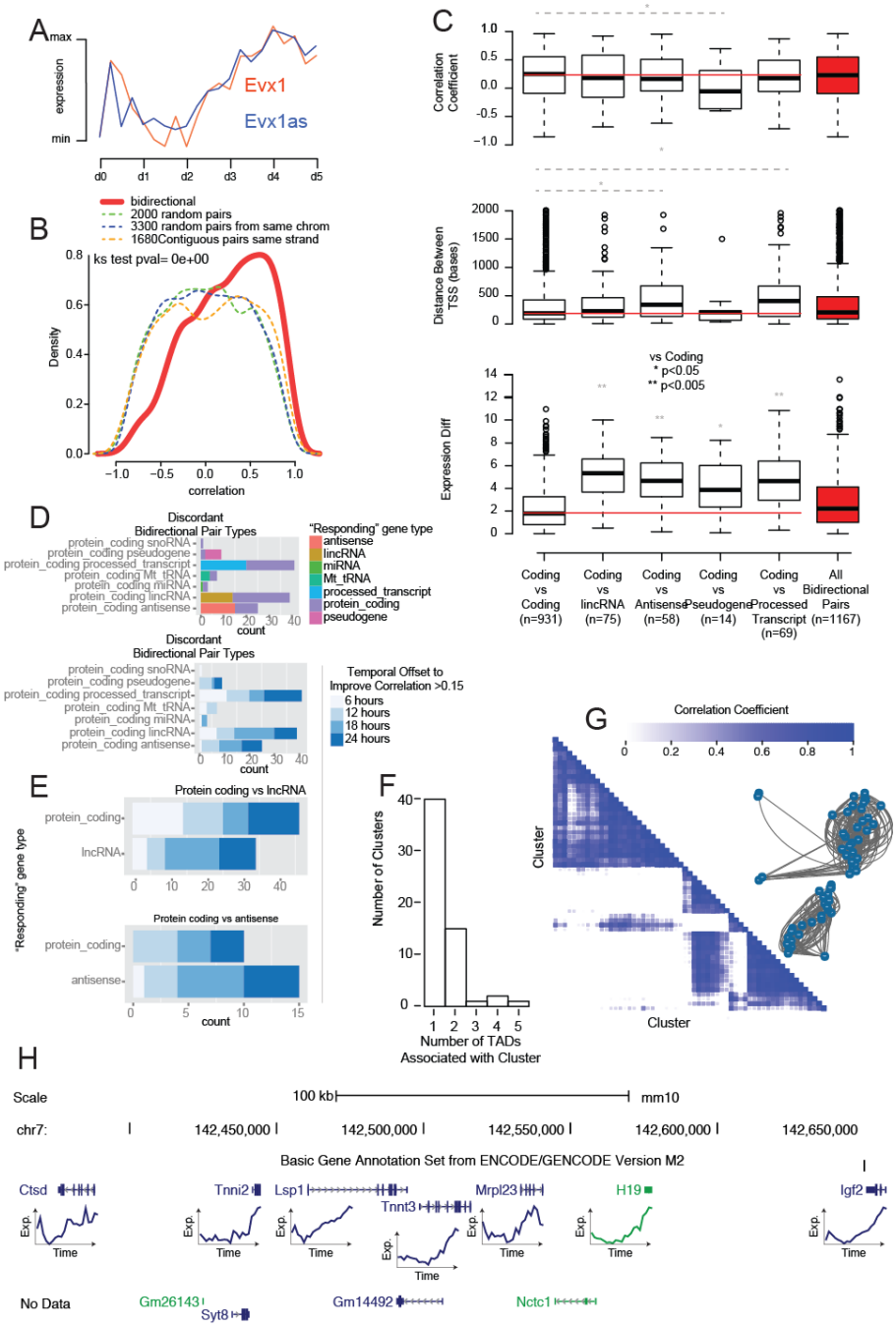
Gloss *et al.* Supplementary Figure S2

Highlighting unique knowledge gained from increased temporal resolution. (A &B) Fully annotated DREM schematic of estimated TF activity of key ESC related TFs at 6hourly (A) vs. 24 hourly (B). **(C)** GO term enrichment (adjusted p<0.05) for genes corresponding branch points designated as early (co-observed with change in POU expression) and late (observed after POU5f1 Expression changes) highlighted by the blue boxes in Figure 2A. Black boxes represent similar terms identified in figure 2D. **(D)** Schematic of differential expression analysis design used to identify slRNAs. **(E)** Correlation of slRNA expression in (De Kumar et al. 2015). **(F)** Comparison of slRNAs and cycRNAs. Venn diagram of the overlap observed and examples from each class.



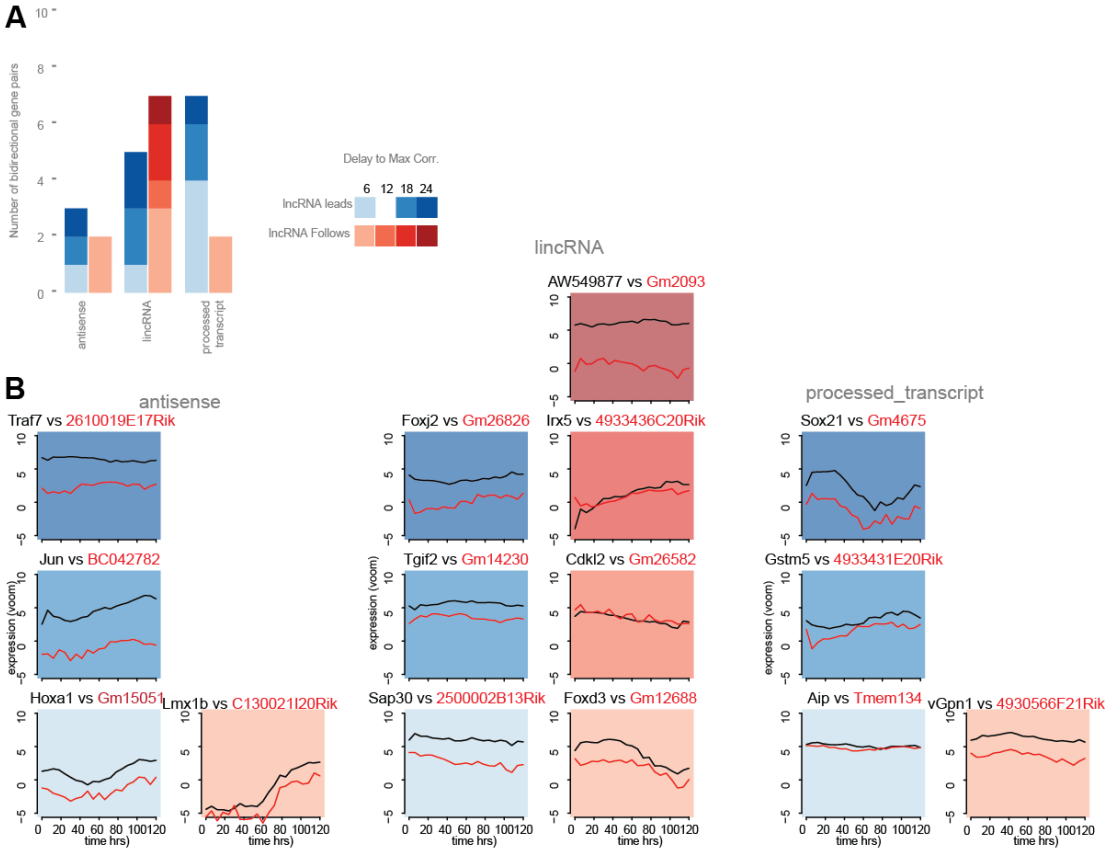
Gloss *et al.* Supplementary Figure S3

Temporal offsets in transcription factor (TF)- target gene expression. **(A)** Curated TF/gene targets were downloaded from chea (http://amp.pharm.mssm.edu/lib/chea.jsp) for Myc, Nanog, Pou5f1, Sox2 and Suz12. Expression of target genes were tested for correlation with their TF at different temporal offsets (0-36 hours) and compared to 500 random selections of the same number of genes (Null). Where absolute correlations of predicted targets exceeded the null distribution (arrow), **(B)** the number of genes achieving a maximal absolute correlation of >0.8 and the offset required to reach these maxima was plotted against the 5th and 95thquantiles of the same results from the null distribution. Where the number of target genes exceeded the null distribution, the lists of genes in each offset were tested for enrichment of Reactome pathways relative to the total predicted target list (enrichment). **(C)** Example expression patterns of genes displaying these attributes were plotted.



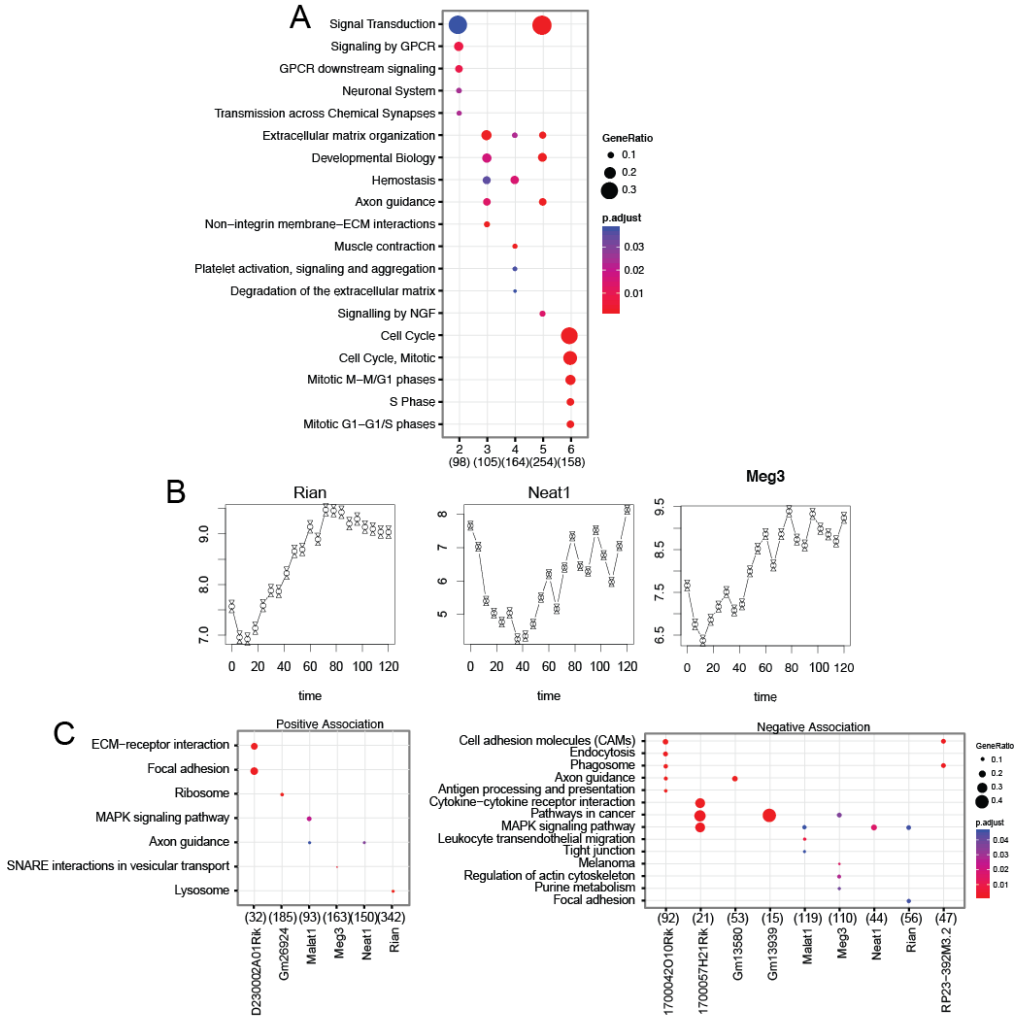
Gloss *et al.* Supplementary Figure S4

Bidirectional and co-expression analysis of mouse ES development. (A) Expression profile of EVX1 and its antisense (and positively correlated) transcript EVX1AS- the peak at 6-18 hours has not been observed previously. **(B)** Distribution of correlation coefficients of bidirectional gene pairs (red) compared to similar numbers of randomly chosen genes pairs, randomly chosen genes from the same chromosome and, randomly selected neighbouring genes (dotted lines). ks=Kolmogorov–Smirnov test (bidirectional vs. random neighbouring gene pairs). **(C)** Characteristics of bidirectional gene pairs (Correlation coefficient, Distance between TSS and Difference of median expression (log scale) based on annotated gene-biotype. **(D)** Counts of bidirectional gene pairs of differing biotypes achieving an improved correlation coefficient of >0.15 (to at lease 0.25) over that at time zero, colored by the biotype of the "following" gene or by the temporal offset required to achieve the improvement. **(E)** Comparison of responding gene biotype to the temporal offset for lincRNA and antisense biotypes. **(F)** Number of topological associated domains (TADS, HindIII data mapped to mm10 using liftOver from mm9) associated with each co-expressed gene cluster. **(G)** Clustering of co-regulated gene clusters by correlation coefficient visualized by network diagram and hierarchical clustering of the correlation matrix. **(H)** The imprinted H19/IGF2 cluster identified as a co-expressed gene cluster with gene expression data for measured genes. Some genes did not have expression data (no data).



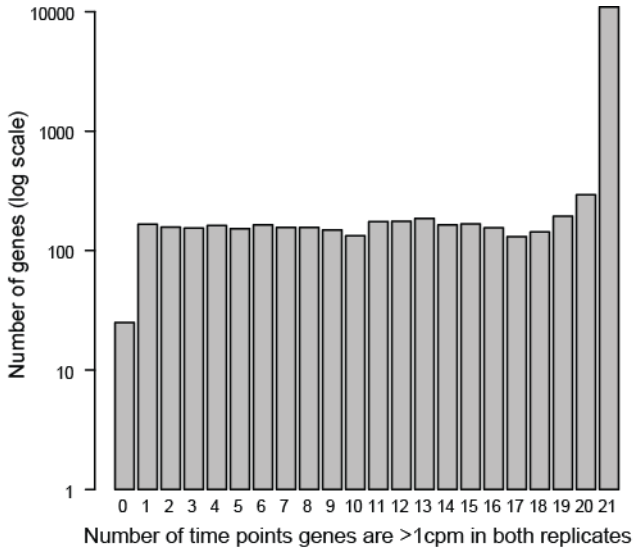
Gloss *et al.* Supplementary Figure S5

Temporal relationships of highly correlated coding-noncoding bidirectional pairs. (A) Bar chart of the temporal offset required to reach a maximum correlation >0.8 and whether the noncoding gene preceded the protein coding gene or vice versa. (**B)** Example gene expression profiles of bidirectional paired gene over the time course. Gene profiles are arranged and colored as the bar chart.



Gloss *et al.* Supplementary Figure S6

LncRNAs and their role in ES development. (A) Reactome pathway enrichment for 5/6 k-means clusters of time-dependent protein coding genes. (**B**) Expression profiles for characterized lncRNAs described in text. (**C**) Reactome pathway enrichment for putative gene targets positively or negatively associated with candidate lncRNAs (top 4 pathways, enrichment adj. pval.<0.05)



Gloss *et al.* Supplementary Figure S7

The number of conditions in which each gene observed is expressed above background in both replicates across the time course. ~150 new genes are observed at a single timepoint.

References

Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, Huber W, Robinson MD. 2013. Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nature protocols* **8**(9): 1765-1786.

Anders S, Pyl PT, Huber W. 2014. HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics*.

Anders S, Reyes A, Huber W. 2012. Detecting differential usage of exons from RNA-seq data. *Genome research* **22**(10): 2008-2017.

Arbeitman MN, Furlong EE, Imam F, Johnson E, Null BH, Baker BS, Krasnow MA, Scott MP, Davis RW, White KP. 2002. Gene expression during the life cycle of Drosophila melanogaster. *Science (New York, NY)* **297**(5590): 2270-2275.

Bar-Joseph Z, Gerber GK, Gifford DK, Jaakkola TS, Simon I. 2003. Continuous representations of time-series gene expression data. *Journal of computational biology : a journal of computational molecular cell biology* **10**(3-4): 341-356.

Bay SD, Chrisman L, Pohorille A, Shrager J. 2004. Temporal aggregation bias and inference of causal regulatory networks. *Journal of computational biology : a journal of computational molecular cell biology* **11**(5): 971-985.

Bergmann JH, Li J, Eckersley-Maslin MA, Rigo F, Freier SM, Spector DL. 2015. Regulation of the ESC transcriptome by nuclear long noncoding RNAs. *Genome research*.

Bertone P, Stolc V, Royce TE, Rozowsky JS, Urban AE, Zhu X, Rinn JL, Tongprasit W, Samanta M, Weissman S et al. 2004. Global identification of human transcribed sequences with genome tiling arrays. *Science (New York, NY)* **306**(5705): 2242-2246.

Boeck ME, Huynh C, Gevirtzman L, Thompson OA, Wang G, Kasper DM, Reinke V, Hillier LW, Waterston RH. 2016. The time resolved transcriptome of C. elegans. *Genome research*.

Bonasio R, Shiekhattar R. 2014. Regulation of Transcription by Long Noncoding RNAs. *Annual review of genetics* **48**: 433-455.

Bruce SJ, Gardiner BB, Burke LJ, Gongora MM, Grimmond SM, Perkins AC. 2007a. Dynamic transcription programs during ES cell differentiation towards mesoderm in serum versus serum-freeBMP4 culture. *BMC genomics* **8**: 365.

Bruce SJ, Rea RW, Steptoe AL, Busslinger M, Bertram JF, Perkins AC. 2007b. In vitro differentiation of murine embryonic stem cells toward a renal lineage. *Differentiation; research in biological diversity* **75**(5): 337-349.

Chen H, Mundra PA, Zhao LN, Lin F, Zheng J. 2014. Highly sensitive inference of time-delayed gene regulation by network deconvolution. *BMC systems biology* **8 Suppl 4**: S6.

Chu LF, Leng N, Zhang J, Hou Z, Mamott D, Vereide DT, Choi J, Kendziorski C, Stewart R, Thomson JA. 2016. Single-cell RNA-seq reveals novel regulators of human embryonic stem cell differentiation to definitive endoderm. *Genome biology* **17**(1): 173.

Clark MB, Johnston RL, Inostroza-Ponta M, Fox AH, Fortini E, Moscato P, Dinger ME, Mattick JS. 2012. Genome-wide analysis of long noncoding RNA stability. *Genome research* **22**(5): 885-898.

Cloonan N, Forrest AR, Kolle G, Gardiner BB, Faulkner GJ, Brown MK, Taylor DF, Steptoe AL, Wani S, Bethel G et al. 2008. Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nature methods* **5**(7): 613-619.

De Kumar B, Parrish ME, Slaughter BD, Unruh JR, Gogol M, Seidel C, Paulson A, Li H, Gaudenz K, Peak A et al. 2015. Analysis of dynamic changes in retinoid-induced transcription and epigenetic profiles of murine Hox clusters in ES cells. *Genome research* **25**(8): 1229-1243.

Dinger ME, Amaral PP, Mercer TR, Pang KC, Bruce SJ, Gardiner BB, Askarian-Amiri ME, Ru K, Solda G, Simons C et al. 2008. Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome research* **18**(9): 1433-1445.

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**(7398): 376-380.

Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F et al. 2012. Landscape of transcription in human cells. *Nature* **489**(7414): 101-108.

Fatica A, Bozzoni I. 2014. Long non-coding RNAs: new players in cell differentiation and development. *Nature reviews Genetics* **15**(1): 7-21.

Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO. 2000. Genomic expression programs in the response of yeast cells to environmental changes. *Molecular biology of the cell* **11**(12): 4241-4257.

Gloss BSD, M. E. 2015. The Specificity of lncRNA expression. *Biochimica et biophysica acta*.

Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP et al. 2009. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**(7235): 223-227.

Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R, Bruhn L et al. 2011. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477**(7364): 295-300.

Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S et al. 2012. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome research* **22**(9): 1760-1774.

Hirst CE, Ng ES, Azzola L, Voss AK, Thomas T, Stanley EG, Elefanty AG. 2006. Transcriptional profiling of mouse and human ES cells identifies SLAIN1, a novel stem cell gene. *Developmental biology* **293**(1): 90-103.

Huang T. 2012. GeneReg: Construct time delay gene regulatory network.

Kaffer CR, Grinberg A, Pfeifer K. 2001. Regulatory mechanisms at the mouse Igf2/H19 locus. *Molecular and cellular biology* **21**(23): 8189-8196.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology* **14**(4): R36.

Kim DH, Marinov GK, Pepke S, Singer ZS, He P, Williams B, Schroth GP, Elowitz MB, Wold BJ. 2015. Single-cell transcriptome analysis reveals dynamic changes in lncRNA expression during reprogramming. *Cell stem cell* **16**(1): 88-101.

Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, Gao G. 2007. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic acids research* **35**(Web Server issue): W345-349.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. *Genome research* **19**(9): 1639-1645.

Lachmann A, Xu H, Krishnan J, Berger SI, Mazloom AR, Ma'ayan A. 2010. ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics* **26**(19): 2438-2444.

Law CW, Chen Y, Shi W, Smyth GK. 2014. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome biology* **15**(2): R29.

Lawrence M, Gentleman R, Carey V. 2009. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* **25**(14): 1841-1842.

Lawrence M, Huber W, Pages H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. 2013. Software for computing and annotating genomic ranges. *PLoS computational biology* **9**(8): e1003118.

Lee MC, Lopez-Diaz FJ, Khan SY, Tariq MA, Dayn Y, Vaske CJ, Radenbaugh AJ, Kim HJ, Emerson BM, Pourmand N. 2014. Single-cell analyses of transcriptional heterogeneity during drug tolerance transition in cancer cells by RNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America* **111**(44): E4726-4735.

Lercher MJ, Urrutia AO, Hurst LD. 2002. Clustering of housekeeping genes provides a unified model of gene order in the human genome. *Nature genetics* **31**(2): 180-183.

Li H, Luan Y, Hong F, Li Y. 2002. Statistical methods for analysis of time course gene expression data. *Frontiers in bioscience : a journal and virtual library* **7**: a90-98.

Li L, Feng T, Lian Y, Zhang G, Garen A, Song X. 2009. Role of human noncoding RNAs in the control of tumorigenesis. *Proceedings of the National Academy of Sciences of the United States of America* **106**(31): 12956-12961.

Liu N, Liu L, Pan X. 2014. Single-cell analysis of the transcriptome and its application in the characterization of stem cells and early embryos. *Cellular and molecular life sciences : CMLS* **71**(14): 2707-2715.

Liu Y, Conaway L, Rutherford Bethard J, Al-Ayoubi AM, Thompson Bradley A, Zheng H, Weed SA, Eblen ST. 2013. Phosphorylation of the alternative mRNA splicing factor 45 (SPF45) by Clk1 regulates its splice site utilization, cell migration and invasion. *Nucleic acids research* **41**(9): 4949-4962.

Martello G, Smith A. 2014. The nature of embryonic stem cells. *Annual review of cell and developmental biology* **30**: 647-675.

Mercer TR, Dinger ME, Sunkin SM, Mehler MF, Mattick JS. 2008. Specific expression of long noncoding RNAs in the mouse brain. *Proceedings of the National Academy of Sciences of the United States of America* **105**(2): 716-721.

mod EC, Roy S, Ernst J, Kharchenko PV, Kheradpour P, Negre N, Eaton ML, Landolin JM, Bristow CA, Ma L et al. 2010. Identification of functional elements and regulatory circuits by Drosophila modENCODE. *Science (New York, NY)* **330**(6012): 1787-1797.

Nueda ACMJ. maSigPro: Significant Gene Expression Profile Differences in Time Course Microarray Data.

Orlando DA, Lin CY, Bernard A, Wang JY, Socolar JE, Iversen ES, Hartemink AJ, Haase SB. 2008. Global control of cell-cycle transcription by coupled CDK and network oscillators. *Nature* **453**(7197): 944-947.

Paronetto MP, Minana B, Valcarcel J. 2011. The Ewing sarcoma protein regulates DNA damage-induced alternative splicing. *Molecular cell* **43**(3): 353-368.

Poirier F, Chan CT, Timmons PM, Robertson EJ, Evans MJ, Rigby PW. 1991. The murine H19 gene is activated during embryonic stem cell differentiation in vitro and at the time of implantation in the developing embryo. *Development (Cambridge, England)* **113**(4): 1105-1114.

Prieto C, Risueno A, Fontanillo C, De las Rivas J. 2008. Human gene coexpression landscape: confident network derived from tissue transcriptomic profiles. *PloS one* **3**(12): e3911.

Quek XC, Thomson DW, Maag JL, Bartonicek N, Signal B, Clark MB, Gloss BS, Dinger ME. 2015. lncRNAdb v2.0: expanding the reference database for functional long noncoding RNAs. *Nucleic acids research* **43**(Database issue): D168-173.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**(6): 841-842.

R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Roberts A, Pachter L. 2011. RNA-Seq and find: entering the RNA deep field. *Genome medicine* **3**(11): 74.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**(1): 139-140.

Rosa A, Brivanlou AH. 2013. Regulatory non-coding RNAs in pluripotent stem cells. *International journal of molecular sciences* **14**(7): 14346-14373.

Rosa BA, Zhang J, Major IT, Qin W, Chen J. 2012. Optimal timepoint sampling in high-throughput gene expression experiments. *Bioinformatics* **28**(21): 2773-2781.

Salomonis N, Schlieve CR, Pereira L, Wahlquist C, Colas A, Zambon AC, Vranizan K, Spindler MJ, Pico AR, Cline MS et al. 2010. Alternative splicing regulates mouse embryonic stem cell pluripotency and differentiation. *Proceedings of the National Academy of Sciences of the United States of America* **107**(23): 10514-10519.

Schulz MH, Devanny WE, Gitter A, Zhong S, Ernst J, Bar-Joseph Z. 2012. DREM 2.0: Improved reconstruction of dynamic regulatory networks from time-series expression data. *BMC systems biology* **6**: 104.

Schwanhausser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M. 2011. Global quantification of mammalian gene expression control. *Nature* **473**(7347): 337-342.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* **13**(11): 2498-2504.

Smyth GK. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical applications in genetics and molecular biology* **3**: Article3.

Strimmer MAaKFaK. 2012. GeneCycle: Identification of Periodically Expressed Genes.

Tan MH, Au KF, Yablonovitch AL, Wills AE, Chuang J, Baker JC, Wong WH, Li JB. 2013. RNA sequencing reveals a diverse and dynamic repertoire of the Xenopus tropicalis transcriptome over development. *Genome research* **23**(1): 201-216.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology* **28**(5): 511-515.

Trinklein ND, Aldred SF, Hartman SJ, Schroeder DI, Otillar RP, Myers RM. 2004. An abundance of bidirectional promoters in the human genome. *Genome research* **14**(1): 62-66.

Wei T. 2013. corrplot: Visualization of a correlation matrix.

Yang M, Elnitski L. 2014. Orthology-driven mapping of bidirectional promoters in human and mouse genomes. *BMC bioinformatics* **15 Suppl 17**: S1.

Yang SH, Kalkan T, Morissroe C, Marks H, Stunnenberg H, Smith A, Sharrocks AD. 2014. Otx2 and Oct4 drive early enhancer activation during embryonic stem cell transition from naive pluripotency. *Cell reports* **7**(6): 1968-1981.

Yu G, Wang LG, Han Y, He QY. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics : a journal of integrative biology* **16**(5): 284-287.

Zhang X, Lian Z, Padden C, Gerstein MB, Rozowsky J, Snyder M, Gingeras TR, Kapranov P, Weissman SM, Newburger PE. 2009. A myelopoiesis-associated regulatory intergenic noncoding RNA transcript within the human HOXA cluster. *Blood* **113**(11): 2526-2534.

Zhang X, Zhou Y, Mehta KR, Danila DC, Scolavino S, Johnson SR, Klibanski A. 2003. A pituitary-derived MEG3 isoform functions as a growth suppressor in tumor cells. *The Journal of clinical endocrinology and metabolism* **88**(11): 5119-5126.