The Long Non-Coding RNA Transcriptome Landscape in CHO Cells Under Batch and Fed-Batch Conditions

Davide Vito and C Mark Smales

Industrial Biotechnology Centre and School of Biosciences, University of Kent, Canterbury, Kent, CT2 7NJ, UK

Email: c.m.smales@kent.ac.uk

ABSTRACT

The role of non-coding RNAs in determining growth, productivity and recombinant product quality attributes in Chinese hamster ovary (CHO) cells has received much attention in recent years, exemplified by studies into microRNAs in particular. However, other classes of non-coding RNAs have received less attention. One such class are the non-coding RNAs known collectively as long non-coding RNAs (lncRNAs). We have undertaken the first landscape analysis of the lncRNA transcriptome in CHO using a mouse based microarray that also provided for the surveillance of the coding transcriptome. We report on those lncRNAs present in a model host CHO cell line under batch and fed-batch conditions on two different days and relate the expression of different lncRNAs to each other. We demonstrate that the mouse microarray was suitable for the detection and analysis of thousands of CHO lncRNAs and validated a number of these by qRT-PCR. We then further analysed the data to identify those lncRNAs whose expression changed the most between growth and stationary phases of culture or between batch and fed-batch culture to identify potential lncRNA targets for further functional studies with regard to their role in controlling growth of CHO cells. We discuss the implications for the publication of this rich dataset and how this may be used by the community.

Keywords: Long non-coding RNAs (lncRNAs); Chinese hamster ovary (CHO) cells; microarray; transcriptome; recombinant protein production

Abbreviations: [**CHO**, Chinese hamster ovary cells; **lncRNAs**, Long non-coding RNAs; **mAbs**, monoclonal antibodies; **ORF**, open reading frame; **ceRNAs**, competing endogenous RNAs; **arcRNAs**, architectural RNAs; **eRNAs**, enhance; **FDR**, false discovery rate; **PCA**, principal component analysis; **CH**, Chinese hamster; **PVT1**, Plasmacytoma Variant Translocation 1; **KIFC1**, kinesin family member C1; **HNRNPs**, Heterogeneous nuclear ribonucleoproteins; **DE**, differentially expressed; **FC**, fold change]

Introduction

Mammalian expression systems are widely used for the production of recombinant protein biopharmaceuticals, largely due to their ability to correctly fold, assemble, post-translationally modify and secrete complex human like proteins [1]. Among these, the Chinese hamster ovary (CHO) cell is the most widely utilised expression platform used in industry, especially for the production of monoclonal antibodies (mAbs), and are used for the expression of more than 60% of biotherapeutic proteins made in mammalian cells [1–5]. The most widely used bioprocess for the production of biopharmaceuticals from CHO cells involves fed-batch culture, this offering an advantage over batch culture in terms of cell growth, viability and product yields due to the supplementation of nutrients resulting in higher biomass accumulation, less build-up of toxic metabolic by-products and enhanced productivity [6–8].

The prominence of CHO cell expression systems has driven innovation in the industry such that CHO expression systems and associated bioprocesses have been developed that can deliver yields of mAb in excess of 5-10 g/L in stably expressing, fed-batch systems [9,10]. Despite this, some biotherapeutic recombinant proteins, and particularly some of the non-mAb novel biotherapeutics in development, remain difficult to express in CHO cell or other mammalian cell expression systems [11,12]. Using *a prior* knowledge about cellular systems, various approaches have been taken to engineer cells to deliver enhanced product yields and quality including the engineering of chaperones [13,14], glycosylation machinery [15–17] and proliferation control strategies, including manipulation of the cell cycle [18,19] , apoptosis [20,21] and autophagy [22,23]. A long side this approach, there have been studies to further our understanding of the potential cellular constraints on the production of mAbs [24,25], other recombinant biotherapeutics [26] and difficult to express recombinant proteins [27] in order to identify bottlenecks in the recombinant gene expression pathway and to develop new bioprocesses or adapt/engineer novel hosts for enhanced production and/or quality of such molecules [28,29]. The majority of these studies to date have focussed upon manipulation of coding genes, however with the discovery and improved understanding of non-coding RNA in the control of cellular processes, there has been much interest in the last decade or so in these non-coding RNAs in CHO cells. In particular, the manipulation of microRNAs to enhance the ability of CHO cells to produce biotherapeutic proteins has attracted much attention [30–33]. An advantage of manipulating such non-coding RNAs for modulating CHO cell phenotypes is that such engineering does not place an addition protein synthetic burden upon the host cell [34], unlike engineering of coding genes, and often such non-coding RNAs can modulate whole pathways rather than individual steps or processes as when manipulating many coding genes.

It has been estimated that at least 75% of transcripts originate from non-coding sequences [35]. Investigations into these transcripts has resulted in the identification of a class of transcript collectively known as Long non-coding RNAs (lncRNAs) [36]. LncRNAs are defined as transcripts longer than 200 nucleotides that lack a significant open reading frame (ORF), are usually transcribed by RNA polymerase II and spliced, with or without 3' polyadenylation [37–39]. A number of lncRNA molecules have been shown to play key regulatory roles in various biological processes including epigenetic regulation [40], transcriptional control [41], splicing events [42], and mRNA translation [43]. Indeed, lncRNAs are capable of modulating a wide range of cellular processes and mechanisms both in the nucleus and the cytoplasm [44]. The majority of our understanding into lncRNAs and the mechanism(s) by which they elicit their responses has come from studies relating to disease [45] and developmental studies [46]. LncRNAs elicit their effects by acting as competing endogenous RNAs (ceRNAs) by binding to and sequestering microRNAs [47], by acting as architectural RNAs (arcRNAs) whereby they form functional structures [48], act as *cis* molecules to enhance (eRNAs) coding gene expression [49] or as *trans* protein binding RNA molecules that can recruit chromatin modifying [50,51], as microRNA precursors [52], modulators of mRNA stability [53] and to impact upon post-translational modifications [54,55].

Despite the importance of lncRNAs in controlling cellular processes, and unlike small non-coding RNAs (e.g. siRNAs, microRNAs), the impact(s) of lncRNA expression on CHO cell bioprocessing with regard to growth/proliferation and recombinant protein yields and quality has barely been explored, with only a small number of studies reported [56,57]. However, these two studies show that manipulation of lncRNAs can impact upon recombinant protein production from CHO cells. One of the reasons for the small number of studies on lncRNAs in CHO cells is the lack of a comprehensive annotation of non-coding transcripts in CHO, hampering their identification, genome wide assessment of their expression and modulation during culture, functional studies and hence identification of target lncRNAs for cell engineering. One way to address this problem is to take advantage of the reported similarities between the Chinese hamster, CHO and mouse genome [58], where the number of well annotated non-coding transcripts is much higher. The similarity between CHO cells and mouse has already been utilised for the identification of 416 ncRNAs based on sequenced transcripts from a pooled CHO sample compared to the fRNAdb database of non-coding RNAs using BLAST, with most hits coming from mouse [59]. Here, we report the first analysis in CHO cells of both the coding and the non-coding transcriptome (specifically the lncRNAs) during batch and fed-batch culture at two different time points. We report on the identification of differentially expressed lncRNAs and mRNA, their interconnections and their potential impact on cellular pathways. This has allowed the mapping of the lncRNA landscape in CHO cells and

identification of lncRNAs targets in CHO for further manipulation with a view to increase proliferation and to sustain viability throughout batch and fed-batch culture.

2. Materials and Methods

2.1 Model Cell Line and Cell Culture Conditions

The CHO-S Freestyle host cell line (ThermoFisher Scientific, MA, USA) was used as a model CHO cell line for this study. Cells were routinely cultured in a Lab-Therm LT-X (Kühner AG, Basel, Switzerland) shaking incubator at 37°C, 5% CO_2, 125 rpm and 70% humidity in chemically defined serum-free growth medium (CD CHO, ThermoFisher Scientific, MA, USA). Fed-batch cultures were supplemented with CHO CD Efficient Feed B Liquid Nutrient Supplement (ThermoFisher Scientific, MA, USA). Initial supplementation testing was conducted in duplicate following Conditions 3 and 9 as described in the CHO CD Efficient Feed manual. Cultures with a viability >98% were used to seed initial fed experiments at $3x10^5$ viable cells/mL in a 50 mL working volume in 250 mL polycarbonate Erlenmeyer flasks with vented caps (Corning, Wiesbaden, Germany). From the initial feeding experiments, Condition 3 was used for all future fed-batch cultures and for generating samples analysed during this study. This adopted feeding strategy consisted of a 15% (v/v) supplementation of CHO CD Efficient Feed B to the CD CHO starting volume immediately on Day 0, followed by 10% (v/v) supplementation on Day 3 and Day 6. Four biological replicates of each culture process (batch and fed-batch) were seeded at $3x10^5$ viable cells/mL from 20 mL cultures with a culture viability ≥ 98.5% in 120 mL CD CHO starting working volume in 500 mL polycarbonate Erlenmeyer flasks with vented caps (Corning). Viable cell concentrations and culture viability were determined daily using a Vi-CELL XR Cell Viability Analyzer (Beckman Coulter, Inc., Brea, CA, USA) on 1 mL culture samples.

2.2 Sampling from Cell Cultures and Subsequent RNA Extraction

Samples of 1×10^7 viable cells were taken from each flask after 96 hours (day 4) and 144 hours (day 7) of culture and total RNA immediately extracted using the commercially available mirVana miRNA Isolation Kit (ThermoFisher Scientific, MA, USA) and treated with RapidOut DNA Removal Kit (ThermoFisher Scientific, MA, USA). The RNA quantity and quality were determined using a NanoDrop ND-1000 (ThermoFisher Scientific) instrument and the integrity of RNA assessed by standard denaturing agarose gel electrophoresis.

2.3 lncRNA and Coding RNA Microarray and Data Analysis

2.3.1 Microarray Details

Three of the 4 cultures were selected for analysis. Analysis of extracted RNA for coding and lncRNAs was undertaken using the commercially available ArrayStar Mouse lncRNA Microarray V3.0 (Rockville, MD, USA). RNA labelling and array hybridization were performed

according to the Agilent One-Color Microarray-Based Gene Expression Analysis protocol (Agilent Technologies, Santa Clara, CA, USA) with minor modifications. Ribosomal RNAs were removed from total RNA using the mRNA-ONLY Eukaryotic mRNA Isolation Kit, (Epicentre, Madison, WI, USA). Each sample was then amplified and transcribed into fluorescent cRNA along the entire length of the transcripts without 3' bias utilizing a mixture of oligo(dT) and random primers using the Arraystar Flash RNA Labelling Kit (Arraystar). The labelled cRNAs were purified by RNeasy Mini Kit (Qiagen, Hilden, Germany). The concentration and specific activity of the labelled cRNAs were determined by NanoDrop ND-1000 (ThermoFisher Scientific). A total of 1 μg of each labelled cRNA was then fragmented by adding 5 μL of 10 × Blocking Agent and 1 μL of 25 × Fragmentation Buffer before heating at 60°C for 30 min. Finally, 25 μL of 2 × GE Hybridization buffer was added to dilute the labelled cRNA. A sample of 50 μL of the hybridization solution was then dispensed into the gasket slide and assembled to the LncRNA expression microarray slide. The slides were incubated for 17 hours at 65°C in a Microarray Hybridization Oven (Agilent Technologies). The hybridized arrays were then washed, fixed and scanned using the G2505C DNA Microarray Scanner (Agilent Technologies).

2.3.2 Microarray Data Analysis

Agilent Feature Extraction software (version 11.0.1.1) was used to analyse acquired array images. Quantile normalization and subsequent data processing were performed using the GeneSpring GX v12.1 software package (Agilent Technologies). After quantile normalization of the raw data, lncRNAs and mRNAs that were present in ≥3 of 12 samples were selected for further data analysis. Raw *p*-values were calculated by unpaired t-test, then the differentially expressed lncRNAs and mRNAs with statistical significance between compared groups were filtered for a FC ≥ 2 and a false discovery rate (FDR) ≤ 0.10 calculated by the Benjamini–Hochberg procedure. GO analysis was performed using the Bioconductor package topGO [60] and with a raw *p*-value cut-off of 0.05 calculated by Fisher's exact test, subsequently filtered for an FDR ≤ 0.10. Pathway analysis for differentially expressed mRNAs was performed based on the KEGG database (http://www.genome.jp/kegg) with a raw *p*-value cut-off of 0.05 calculated by Fisher's exact test, subsequently filtered for an FDR ≤ 0.10.

2.4 RT-qPCR Validation of lncRNAs Identified as Differentially Expressed by Microarray

Primers for RT-qPCR were designed using OligoPerfect Designer (ThermoFisher Scientific) and synthetized by Eurofins Scientific (Luxembourg) (described in Supplementary Table 1). RT-qPCR reactions were conducted using a Mastercycler EP Realplex instrument (Eppendorf) following the QuantiFast SYBR Green RT-PCR Kit protocol (Qiagen). Specificity of amplification was checked by the generation of T_m curves and by analysis of the reaction products by 2% agarose gel electrophoresis to confirm the presence of a single amplicon of

the expected size. The results were analysed applying the standard ΔCt method and normalized to GAPDH, β-actin and B2M housekeeping gene expression.

3. RESULTS

3.1 Growth Characteristics and Sampling of CHO-S Cells Throughout Batch and Fed-Batch Culture for (Non)-Coding RNA Analysis

The steps followed to generate the results presented in this work are summarized in Fig. 1. Initially, growth comparisons and supplementation testing were undertaken using CD CHO and the associated commercially available feds (Efficient Feds A and B). From these preliminary experiments, we selected Efficient Feed B Liquid Nutrient Supplement for experimental Fed-batch cultures as this gave the highest viable cell concentrations across a 10-day process from the feed strategies investigated (data not shown). Cells under Fed or Batch culture grew almost identically up until day 4 of culture whereupon their growth characteristics differed (Fig. 2A). Whilst growth of the Batch cultures slowed dramatically after day 4, with a mean peak viable cell concentration on day 6 of approximately 1 x 10^7 cells/mL, in the Fed-batch cultures growth and proliferation continued until day 6 where a mean peak viable cell concentration of almost 2 x 10^7 cells/mL was obtained (Fig. 2A). After day 6 the viable cell number and viability of both cultures decreased with time, however while culture viability rapidly declined in the Batch culture such that by day 9 culture viability was close to 0%, in the Fed-batch mode cultures had a mean viability of 83.6% at day 10 (Fig. 2A). The first-time point at which samples were harvested was Day 4, when both types of culture had grown in a similar way and were still in the exponential growth phase (late exponential for batch), whilst the second time point at which cells were harvested for RNA analysis was on Day 7, corresponding to the end of the stationary phase/first day of decline for both cultures. In the case of the Fed-batch cultures, these two sampling points were also 24 hours after addition of the Efficient Feed B.

3.2 Microarray Analysis of mRNA and lncRNA Transcripts in Batch and Fed-Batch Culture

3.2.1 Differential expression of lncRNAs and mRNAs in Fed-batch v Batch comparisons

The Arraystar Mouse LncRNA Microarray V3.0 used in this study is based on publicly available databases and publications, allowing for the potential simultaneous surveillance of 35923 lncRNAs and 24881 coding transcripts. The lncRNAs collection is based on the NCBI Refseq, UCSC Known Gene 6.0, Ensembl 38.71, Fantom3, RNAdb 2.0, NRED databases, a number of literature publications, T-UCRs, and evolutionary constrained lncRNAs [61–68]. Positive probes for housekeeping genes and negative probes are printed onto the array for hybridization quality control. After quantile normalization of the raw data, 24603 unique

lncRNAs and 19617 mRNAs were selected for analysis. Firstly, we implemented Principal Component Analysis (PCA) on the log2 transformed intensities and plotted the first 3 components against each other, where lncRNAs and mRNAs showed very comparable results, especially where for the first two components samples from Batch cultures at both days grouped closely compared to Fed-batch samples, which showed a wider separation even between the same condition (Supplementary Figure 1). Then we applied hierarchical clustering, arranging samples into groups based on their averaged log-normalized expression levels to show the relationships among gene expression patterns of samples for both lncRNAs and mRNAs (Fig. 2B). We continued performing differential expression analysis (DE) where we compared Fed-batch against Batch at Day 4, Fed-batch against Batch at Day 7, Fed-batch at Day 7 against Fed-batch at Day 4 and Batch at Day 7 against Batch at Day 4. At this point, the expression of a selected number of the identified DE transcripts was confirmed by RT-qPCR, that showed these transcripts were present but that lower fold-changes were observed between samples than from the array analysis (Supplementary Table 1). The genes identified as DE were then filtered based on a threshold fold change (FC) ≥2 and a false discovery rate (FDR) ≤ 0.10 (Supplementary Tables 2 & 3). When we compared Batch Day 7 to Batch Day 4 samples, we found surprisingly low numbers of mRNAs or lncRNAs where the expression changed beyond the set thresholds, with 0 mRNAs being up and 19 down regulated between the two days whilst for the lncRNAs there were no transcripts that changed about the thresholds set (Fig. 2C). To check whether this lack of identified transcripts changing was due to the FC ≥ 2 threshold, we repeated the analysis for FC ≥ 1.5 and FC ≥ 1.25, FDR ≤ 0.10 (Supplementary Tables 4-11). For a FC ≥ 1.25, 99 mRNAs were down and 30 were up-regulated while none met the threshold for the lncRNAs. We then carried on with the FC ≥ 2, FDR ≤ 0.10 thresholds for the remaining comparisons. Fed-batch Day 7 vs Day 4 revealed 693 mRNAs up-regulated and 421 down-regulated whilst for the lncRNAs we found 545 up-regulated and 200 down-regulated genes (Fig. 2C). When we compared Fed-batch against Batch at Day 4 and Day 7, 1048 and 1875 lncRNAs were up-regulated respectively in addition to 856 down-regulated at day 4 and 1018 down-regulated at day 7. For the same comparisons we saw more down-regulated mRNAs at day 4 (1538) and day 7 (1397) while there were 716 up-regulated mRNAs at day 4 and 1121 at day 7 (Fig. 2C).

3.3 Identification of lncRNAs Differentially Expressed as Potential Engineering Targets for Modulation of Cell Growth

As one of the aims of this work was to identify new transcripts for manipulation in CHO, we reduced the list of potential targets to a manageable group for further experimental validation. Firstly, we aligned all the 60 nucleotide mouse probes corresponding to the differentially expressed genes identified for each comparison against the Chinese hamster (CH) genome

using the discontiguous megablast algorithm to check how many transcripts had an annotation in CHO. Only 16-28% of the probes corresponding to differentially expressed lncRNAs had a matching transcript in CH, as opposed to 58-80% of the mRNAs (Supplementary Table 12). This is likely due to the poor annotation of the CH transcriptome compared to mouse, where the number of described lncRNAs is significantly higher. If we consider the ENSEMBL 91 database release alone, 2563 lncRNAs and 446 pseudogenes are listed for Chinese hamster as opposed to 9308 lncRNAs and 12363 pseudogenes in mouse. To address this issue we complemented this approach with literature and databases mining (NCBI, ENSEMBL, lncRNAdb, LNCipedia) to identify differentially expressed lncRNAs already described in mouse but not in Chinese hamster, revealing genomic positions that align with potential lncRNAs (Supplementary Figures 2-4). The identified sequences were then examined using the Rfam database [69] to assess their resemblance to existing non-coding RNA families. A list of potential lncRNAs found using the described approaches is reported in Table 1, along with literature references describing their biological function.

The role of lncRNAs in diseases is well established, especially linked to cell proliferation in cancer [70] and several of those identified here have been investigated in such systems. NEAT1 (nuclear-enriched abundant transcript 1) and MALAT1 (also known as NEAT2), found to be up-regulated at both days in our Fed-batch cultures (Table 1), are among the most well characterized lncRNAs and have been reported to promote cell proliferation through regulation of gene expression at the nuclear level. MALAT1 is a mostly un-spliced transcript around 6.7 kb in mouse, with a long half-life due to the tRNA-like structure adopted at the 3' end [37]. This structure is cleaved to generate a 61 nt mascRNA (MALAT1-associated small cytoplasmic RNA) exported to the cytoplasm where it is subjected to canonical CCA nucleotides addition and accumulates in the cytoplasm [71,72]. We were able to identify the JH002628.1 genomic region in Chinese hamster using the abovementioned complementary approaches and we could recognize all of its main conserved domains (Supplementary Figure 2). The structure of this genomic region is conserved in the most recent CHOK1GS_HDv1 genome and resembles the pattern found in human and mouse, where close to MALAT1 is the NEAT1 locus. The NEAT1 locus is regulated by alternative 3′ end processing, where the primary transcript can be cleaved and polyadenylated to generate a 3.2 kb long MENε isoform or cleaved and non-polyadenylated to generate the 20.8 kb long MENβ isoform [37]. We were able to identify a predicted 3.2 kb long non-coding transcript (NCBI Ref XR_478750.2) arising from a locus (LOC103159497, C_griseus_v1.0) with similar primary structure to both the 3.2 kb and the 20.8 kb isoforms (E-value 0.0, Identity 80%, algorithm megablast). MALAT1 and NEAT1 both localize to the paraspeckles stress-responsive nuclear bodies in the cell [73], where they are reported to influence the splicing machinery [70,74] and the DNA repair machinery [75]. The

Plasmacytoma Variant Translocation 1 (PVT1) lncRNA is considered a biomarker for various cancers due to its ability to promote cell proliferation with a range of proposed mechanisms [76–79]. We identified a predicted non-coding RNA in Chinese hamster (NCBI Ref XR_478426.2) which contained the PVT1_3 RFAM domain and was up-regulated at both Day 4 (FDR = 1.2E-01) and Day 7 in the Fed-batch compared to Batch (Table 1).

3.4 Comparison of mRNAs with Existing Datasets

We compared our Fed-batch against Batch dataset with previous work to identify coding genes related to growth in CHO. Although all these works used different approaches and cell lines, we saw the opportunity to find common patterns of expression in CHO across various conditions. The first dataset compared was Clarke 2011 [80], where we observed 10 of their reported down-regulated genes at Day 4 and 7 at Day 7 of our dataset (Supplementary Table 13). On comparing our data to Clarke 2012 [81], only the kinesin family member C1 (KIFC1) gene, up-regulated at both Day 4 and Day 7, was observed to behave in the same way. Lastly, when we considered the translatome analysis in Courtes 2013 we identified 3 genes at Day 4 and 8 genes at Day 7 among our down-regulated transcripts, together with 1 gene at Day 4 and 2 genes at Day 7 among the up-regulated transcripts [82]. Interestingly, a group of the common genes between the selected works and our dataset (CDC20, MAD2L1, MCM7, MCM4, GTF2H4) are involved in cell cycle and DNA replication, supporting the findings in the pathway enrichment analysis (described in section 3.5) [83–85] . In addition, we found single genes in consistently enriched pathways in our pathway analysis such as HNRNPC, involved in RNA molecule binding or LGMN, participating in protein degradation in the lysosome [86]. Interestingly, HNRNPs proteins are well known to bind and mediate the functions of lncRNAs, and HNRNPC in particular has been reported to interact with MALAT1 in a tightly regulated N^6-methyladenosine-dependent manner [87,88].

3.5 GO analysis and Pathway Enrichment

We performed GO term analysis on the differentially expressed mRNAs with an FDR cut-off of 0.10 calculated by the BH method, followed by pathway enrichment analysis based on the KEGG database (Supplementary Tables14-29), allowing the determination of the significantly enriched biological pathways filtered for an FDR ≤ 0.10 and grouped based on KEGG class annotation (Table 2). The pathway enrichment identified 22 pathways containing down-regulated genes at Fed-batch vs Batch Day 4 and 24 at Fed-batch vs Batch Day 7 while only 1 contained up-regulated genes (Lysosome, ID: 04142) at either day. This is most likely a reflection of the lower number of up-regulated genes compared to down-regulated (Figure 3). The Metabolism domain includes 7 enriched pathways at Day 4 while only 1 of these was still enriched at Day 7, suggesting a central role of metabolism together with p53 signalling

predominantly during the exponential growth phase. On the contrary, towards Day 7 we see the prevalent enrichment of pathways related to translation regulation and RNA interaction at different levels, from transport to splicing. The most evident pattern of enrichment between Day 4 and Day 7 is represented by the Replication and Repair class, where the majority of the pathways involved in genome maintenance and diverse repair mechanisms are consistently enriched, indicating an early and sustained regulation of these genes throughout culture. DNA damage is reported to stimulate the expression of NEAT1 and, together with MALAT1, to promote the formation of paraspeckles, which regulate alternative splicing and promote proliferation [73,75,89]. We then compared our results with relevant KEGG pathway enrichment datasets available for CHO [90] and found 4 common pathways, which were enriched exclusively at Day 7 in our dataset: RNA Transport (ID: 03013), mRNA surveillance (ID: 03015), RNA degradation (ID: 03018), Spliceosome (ID: 03040). Taken together, these results suggest the importance of cell cycle and genome repair mechanism control likely due to the high proliferation of the fed-batch system. In addition, the active regulation of RNA transport, RNA maintenance and splicing seems to be particularly important towards the later stages of our Fed-batch cultures.

4. DISCUSSION

By taking advantage of a comprehensive commercially available mouse microarray containing 35923 lncRNAs and 24881 mRNAs, we have provided the first mapping of the CHO lncRNA landscape, together with the coding transcriptome. Previous reports have shown that more than 70% of the assembled CHO transcriptome is similar to mouse (*Mus musculus*) and closely related to rat (*Rattus norvegicus*) transcriptomes [58] suggesting that this approach was likely to be valid for lncRNAs as well. Due to the species and tissue-specificity of lncRNAs compared to mRNAs, the number of detectable lncRNAs in CHO is likely to be lower than the 35923 probes included in the array. Nevertheless, using this approach we were able to detect 24603 lncRNAs (68.5% of the total probes) and 19617 mRNAs (78.8% of the total probes), and found that t several hundreds of lncRNAs exhibit changing expression profiles on different days of culture and between Batch and Fed-batch culture in a model CHO-S system. This was especially true for the Fed system, where comparing Day 4 and Day 7 we observed 1114 differentially expressed mRNAs and 745 lncRNAs, as opposed to the Batch, were we saw only 19 differentially expressed mRNAs and no lncRNA for the same comparison, suggesting a prevalent variability induced by the Fed supplement as compared to time only.

Among the differentially expressed genes, between 16-28% of the lncRNAs probes had a matching transcript in CH, as opposed to 58-80% of the mRNAs. This required a specific approach where the Chinese hamster genome, literature search and databases mining were

combined to detect lncRNAs differentially expressed in our system with an established biological function. Within these lncRNAs, we focused on MALAT1, NEAT1 and PVT1 to provide a comparison between the mouse gene and the Chinese hamster putative homologues. The number of lncRNAs with a fully understood role in the cell remains small, however these three non-coding genes are among the most well characterized [91]. MALAT1 and NEAT1 in particular are associated with increased proliferation and participate in the regulation of alternative splicing and DNA repair, which we found to be strongly regulated in our pathway enrichment analysis. This suggests a potential role for these lncRNAs in CHO, although further experimental studies on the single genes are now required to assess the actual effects on the cell under different conditions.

Several approaches to investigate and confirm the functional annotation of lncRNAs in other organisms have been described, including the perturbation of lncRNA expression by overexpression, knockout or knockdown [92,93], in addition to complementary strategies [39]. Future developments in CHO will have to proceed with a mix of functional prediction tools to assess the properties of the transcriptome and evaluate the degree of conservation with other species [94–96] and of targeted approaches to ameliorate the annotation and propose mechanisms of action for the single transcripts [97–100]. Further, the data reported here is for both lncRNAs and mRNAs and hence will allow investigators to further probe the relationships between the expression and regulation of these two classes of RNA. As the majority of the lncRNAs reported in the literature are discussed and related to human or model organism systems, our work aimed at unveiling the role of lncRNAs in CHO under industrially relevant conditions to identify new targets for manipulation to sustain proliferation. Examples of successful cell engineering of lncRNAs to selectively enhance translation [101] and product yield [57] have already been reported in CHO, demonstrating the potential of manipulation of lncRNAs for enhancing industrial processes. Moreover, since it was reported up to 15% of the total ribosome occupancy can be occupied by a single recombinant mRNA, the intrinsic characteristics of lncRNAs place them as ideal candidates for cell line engineering of protein production cell factories, as they do not add any translational burden on top of the coding gene of interest [102].

Our work has identified potential lncRNA targets differentially expressed in Fed-batch compared to Batch culture from which we selected a group of molecules to be experimentally studied (Table 1). In addition to the expression of lncRNAs we also looked at expression of mRNAs (coding transcripts) and found a consistent change in differentially expressed mRNAs when comparing batch and fed-batch cultures. Pathway enrichment analysis (Figure 3) underlined the importance of genes involved in cell cycle and genome maintenance pathways along with the regulation of lysosome formation as potential targets for cell engineering to

enhance proliferation. Our approach allowed the identification of previously undescribed lncRNAs in CHO along with mRNAs to identify the connections between them and compared these with existent literature. This network of reciprocal interactions is beginning to be unveiled in other organisms [103–105] and our work will help pave the way for the definition of new layers of regulation involving single transcripts or even entire pathways in CHO.

5. CONCLUSIONS

He we report on the lncRNA landscape and how this changes in CHO cells, presenting a full dataset of those lncRNAs present as determined from an array study and how these change through a Batch and Fed-batch culture and between the two culture systems. From analysis of the data, we have determined those lncRNAs whose expression changes the most between 2 days in culture and between fed and batch culture that are attractive targets for cell engineering. This resource will now provide the community with the opportunity to undertake functional validation studies by undertaking single or multiple knock downs/outs, or by the up-regulation of target lncRNAs, and determine the impact on growth, and productivity, characteristics of CHO cells. Ultimately, we anticipate such a resource will be incorporated into wider genome analysis datasets including coding mRNAs and other non-coding RNAs to develop a wider appreciation of the role of RNAs in controlling recombinant CHO cell line growth and productivity characteristics.

6.0 Acknowledgements

The authors acknowledge the European Commission for funding for DV and supporting this work (EC - Horizon 2020 MSCA ITN 2014 - 642663). DV undertook experiments, analysed the data and co-wrote the paper; CMS devised the project and with DV devised experiments, aided in experimental design, analysed data and co-wrote the paper. The authors acknowledge the Noun Project for sharing the icons used in the graphical abstract.

Conflict of Interest Statement

The authors declare no commercial or financial conflict of interest.

REFERENCES

[1] G. Walsh, *Nat. Biotechnol.* **2014**, 32, 992–1000.
[2] J. Y. Kim, Y. G. Kim, G. M. Lee, *Appl. Microbiol. Biot.* 2012. 93, 917.
[3] K. P. Jayapal, K. F. Wlaschin, W. S. Hu, M. G. S. Yap, *Chem. Eng. Prog.* **2007**, 103, 40.
[4] J. Dumont, D. Euwart, B. Mei, S. Estes, R. Kshirsagar, *Crit. Rev. Biotechnol.* **2015**, 1, 1.
[5] R. Kunert, D. Reinhart, *Appl. Microbiol. Biotechnol.* **2016**, 100, 3451.

[6] X. Pan, M. Streefland, C. Dalm, R. H. Wijffels, D. E. Martens, *Cytotechnology*, **2017**, 69, 39.
[7] Y. Durocher, M. Butler, *Curr. Opin. Biotechnol.*, **2009**, vol. 20, no. 6, pp. 700–707.
[8] D. C. F. Wong, K. T. K. Wong, Y. Y. Lee, P. N. Morin, C. K. Heng, M. G. S. Yap, *Biotechnol. Bioeng.* **2006**, 94, 373.
[9] J. F. Povey, C. J. O'Malley, T. Root, E. B. Martin, G. A. Montague, M. Feary, C. Trim, D. A. Lang, R. Alldread, A. J. Racher, C. M. Smales, *J. Biotechnol.* **2014**, 184, 84.
[10] D. Reinhart, L. Damjanovic, C. Kaisermayer, R. Kunert, *Appl. Microbiol. Biotechnol.* **2015**, 99, 4645.
[11] C. S. Alves, T. M. Dobrowsky, in *Heterologous Protein Production in CHO Cells: Methods and Protocols* (Ed: P. Meleady), Springer, New York **2017**, Ch. 1.
[12] L. Jossé, J. Xie, C. G. Proud, C. M. Smales, *Biochem. J.* **2016**, 473, 4651.
[13] L. Josse, C. M. Smales, M. F. Tuite, *Biotechnol. Bioeng.* **2010**, 105, 556.
[14] G. M. Lee, *J. Biotechnol.* **2008**, 136, S130.
[15] C. Ferrara, P. Brünker, T. Suter, S. Moser, U. Püntener, P. Umaña, *Biotechnol. Bioeng.* **2006**, 93, 851.
[16] N. Yamane-Ohnuki, S. Kinoshita, M. Inoue-Urakubo, M. Kusunoki, S. Iida, R. Nakano, M. Wakitani, R. Niwa, M. Sakurada, K. Uchida, K. Shitara, M. Satoh, *Biotechnol. Bioeng.* **2004**, 87, 614.
[17] L. Malphettes, Y. Freyvert, J. Chang, P.-Q. Liu, E. Chan, J. C. Miller, Z. Zhou, T. Nguyen, C. Tsai, A. W. Snowden, T. N. Collingwood, P. D. Gregory, G. J. Cost, *Biotechnol. Bioeng.* **2010**, 106, 774.
[18] J.-X. Bi, J. Shuttleworth, M. Al-Rubeai, *Biotechnol. Bioeng.* **2004**, 85, 741.
[19] M. Fussenegger, X. Mazur, J. E. Bailey, *Biotechnol. Bioeng.* **1997**, 55, 927.
[20] N. S. Kim, G. M. Lee, *Biotechnol. Bioeng.*, **2002**, 78, 217.
[21] S. S. Choi, W. J. Rhee, E. J. Kim, T. H. Park, *Biotechnol. Bioeng.* **2006**, 95, 459.
[22] S. O. Hwang, G. M. Lee, *J. Biotechnol.* **2009**, 139, 89.
[23] J. S. Lee, T. K. Ha, J. H. Park, G. M. Lee, *Biotechnol. Bioeng.* **2013**, 110, 2195.
[24] M. Mason, B. Sweeney, K. Cain, P. Stephens, S. T. Sharfstein, *Biotechnol. Prog.* **2012**, 28, 846.
[25] L. P. Pybus, G. Dean, N. R. West, A. Smith, O. Daramola, R. Field, S. J. Wilkinson, D. C. James, *Biotechnol. Bioeng.* **2014**, 111, 372.
[26] Y. B. Johari, S. D. Estes, C. S. Alves, M. S. Sinacore, D. C. James, *Biotechnol. Bioeng.* **2015**, 112, 2527.
[27] L. Thoring, D. A. Wustenhagen, M. Borowiak, M. Stech, A. Sonnabend, S. Kubick, *PLoS ONE* **2016**, 11, e0163670.
[28] E. J. Mead, L. M. Chiverton, S. K. Spurgeon, E. B. Martin, G. A. Montague, C. M. Smales, T. von der Haar, *PLOS ONE* **2012**, 7, e47422.
[29] E. J. Mead, R. J. Masterton, M. Feary, O. Obrezanova, L. Zhang, R. Young, C. M. Smales, *Biochem. J.* **2015**, 472, 261.
[30] F. Stiefel, S. Fischer, A. Sczyrba, K. Otte, F. Hesse, *J. Biotechnol.* **2016**, 225, 31.
[31] J. A. H. Bort, M. Hackl, H. Höflmayer, V. Jadhav, E. Harreither, N. Kumar, W. Ernst, J. Grillari, N. Borth, *Biotechnol. J.* **2012**, 7, 500.
[32] S. Fischer, K. F. Marquart, L. A. Pieper, J. Fieder, M. Gamer, I. Gorr, P. Schulz, H. Bradl, *Biotechnol. Bioeng.* **2017**, 114, 1495.
[33] N. Barron, N. Kumar, N. Sanchez, P. Doolan, C. Clarke, P. Meleady, F. O'Sullivan, M. Clynes, *J. Biotechnol.* **2011**, 151, 204.
[34] M. Hackl, N. Borth, J. Grillari, *Trends Biotechnol.* **2012**, 30, 405.
[35] S. Djebali, C. A. Davis, A. Merkel, A. Dobin, T. Lassmann, A. Mortazavi, A. Tanzer, J. Lagarde, W. Lin, F. Schlesinger, C. Xue, G. K. Marinov, J. Khatun, B. A. Williams, C. Zaleski, J. Rozowsky, M. Röder, F. Kokocinski, R. F. Abdelhamid, T. Alioto, I. Antoshechkin, M. T. Baer, N. S. Bar, P. Batut, K. Bell, I. Bell, S. Chakrabortty, X. Chen, J. Chrast, J. Curado, T. Derrien, J. Drenkow, E. Dumais, J. Dumais, R. Duttagupta, E. Falconnet, M. Fastuca, K. Fejes-Toth, P. Ferreira, S. Foissac, M. J. Fullwood, H. Gao,

D. Gonzalez, A. Gordon, H. Gunawardena, C. Howald, S. Jha, R. Johnson, P. Kapranov, B. King, C. Kingswood, O. J. Luo, E. Park, K. Persaud, J. B. Preal, P. Ribeca, B. Risk, D. Robyr, M. Sammeth, L. Schaffer, L. H. See, A. Shahab, J. Skancke, A. M. Suzuki, H. Takahashi, H. Tilgner, D. Trout, N. Walters, H. Wang, J. Wrobel, Y. Yu, X. Ruan, Y. Hayashizaki, J. Harrow, M. Gerstein, T. Hubbard, A. Reymond, S. E. Antonarakis, G. Hannon, M. C. Giddings, Y. Ruan, B. Wold, P. Carninci, R. Guigó, T. R. Gingeras, *Nature* **2012**, 489, 101.

[36] P. Kapranov, J. Cheng, S. Dike, D. A. Nix, R. Duttagupta, A. T. Willingham, P. F. Stadler, J. Hertel, J. Hackermuller, I. L. Hofacker, I. Bell, E. Cheung, J. Drenkow, E. Dumais, S. Patel, G. Helt, M. Ganesh, S. Ghosh, A. Piccolboni, V. Sementchenko, H. Tammana, T. R. Gingeras, *Science* **2007**, 316, 1484–1488.

[37] J. E. Wilusz, *Biochim. Biophys. Acta* **2016**, 1859, 128.

[38] J. T. Y. Kung, D. Colognori, J. T. Lee, *Genetics* **2013**, 193, 651.

[39] K. Kashi, L. Henderson, A. Bonetti, P. Carninci, *Biochim. Biophys. Acta* **2016**, 1859, 3.

[40] J. G. Betancur, *Biochim. Biophys. Acta* **2016**, 1859, 93.

[41] T. Trimarchi, E. Bilal, P. Ntziachristos, G. Fabbri, R. Dalla-Favera, A. Tsirigos, I. Aifantis, *Cell* **2014**, 158, 593.

[42] I. Gonzalez, R. Munita, E. Agirre, T. A. Dittmer, K. Gysling, T. Misteli, R. F. Luco, *Nat. Struct. Mol. Biol.* **2015**, 22, 1.

[43] C. Carrieri, L. Cimatti, M. Biagioli, A. Beugnet, S. Zucchelli, S. Fedele, E. Pesce, I. Ferrer, L. Collavin, C. Santoro, A. R. R. Forrest, P. Carninci, S. Biffo, E. Stupka, S. Gustincich, *Nature* **2012,** 491, 454.

[44] S. Geisler, J. Coller, *Nat. Rev. Mol. Cell Biol.* **2013**, 14, 699.

[45] A. M. Schmitt, H. Y. Chang, *Cancer Cell* **2016**, 29, 452.

[46] R. B. T. Perry, I. Ulitsky, *Development* **2016**, 143, 3882.

[47] Y. Tay, J. Rinn, P. P. Pandolfi, *Nature* **2014**, 505, 344.

[48] T. Chujo, T. Yamazaki, T. Hirose, *Biochim. Biophys. Acta* **2016**, 1859, 1.

[49] C. A. Melo, J. Drost, P. J. Wijchers, H. van de Werken, E. de Wit, J. A. F. O. Vrielink, R. Elkon, S. A. Melo, N. Léveillé, R. Kalluri, W. de Laat, R. Agami, *Mol. Cell* **2013**, 49, 524.

[50] M. Koziol, J. Rinn, *Curr. Opin. Genet. Dev.* **2010**, 20, 142.

[51] J. A. Gomez, O. L. Wapinski, Y. W. Yang, J. F. Bureau, S. Gopinath, D. M. Monack, H. Y. Chang, M. Brahic, K. Kirkegaard, *Cell* **2013**, 152, 743.

[52] A. Keniry, D. Oxley, P. Monnier, M. Kyba, L. Dandolo, G. Smits, W. Reik, *Nat. Cell Biol.* **2012**, 14, 659.

[53] F. Rashid, A. Shah, G. Shan, *Genomics Proteomics Bioinformatics*, **2016**, 14, 73.

[54] B. Liu, L. Sun, Q. Liu, C. Gong, Y. Yao, X. Lv, L. Lin, H. Yao, F. Su, D. Li, M. Zeng, E. Song, *Cancer Cell* **2015**, 27, 370.

[55] P. Wang, Y. Xue, Y. Han, L. Lin, C. Wu, S. Xu, Z. Jiang, J. Xu, Q. Liu, *Science* **2014**, 344, 310.

[56] L. Patrucco, A. Chiesa, M. F. Soluri, F. Fasolo, H. Takahashi, P. Carninci, S. Zucchelli, C. Santoro, S. Gustincich, D. Sblattero, D. Cotella, *Gene* **2015**, 569, 287.

[57] S. Zucchelli, L. Patrucco, F. Persichetti, S. Gustincich, D. Cotella, *Comput. Struct. Biotechnol. J.* **2016**, 14, 404.

[58] J. Becker, M. Hackl, O. Rupp, T. Jakobi, J. Schneider, R. Szczepanowski, T. Bekel, N. Borth, A. Goesmann, J. Grillari, C. Kaltschmidt, T. Noll, A. Pühler, A. Tauch, K. Brinkrolf, *J. Biotechnol.* **2011**, 156, 227.

[59] X. Xu, H. Nagarajan, N. E. Lewis, S. Pan, Z. Cai, X. Liu, W. Chen, M. Xie, W. Wang, S. Hammond, M. R. Andersen, N. Neff, B. Passarelli, W. Koh, H. C. Fan, J. Wang, Y. Gui, K. H. Lee, M. J. Betenbaugh, S. R. Quake, I. Famili, B. O. Palsson, J. Wang, *Nat. Biotechnol*. **2011**, 29, 735.

[60] A. Alexa, J. Rahnenfuhrer, topGO: Enrichment Analysis for Gene Ontology. R package version 2.28.0. 2016.

[61] F. Cunningham, M. R. Amode, D. Barrell, K. Beal, K. Billis, S. Brent, D. Carvalho-Silva, P. Clapham, G. Coates, S. Fitzgerald, L. Gil, C. G. Giron, L. Gordon, T. Hourlier, S. E. Hunt, S. H. Janacek, N. Johnson, T. Juettemann, A. K. Kahari, S. Keenan, F. J. Martin, T. Maurel, W. McLaren, D. N. Murphy, R. Nag, B. Overduin, A. Parker, M. Patricio, E. Perry, M. Pignatelli, H. S. Riat, D. Sheppard, K. Taylor, A. Thormann, A. Vullo, S. P. Wilder, A. Zadissa, B. L. Aken, E. Birney, J. Harrow, R. Kinsella, M. Muffato, M. Ruffier, S. M. Searle, G. Spudich, S. J. Trevanion, A. Yates, D. R. Zerbino, P. Flicek, *Nucleic Acids Res.* **2015**, 43, D662.
[62] K. D. Pruitt, T. Tatusova, D. R. Maglott, *Nucleic Acids Res.* **2007**, 35, D61.
[63] P. P. Amaral, M. B. Clark, D. K. Gascoigne, M. E. Dinger, J. S. Mattick, *Nucleic Acids Res.* **2011**, 39, D146.
[64] G. Bejerano, M. Pheasant, I. Makunin, S. Stephen, W. J. Kent, J. S. Mattick, D. Haussler, *Science* **2004**, 304, 1321.
[65] J. L. Rinn, M. Kertesz, J. K. Wang, S. L. Squazzo, X. Xu, S. A. Brugmann, L. H. Goodnough, J. A. Helms, P. J. Farnham, E. Segal, H. Y. Chang, *Cell* **2007**, 129, 1311.
[66] T. R. Mercer, M. E. Dinger, S. M. Sunkin, M. F. Mehler, J. S. Mattick, *Proc. Natl. Acad. Sci. U. S. A.* **2008**, 105, 716.
[67] M. Guttman, I. Amit, M. Garber, C. French, M. F. Lin, D. Feldser, M. Huarte, O. Zuk, B. W. Carey, J. P. Cassady, M. N. Cabili, R. Jaenisch, T. S. Mikkelsen, T. Jacks, N. Hacohen, B. E. Bernstein, M. Kellis, A. Regev, J. L. Rinn, E. S. Lander, *Nature* **2009**, 458, 223.
[68] D. A. Benson, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, D. L. Wheeler, *Nucleic Acids Res.* **2004**, 32, D23.
[69] I. Kalvari, J. Argasinska, N. Quinones-Olvera, E. P. Nawrocki, E. Rivas, S. R. Eddy, A. Bateman, R. D. Finn, A. I. Petrov, *Nucleic Acids Res.* **2018**, 46, D335.
[70] W. Sun, Y. Yang, C. Xu, J. Guo, *Cancer Genet.* **2017**, 216–217, 105.
[71] J. E. Wilusz, *Wiley Interdiscip. Rev. RNA*, **2015**, 6, 453.
[72] J. E. Wilusz, S. M. Freier, D. L. Spector, *Cell* **2008**, 135, 919.
[73] S. Nakagawa, T. Hirose, *Cell. Mol. Life Sci.* **2012**, 69, 3027.
[74] V. Tripathi, J. D. Ellis, Z. Shen, D. Y. Song, Q. Pan, A. T. Watt, S. M. Freier, C. F. Bennett, A. Sharma, P. A. Bubulya, B. J. Blencowe, S. G. Prasanth, K. V. Prasanth, *Mol. Cell* **2010**, 39, 925.
[75] C. Adriaens, L. Standaert, J. Barra, M. Latil, A. Verfaillie, P. Kalev, B. Boeckx, P. W. G. Wijnhoven, E. Radaelli, W. Vermi, E. Leucci, G. Lapouge, B. Beck, J. van den Oord, S. Nakagawa, T. Hirose, A. A. Sablina, D. Lambrechts, S. Aerts, C. Blanpain, J-C. Marine, *Nat. Med.* **2016**, 22, 861.
[76] S. Zhu, P. Shuai, C. Yang, Y. Zhang, S. Zhong, X. Liu, K. Chen, Q. Ran, H. Yang, Y. Zhou, S. Zhu, P. Shuai, C. Yang, Y. Zhang, S. Zhong, X. Liu, K. Chen, Q. Ran, H. Yang, Y. Zhou, *Oncotarget* **2017**, 8, 113174.
[77] F. Wang, J.-H. Yuan, S.-B. Wang, F. Yang, S.-X. Yuan, C. Ye, N. Yang, W.-P. Zhou, W.-L. Li, W. Li, S.-H. Sun, *Hepatology* **2014**, 60, 1278.
[78] Y.-Y. Tseng, A. Bagchi, *Mol. Cell. Oncol.* **2015**, 2, e974467.
[79] T. Colombo, L. Farina, G. Macino, P. Paci, *Biomed Res. Int.* **2015**, 2015, 304208.
[80] C. Clarke, P. Doolan, N. Barron, P. Meleady, F. O'Sullivan, P. Gammell, M. Melville, M. Leonard, M. Clynes, *J. Biotechnol.* **2011**, 155, 350.
[81] C. Clarke, M. Henry, P. Doolan, S. Kelly, S. Aherne, N. Sanchez, P. Kelly, P. Kinsella, L. Breen, S. F. Madden, L. Zhang, M. Leonard, M. Clynes, P. Meleady, N. Barron, *BMC Genomics* **2012**, 13, 656.
[82] F. C. Courtes, J. Lin, H. L. Lim, S. W. Ng, N. S. C. Wong, G. Koh, L. Vardy, M. G. S. Yap, B. Loo, D. Y. Lee, *J. Biotechnol.* **2013**, 167, 215.
[83] Y. Zhai, N. Li, H. Jiang, X. Huang, N. Gao, B. K. Tye, *Mol. Cell* **2017**, 67, 168.
[84] H. Yu, *Curr. Opin. Cell Biol.* **2002**, 14, 706.
[85] S. C. Buch, B. Diergaarde, T. Nukui, R. S. Day, J. M. Siegfried, M. Romkes, J. L.

Weissfeld, *Mol. Carcinog.* **2012**, 51, E11.
[86] E. Dall, H. Brandstetter, *Biochimie* **2016**, 122, 126.
[87] X. Sun, M. S. S. Haider Ali, M. Moran, *Biochem. J.* **2017**, 474, 2925.
[88] N. Liu, Q. Dai, G. Zheng, C. He, M. Parisien, T. Pan, *Nature* **2015**, 518, 560.
[89] C. Adriaens, J.-C. Marine, *Cell Cycle* **2017**, 16, 137.
[90] E. Harreither, M. Hackl, J. Pichler, S. Shridhar, N. Auer, P. P. Łabaj, M. Scheideler, M. Karbiener, J. Grillari, D. P. Kreil, N. Borth, *Biotechnol. J.* **2015**, 10, 1625.
[91] F. Kopp, J. T. Mendell, *Cell* **2018**, 172, 393.
[92] S. Zhu, W. Li, J. Liu, C.-H. Chen, Q. Liao, P. Xu, H. Xu, T. Xiao, Z. Cao, J. Peng, P. Yuan, M. Brown, X. S. Liu, W. Wei, *Nat. Biotechnol.* **2016**, 34, 1279.
[93] S. J. Liu, M. A. Horlbeck, S. W. Cho, H. S. Birk, M. Malatesta, D. He, F. J. Attenello, J. E. Villalta, M. Y. Cho, Y. Chen, M. A. Mandegar, M. P. Olvera, L. A. Gilbert, B. R. Conklin, H. Y. Chang, J. S. Weissman, D. A. Lim, *Science* **2017**, 355, 6320.
[94] B. Signal, B. S. Gloss, M. E. Dinger, *Trends Genet.* **2016**, 32, 620.
[95] J. Iwakiri, M. Hamada, K. Asai, *Biochim. Biophys. Acta* **2016**, 1859, 23.
[96] I. Ulitsky, *Nat. Rev. Genet.* **2016**, 17, 601.
[97] M. B. Clark, T. R. Mercer, G. Bussotti, T. Leonardi, K. R. Haynes, J. Crawford, M. E. Brunck, K.-A. L. Cao, G. P. Thomas, W. Y. Chen, R. J. Taft, L. K. Nielsen, A. J. Enright, J. S. Mattick, M. E. Dinger, *Nat. Methods* **2015**, 12, 339.
[98] T. R. Mercer, M. B. Clark, J. Crawford, M. E. Brunck, D. J. Gerhardt, R. J. Taft, L. K. Nielsen, M. E. Dinger, J. S. Mattick, *Nat. Protoc.* **2014**, 9, 89.
[99] F. Lai, E. Blumenthal, R. Shiekhattar, *Methods Enzymol.* **2016**, 573, 421.
[100] M. N. Cabili, M. C. Dunagin, P. D. McClanahan, A. Biaesch, O. Padovan-Merhar, A. Regev, J. L. Rinn, A. Raj, *Genome Biol.* **2015**, 16, 20.
[101] S. Zucchelli, L. Patrucco, F. Persichetti, S. Gustincich, D. Cotella, *Comput. Struct. Biotechnol. J.* **2016**, 14, 404.
[102] T. B. Kallehauge, S. Li, L. E. Pedersen, T. K. Ha, D. Ley, M. R. Andersen, H. F. Kildegaard, G. M. Lee, N. E. Lewis, *Sci. Rep.*, **2017**, 7, 40388.
[103] P. Han, C.-P. Chang, *RNA Biol.* **2015**, 12, 1094.
[104] A. E. Kornienko, P. M. Guenzl, D. P. Barlow, F. M. Pauler, *BMC Biol.* **2013**, 11, 59.
[105] A. C. Mallory, A. Shkumatava, *Biochimie* **2015**, 117, 3.

Table 1: Summary of the lncRNAs identified using the described approaches, consisting in a direct alignment of the differentially expressed gene probe against the Chinese hamster genome to literature search and RNA families conservation in RFAM. From left to right, the columns show the gene symbol, the NCBI accession reference, the fold-change measured at Day 4 and Day 7 with the corresponding FDRs, the ID of the non-coding RNA family found for the sequence in Rfam database with the corresponding ID and E-value and the references describing the biological function of the gene.

Gene symbol	NCBI Ref	FC Day 4	FC Day 7	FDR Day 4	FDR Day 7	Rfam Family	Rfam ID	E-value	References
MALAT1	JH002628.1	3.5	4.5	2.4E-02	1.0E-02	MALAT1	RF01871	7.5E-13	*Sun* 2017, *Wilusz* 2016
						mascRNA-menRNA	RF01684	7.0E-16	
MEG3	JH001208.1	2.9	2.8	5.9E-02	4.7E-02	MEG3_2	RF01872	3.5E-14	*He* 2017
MIAT	XM_007625231.2	2.6	0.0	2.6E+00	1.0E-02	MIAT_exon1	RF01874	7.0E-12	*Liao* 2016
NEAT1	XR_478750.2	6.1	3.3	3.9E-02	6.3E-02	NEAT1_1	RF01955	5.4E-18	*Adriaens* 2016, *Hirose* 2014
						NEAT1_2	RF01956	1.2E-17	
						NEAT1_3	RF01957	2.2E-19	
PVT1	XR_478426.2	2.5	3.5	1.2E-01	2.7E-02	PVT1_3	RF02166	1.7E-16	*Colombo* 2015, *Zhu* 2017
TERC	AF221928.1	-	6.3	-	4.0E-02	Telomerase-vert	RF00024	5.7E-58	*Engreitz* 2016
TUG1	XR_483407.1	-	2.1	-	8.0E-02	TUG1_3	RF01891	1.6E-46	*Li* 2016
						TUG1_1	RF01882	1.3E-21	
						TUG1_4	RF01892	2.1E-20	
						TUG1_2	RF01883	3.2E-17	

Table 2: Summary of the enriched pathways based on the list of up-regulated (lysosome) and down-regulated (lower tables) transcripts in Fed-batch vs Batch comparison at Day 4 (left panel) and Day 7 (right panel). The columns show from left to right the ID from KEGG, the pathway name, the FDR associated with the enrichment (FDR ≤ 0.10 threshold), the number of differentially expressed genes in the pathway, the number of total genes listed in the pathway, the differentially expressed genes over total genes ratio.

UP fed D4 v batch D4

KEGG ID	Pathway	FDR	Diff. Expressed (DE) genes	Total Genes (TG)	DE:TG ratio
04142	**Lysosome**	6.3E-02	14	124	0.11

UP fed D7 v batch D7

KEGG ID	Pathway	FDR	Diff. Expressed (DE) genes	Total Genes (TG)	DE:TG ratio
04142	**Lysosome**	7.1E-04	21	124	0.17

DOWN fed D4 v batch D4

KEGG ID	Pathway	FDR	Diff. Expressed (DE) genes	Total Genes (TG)	DE:TG ratio
05034	**Alcoholism**	8.5E-11	48	203	0.24
04110	**Cell cycle**	2.3E-09	34	125	0.27
05322	**Systemic lupus erythematosus**	6.8E-07	33	147	0.22
03030	**DNA replication**	3.5E-04	12	35	0.34
03460	**Fanconi anemia pathway**	8.5E-04	14	51	0.27
05203	**Viral carcinogenesis**	1.1E-03	37	243	0.15
03410	**Base excision repair**	1.3E-03	11	35	0.31
03430	**Mismatch repair**	4.4E-03	8	22	0.36
01200	**Carbon metabolism**	4.5E-03	21	117	0.18

00480	**Glutathione metabolism**	5.7E-03	13	56	0.23
00900	**Terpenoid backbone biosynthesis**	6.4E-03	8	24	0.33
04114	**Oocyte meiosis**	8.5E-03	20	116	0.17
05200	**Pathways in cancer**	1.3E-02	48	397	0.12
03440	**Homologous recombination**	1.6E-02	8	28	0.29
04914	**Progesterone-mediated oocyte maturation**	1.8E-02	16	90	0.18
01212	**Fatty acid metabolism**	2.1E-02	11	51	0.22
04540	**Gap junction**	2.8E-02	15	86	0.17
00280	**Valine, leucine and isoleucine degradation**	4.2E-02	11	56	0.20
04115	**p53 signaling pathway**	6.4E-02	12	68	0.18
00240	**Pyrimidine metabolism**	6.4E-02	16	104	0.15
03420	**Nucleotide excision repair**	6.5E-02	9	44	0.20
00630	**Glyoxylate and dicarboxylate metabolism**	7.6E-02	7	30	0.23

DOWN fed D7 v batch D7

KEGG ID	Pathway	FDR	Diff. Expressed (DE) genes	Total Genes (TG)	DE:TG ratio
04110	**Cell cycle**	7.9E-14	38	125	0.30
03030	**DNA replication**	1.2E-09	17	35	0.49
03013	**RNA transport**	1.2E-09	38	170	0.22
03040	**Spliceosome**	1.3E-07	30	134	0.22
05203	**Viral carcinogenesis**	2.8E-06	40	243	0.16
03430	**Mismatch repair**	2.2E-05	10	22	0.45
03460	**Fanconi anemia pathway**	2.3E-05	15	51	0.29
03410	**Base excision repair**	3.3E-04	11	35	0.31

05322	**Systemic lupus erythematosus**	9.3E-04	24	147	0.16
03440	**Homologous recombination**	1.4E-03	9	28	0.32
03420	**Nucleotide excision repair**	2.5E-03	11	44	0.25
05034	**Alcoholism**	3.5E-03	28	203	0.14
04114	**Oocyte meiosis**	4.2E-03	19	116	0.16
03008	**Ribosome biogenesis in eukaryotes**	2.0E-02	14	83	0.17
00240	**Pyrimidine metabolism**	2.3E-02	16	104	0.15
03015	**mRNA surveillance pathway**	2.6E-02	15	96	0.16
00310	**Lysine degradation**	3.1E-02	10	52	0.19
04914	**Progesterone-mediated oocyte maturation**	3.4E-02	14	90	0.16
03018	**RNA degradation**	4.4E-02	13	83	0.16
05166	**HTLV-I infection**	4.8E-02	32	294	0.11
05222	**Small cell lung cancer**	4.9E-02	13	85	0.15
00020	**Citrate cycle (TCA cycle)**	5.5E-02	7	32	0.22
05215	**Prostate cancer**	6.7E-02	13	89	0.15
01210	**2-Oxocarboxylic acid metabolism**	1.0E-01	5	20	0.25

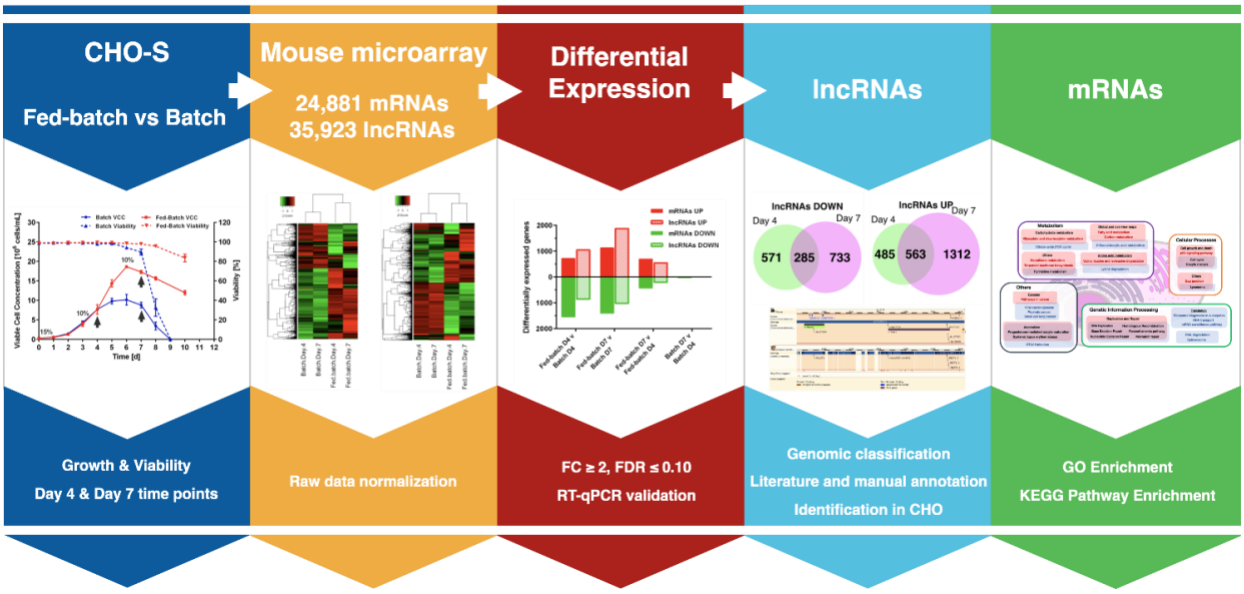

Figure 1: Summary of the experimental workflow. Growth and culture viability of a CHO-S cell line in Batch and Fed-batch cultures were measured for 10 days and samples for RNA extraction taken at Day 4 and at Day 7. The samples were analysed on a mouse array containing all the coding and non-coding transcripts stored in the main public databases. The measured intensities were log-normalized and differentially expressed transcripts/genes were filtered for a fold change (FC) ≥ 2 and an FDR ≤ 0.10. A selected group of genes was validated through RT-qPCR (Supplementary Material). Due to the poor annotation of lncRNAs in CHO, the identification of potential targets with a described biological role required the comparison of human and mouse literature and databases, followed by alignment against the Chinese hamster genome, leading to predicted lncRNAs transcripts and previously un-annotated genomic regions (Table 1). At the same time, GO and pathway enrichment was implemented on mRNAs.

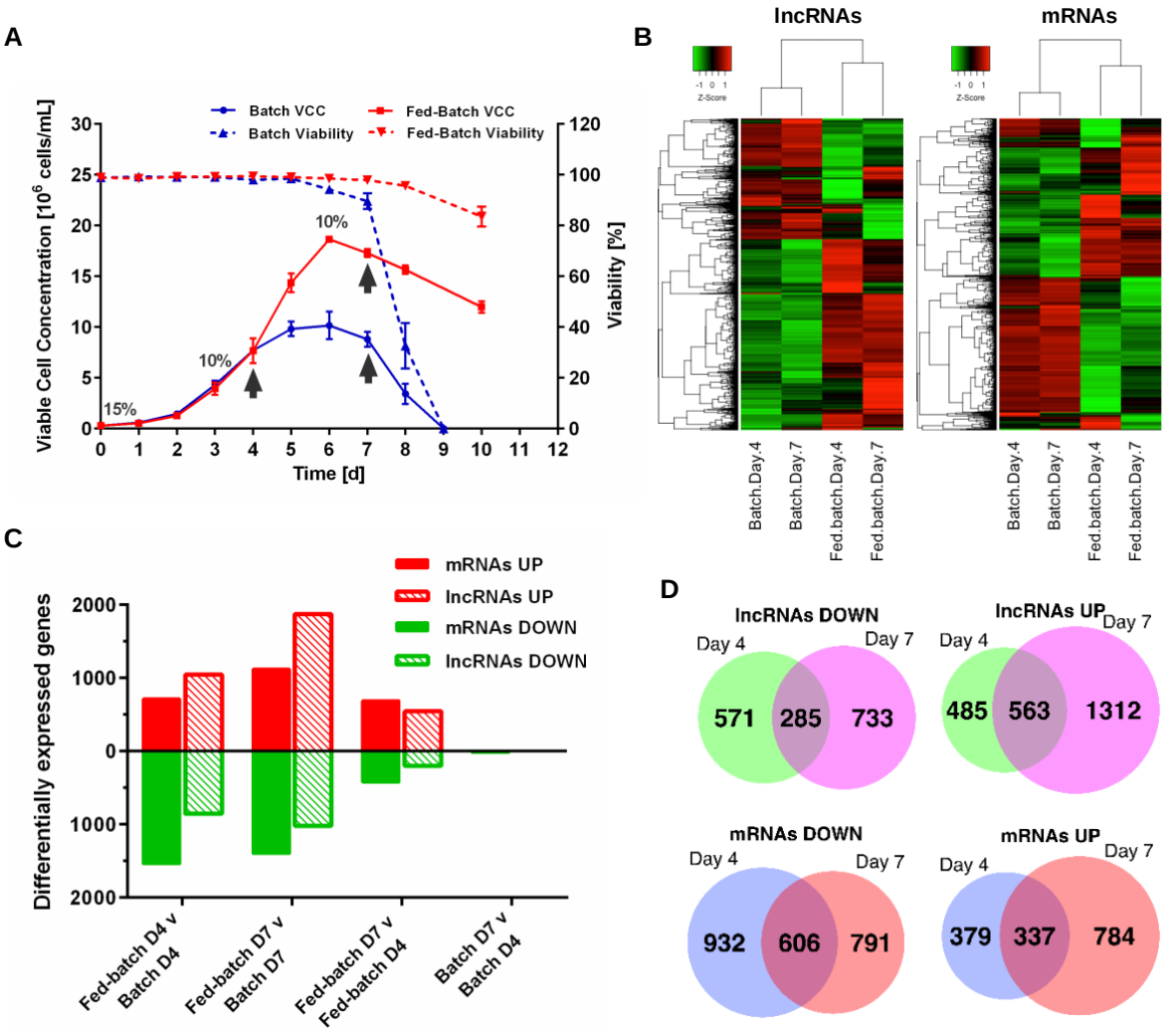

Figure 2. A) Growth profiles for the model CHO-S host cell line throughout Batch and Fed-batch culture. Viable cell concentration (VCC) and culture viability are shown (red for Batch and blue for Fed-batch). The percentage of CHO CD Efficient Feed B Liquid Nutrient Supplement added to the existing working volume are shown for each feeding day (days 0, 3 and 6). The arrows indicate when samples for microarray analysis were harvested (Day 4 and Day 7). B) Hierarchical clustering heatmaps arranging samples into groups based on their averaged log-normalized expression levels. Only transcripts/genes with an expression variance between each group above the 80th percentile are shown. The dendrogram shows the relationships for lncRNAs (left panel) and mRNAs (right panel). C) Down-regulated (green bars) and up-regulated (red bars) mRNAs (full bars) and lncRNAs (textured bars) in CHO-S cells during batch and fed-batch culture for each of the compared pairs showing a fold change ≥ 2 and an FDR ≤ 0.10. D) Venn diagrams showing the number of lncRNAs and mRNAs differentially expressed (DE) in Fed-batch vs Batch (FC ≥ 2, FDR ≤ 0.10). Genes DE at both Day 4 and Day 7 are represented in the overlaps while genes DE only at one time point are represented inside their corresponding circle.

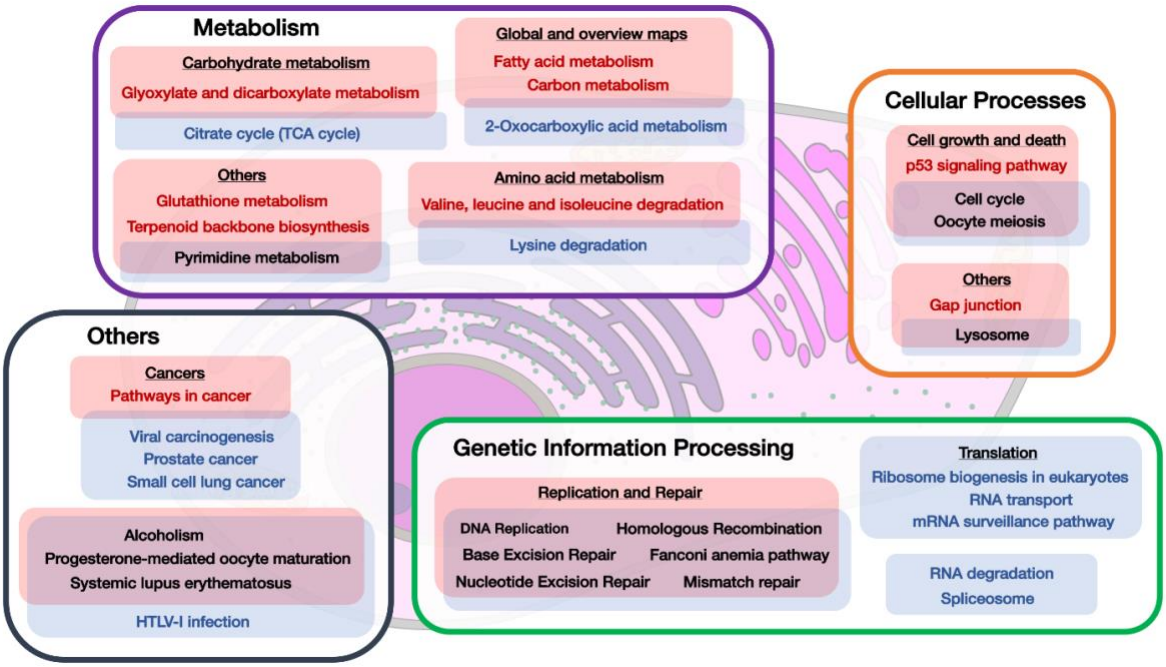

Figure 3: Graphical representation of the enriched KEGG pathways listed in Table 2, hierarchically grouped based on KEGG Pathway Maps. Each filled rectangle (in red for Day 4, in blue for Day 7) contains the corresponding enriched pathways, with pathways enriched at both Day 4 and Day 7 enclosed in the overlaps between the filled rectangles. The enrichment is based on genes differentially expressed in Fed-batch vs Batch at Day 4 and Day 7 with an FDR ≤ 0.10.