Multi-omics profiling of CHO parental hosts reveals cell line-specific variations in bioprocessing traits — Source link ↗

Meiyappan Lakshmanan, Yee Jiun Kok, Alison P. Lee, Sarantos Kyriakopoulos ...+12 more authors

Institutions: Agency for Science, Technology and Research, Sungkyunkwan University

Related papers:

- Multi-omics profiling of CHO parental hosts reveals cell line-specific variations in bioprocessing traits

- A Consensus Genome-scale Reconstruction of Chinese Hamster Ovary Cell Metabolism.

- Biopharmaceutical benchmarks 2018.

- The emerging CHO systems biology era: harnessing the 'omics revolution for biotechnology.

- Application of a curated genome-scale metabolic model of CHO DG44 to an industrial fed-batch process

Multi-omics profiling of CHO parental hosts reveals cell line-specific variations in bioprocessing traits

Meiyappan Lakshmanan[1], Yee Jiun Kok[1], Alison P. Lee[1], Sarantos Kyriakopoulos[1], Hsueh Lee Lim[1], Gavin Teo[1], Swan Li Poh[1], Wen Qin Tang[1], Jongkwang Hong[1], Andy Hee-Meng Tan[1],

Xuezhi Bi[1], Ying Swan Ho[1], Peiqing Zhang[1], Say Kong Ng[1], Dong-Yup Lee[1,2]

[1]Bioprocessing Technology Institute, Agency for Science, Technology and Research (A*STAR), 20 Biopolis Way, #06-01, Centros 138668, Singapore

[2]School of Chemical Engineering, Sungkyunkwan University, 2066, Seobu-ro, Jangan-gu, Suwon-si, Gyeonggi-do 16419, Republic of Korea; telephone: +82-31-290-7253; fax: +82-31-290-7272; e-mail: dongyuplee@skku.edu

Correspondence to: D.-Y. Lee

Running Head: Multi-omics characterization of CHO parental cell lines

26 **Abstract**

27 Chinese hamster ovary (CHO) cells are the most prevalent mammalian cell factories for producing

28 recombinant therapeutic proteins due to their ability to synthesize human-like post-translational

29 modifications and ease of maintenance in suspension cultures. Currently, a wide variety of CHO

30 host cell lines have been developed; substantial differences exist in their phenotypes even when

31 transfected with the same target vector. However, relatively less is known about the influence of

32 their inherited genetic heterogeneity on phenotypic traits and production potential from the

33 bioprocessing point of view. Herein, we present a global transcriptome and proteome profiling of

34 three commonly used parental cell lines (CHO-K1, CHO-DXB11 and CHO-DG44) in suspension

35 cultures and further report their growth-related characteristics, and N- and O-glycosylation

36 patterns of host cell proteins (HCPs). The comparative multi-omics analysis indicated that some

37 physiological variations of CHO cells grown in the same media are possibly originated from the

38 genetic deficits, particularly in the cell cycle progression. Moreover, the dihydrofolate reductase

39 deficient DG44 and DXB11 possess relatively less active metabolism when compared to K1 cells.

40 The protein processing abilities and the N- and O-glycosylation profiles also differ significantly

41 across the host cell lines, suggesting the need to select host cells in a rational manner for the cell

42 line development on the basis of recombinant protein being produced.

43

44 **Keywords:** multi-omics analysis, mammalian systems biotechnology, critical quality attributes,

45 N-glycosylation, CHO parental cell lines,

1 INTRODUCTION

Mammalian expression systems are the preferred choice for the industrial production of recombinant glycoprotein therapeutics due to their humanized post-translational modifications, i.e. N-glycosylation and appropriate protein folding, and the ability to secrete the product naturally outside of the cell (Wurm, 2004). Among several mammalian host cells, Chinese hamster ovary (CHO) cells are commonly used since they can grow in suspension culture with chemically defined media, and are highly resistant to viral susceptibility. Overall, 84% of the biopharmaceutical drugs are currently produced using CHO cells (Walsh, 2018).

Generally, bioprocess development includes two major stages: 1) establishment of efficient cell lines producing target proteins, and 2) development of optimal cell culture media and/or process conditions (Hong et al., 2018). Currently, a multitude of CHO cells with various genetic backgrounds have been utilized since the first approval of a drug produced in CHO cells in 1984 (Kaufman et al., 1985). Of them, the CHO-K1, -DXB11 and -DG44 are the most relevant parental cell lines to manufacture target products industrially (Golabgir et al., 2016). Historically, the proline-deficient K1 cells was established in 1968 (Kao and Puck, 1968), about 10 years after its first isolation from the Chinese hamster ovaries. Later, the DXB11 was generated by mutagenesis targeting the dihydrofolate reductase gene (*Dhfr*) from K1 cells, resulting in missense mutation in one allele and complete deletion in the other (Urlaub and Chasin, 1980; **Figure 1a**). Note that the *Dhfr* gene is responsible for reducing folate, which is an essential step in nucleotide biosynthesis. Even though both alleles of the *Dhfr* gene were modified in the -DXB11 cells, it still presented some minor activity of the gene. Thus, the deletion of both alleles of this gene led to the generation of the -DG44 cell line in the mid-1980s (Urlaub et al., 1983). Here, it should be noticed that DG44 was not derived from the K1 cell line, but from the very first isolated CHO cell lines (**Figure 1a**).

3

69   Interestingly, although these cell lines are derived from a common ancestor, extensive mutagenesis

70   and the clonal diversification has resulted in considerable genetic heterogeneity as revealed by

71   their genome sequences (Kaas et al., 2015; Lewis et al., 2013; Xu et al., 2011). Significant

72   differences in the gene copy numbers, single nucleotide polymorphisms, structural variants and

73   chromosomal rearrangements were observed.

74   Due to the presence of such "quasispecies" within the CHO parental cell lines (Wurm,

75   2013), there exist considerable dissimilarities in their phenotypes and possibly even variations in

76   the critical quality attributes, e.g. N- and O-glycosylation profiles, of the recombinant proteins

77   produced from them. For example, when the same antibody was produced from different CHO

78   cells, the viable cell densities and final titer across several bioprocess conditions varied

79   significantly, depending on the parental cell line used (Hu et al., 2013; Reinhart et al., 2018).

80   Furthermore, there are appreciable variations among the N-glycosylation patterns and other quality

81   attributes of the recombinant proteins produced (Könitzer et al., 2015; Reinhart et al., 2018).

82   Therefore, it is highly required to understand the diversity in the phenotypes of CHO cells and

83   associate it with known genetic heterogeneity for the rational selection of suitable parental cells.

84   Moreover, the characterization of N- and O-glycosylation preferences of various CHO host cells

85   is a critical aspect within the "Quality by Design (QbD)" framework (Butler and Spearman, 2014).

86   Towards the goal of unraveling the heterogeneity of the various CHO parental cell lines

87   and linking it with their genetic diversity, here we have accomplished an integrative multi-omics

88   analysis, at the transcriptomic, proteomic and glycomic levels, of the K1, DG44 and DXB11 cells

89   during the exponential growth phase in suspension cultures. To our knowledge, although a recent

90   study reported proteomics analysis to compare the differential protein expression in CHO-K1, -S

91   and *Dhfr⁻* cell lines (Xu et al., 2017), the current work provides more comprehensive multi-omics

92 analysis for various CHO host cells, thus serving as a basis to further rationalize cell line

93 development and engineering.

94

2 MATERIALS AND METHODS

96 **2.1 Cell lines, cultivation and characterization**

97 CHO-K1 (ATCC No. CCL-61, American Type Culture Collection, USA), CHO-DXB11 (ATCC

98 No. CRL-9096, American Type Culture Collection) and CHO-DG44 (Gibco™ catalog number

99 12609-012, Invitrogen, USA) previously adapted to serum free suspension culture were

100 propagated in a DMEM/F12-based protein free chemically defined medium (PFCDM)

101 supplemented with Soybean peptone (Catalog number P0521, Sigma-Aldrich, USA) and HT

102 supplement (Gibco™ catalog number 11067-030, ThermoFisher Scientific, USA). To compare cell

103 growth over 6 days, all three cell lines were cultivated in batch mode by seeding 2×10^5 cells/mL

104 into orbitally agitated disposable Erlenmeyer flasks (Corning, USA) in duplicates. Replicate shake

105 flasks were cultivated similarly and harvested at Day 2 for the various –omics analyses. Cell

106 density and viability were measured daily by the trypan blue dye exclusion method using Vi-Cell

107 XR (Beckman Coulter, USA). Glucose, lactate, ammonium, glutamine and glutamate

108 concentrations in the culture supernatant were measured daily using BioProfile 100 Plus (Nova

109 Biomedical, USA).

110

111 **2.2 Analysis of extracellular nutrients/metabolites**

112 Glucose, lactate, glutamine and glutamate concentrations in the culture supernatant were measured

113 using a 2700 Biochemistry Analyser (Yellow Springs Instruments, USA). Ammonia concentration

114 was measured using a commercial kit (Sigma, USA).

2.3 Transcriptome profiling and data processing

Cell pellets (1×10^7 cells) were collected from three replicate flasks for each of the cell lines. Total RNA was isolated using RNeasy Mini Kit (Qiagen, Germany). RNA quality and integrity were verified using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific) and Agilent Bioanalyzer 2100 (Agilent Technologies, USA). cDNA targets were generated from 250 ng total RNA using the GeneChip 3' IVT PLUS Reagent Kit (Thermo Fisher Scientific). The targets were then hybridized onto a proprietary CHO 3' IVT microarray (CHO V3) consisting of 43,856 probe sets that correspond to 13,358 mouse genes based on CHO EST Sequence and Annotation Release 06 (Affymetrix, Singapore). Quality assessment and data analysis were carried out on the CEL files obtained, using Partek Genomics Suite version 6.6 (Partek Inc., USA).

2.4 Proteome profiling and data processing

1×10^7 cells were collected for each experimental condition in set of biological triplicates. Cells were washed twice in $1 \times$ PBS and lysed in 8 M urea supplemented with $1 \times$ Halt Protease Inhibitor Cocktail (Thermo Fisher Scientific, USA). Insoluble cellular material was removed by centrifugation at 18,000 ×g for 20 min at 4°C, and protein concentration in each sample was determined by Bradford assay using Pierce™ Coomassie Plus Assay Reagent (Thermo Fisher Scientific). Subsequently, 100 μg of protein from each sample was digested with trypsin and labelled with TMTsixplex™ Isobaric Mass Tagging Kit (Thermo Fisher Scientific) according to manufacturer's instructions. Labelled peptides from each set of replicates were then combined and cleaned up using SCX cartridge system (Applied Biosystems, USA) prior to 2D-LC-MS analysis.

First dimension LC separation consisted of high pH reverse-phase fractionation at a flow rate of 1 ml/min from a gradient of 3.5-54% acetonitrile in 20 mM ammonium formate, pH 9.6

6

138   over 30 min on an XBridge BEH C18 column, 130Å, 3.5 μm, 3 mm × 150 mm (Waters, USA) in

139   an Ultimate 3000 UHPLC system (Thermo Fisher Scientific), with a total of 10 pooled peptide

140   fractions from each triplicate set collected. Second dimension separation was carried out by

141   analyzing 3 μg of each sample by nanoLC-MS/MS at a flow rate 300 nl/min 2-50% acetonitrile in

142   0.1 % formic acid over 100 min on an ACQUITY UPLC Peptide BEH C18 column, 130Å, 1.7 μm,

143   75 μm × 200 mm (Waters) in a nanoACQUITY UPLC system (Waters, USA) coupled to an LTQ-

144   Orbitrap Velos mass spectrometer (Thermo Fisher Scientific). The LTQ-Pribtirap Velos was

145   operated in data-dependent mode with full mass scan acquired for mass range of 400-1,400 m/z.

146   The top 10 most intense peaks with charged state ≥ +2 were fragmented by HCD with normalized

147   collision energy of 45%, minimum signal threshold of 500, isolation width of 2 Da and activation

148   time of 0.1 ms.

149   Raw MS data was analyzed by Sequest HT and MS Amanda search engines against NCBI

150   CHO refseq database using Proteome Discoverer v. 2.0 (Thermo Fisher Scientific). Precursor mass

151   range of 350-5,000 Da was used, with MS1 tolerance of 10 ppm and MS2 tolerance of 0.02 Da,

152   static modifications of carbamidomethylation on cysteine and TMTsixplex label on lysine,

153   dynamic modifications of oxidation on methionine, TMTsixplex on peptide N-terminus and

154   acetylation on protein N-terminus, and PSM validation by percolator node. Quantification was

155   performed without replacement of missing value and without application of isotope correction

156   factors. Co-isolation threshold was set at 50% with peptide abundance for each channel normalized

157   by total peptide amount and scaled on channels average.

158

159   **2.5 Metabolome profiling and data processing**

160   1×10^7 cells were collected from each experimental condition and quenched in 5 volumes of ice-

161 cold 150 mM sodium chloride (Sigma-Aldrich) solution. Subsequently, the quenched cell pellet

162 was processed using a two-phase liquid extraction protocol involving methanol, tricine solution

163 and chloroform as previously described (Yusufi et al., 2017). Polar metabolite extracts were stored

164 at -80°C. Prior to analysis, the extracts were dried and reconstituted in cold methanol and water

165 (5:95 v/v).

166   Untargeted LC-MS analysis of the polar metabolites was then performed using an ultra-

167 high performance liquid chromatography (UPLC) system (Acquity; Waters, USA) coupled to a

168 mass spectrometer (QExactive MS; Thermo Scientific, USA). A reversed phase (C18) UPLC

169 column with polar end-capping (Acquity UPLC HSS T3 column, 2.1mm, 100 mm, 1.8 mm; Waters)

170 was used with two solvents: 'A' being water with 0.1% formic acid (Fluka, Sigma-Aldrich, USA),

171 and 'B' being methanol (Optima grade, Thermo Fisher Scientific, USA) with 0.1% formic acid.

172 The LC program was as follows: the column was first equilibrated for 0.5 min at 0.1% B. The

173 gradient was then increased from 0.1% B to 50% B over 8 min before being held at 98% B for 3

174 min. The column was washed for a further 3 min with 98% acetonitrile (Optima grade, Fisher

175 Scientific) with 0.1% formic acid and finally equilibrated with 0.1% B for 1.5 min. The solvent

176 flow rate was set at 400 mL/min; a column temperature of 30°C was used. The eluent from the

177 UPLC system was directed into the MS. High-resolution mass spectrometry was then performed

178 in both positive and negative electrospray ionization (ESI) modes, with a mass range of 70 to 1,050

179 m/z and a resolution of 70,000. Sheath and auxiliary gas flow was set at 30.0 and 20.0 (arbitrary

180 units) respectively, with a capillary temperature of 400°C. The spray voltages were 1.25 kV for

181 positive and 1.5 kV negative mode ionization. Mass calibration was performed using standard

182 calibration solution (Thermo Scientific) prior to injection of the samples. A quality control (QC)

183 sample comprising of equal aliquots of each sample was run at regular intervals during the batch

8

184   LC-MS runs.

185   The raw LC-MS data obtained was pre-processed based on the XCMS peak finding

186   algorithm (Smith et al., 2006) and the QC samples were used to adjust for instrumental drift. Total

187   area normalization was then applied to the pre-processed data prior to statistical analysis using

188   multivariate (SIMCA-P+ software, version 13.0.3, Umetrics, Sweden) and univariate tools,

189   including relative ratios, student's t-test (Welch's correction) and hierarchical clustering for the

190   classification of common trends. Metabolite identities were confirmed by MS-MS spectral

191   comparison with commercially available metabolite standards or with available mass spectral

192   libraries (Wishart et al., 2018).

193

194   **2.6 N- and O-glycome profiling**

195   Glycomics analysis of CHO cells was performed according to a previous study by North et al.,

196   (2010) with slight modifications. Briefly, N-glycans were released by PNGase F (Prozyme, USA)

197   from post nuclear fraction of CHO cell lysate which had been subjected to reduction, S-

198   carboxylmethylation, and tryptic digestion. O-glycans were released by sodium borohydride-

199   mediated beta-elimination, from de-N-glycosylated peptides. Both N- and O-glycan samples were

200   permethylated and purified by C18 solid phase extraction. N-glycan samples were analyzed by

201   5800 TOF/TOF mass spectrometer (AB SCIEX, USA) and O-glycans were analyzed by 5500 Q-

202   TOF system online coupled downstream to a nano-LC system with an integrated chip-based nano

203   C18 column (both from Agilent Technologies, USA). Deconvoluted mass spectra were analyzed

204   for data obtained from both MALDI and QTOF systems. Neutral masses were extracted and parsed

205   against theoretical N- and O-glycan mass libraries (permethylated for N-glycans, reduced and

206   permethylated for O-glycans) for CHO cells. Putative N- and O-glycan structures were assigned

9

207 based on the knowledge of glycan biosynthetic pathway in CHO cells (North et al., 2010).

208

2.7 Identification of differentially expressed genes in the transcripts and proteins

210 Using the normalized gene expression data, the differentially expressed genes was identified with

211 CHO-K1 as the reference using empirical Bayes-moderated t-statistics via "limma" package in R

212 (Wettenhall and Smyth, 2004). Genes showing $\geq \pm 1.5$ fold difference with a FDR-adjusted p-

213 values ≤ 0.01 were classified as differentially expressed.

214   Quantifiable proteins were identified at 1% FDR with ≥ 2 unique peptides and had

215 abundance values reported for at least two of the triplicate samples; any missing protein abundance

216 value was replaced by the average abundance values reported for the other two of the triplicate

217 samples where applicable. Proteins showing $\geq \pm 1.5$ fold difference with a FDR-adjusted p-values

218 ≤ 0.1 were classified as differentially expressed.

219

2.8 Metabolic pathway enrichment analysis of differentially expressed transcripts and proteins

222 Enrichment analysis of the metabolic genes was performed using omics data in conjunction with

223 the CHO genome-scale metabolic network (Hefzi et al., 2016). The gene-protein-reaction (GPR)

224 associations in the metabolic network and the reaction pathway classifications were used to assign

225 the metabolic genes to different reaction pathways. To identify enriched metabolic pathways, the

226 differentially expressed genes are first identified based on the p-value (<0.01; FDR-corrected) and

227 fold change (>1.5). The enrichment of differentially regulated genes in each pathway was then

228 computed using the hypergeometric distribution as follows:

10

229
$$p_x = \frac{\binom{K \quad N-K}{k \quad n-k}}{\binom{N}{n}}$$

230   where N is the total number of genes in CHO cells, K is the total number of up/downregulated

231   genes, n is the total number of genes in pathway "x" and k is the number of up/downregulated

232   genes in pathway "x". The pathway with lowest p-value is classified as highly enriched in

233   differentially expressed genes.

234

235   **2.9 Estimation of gene copy number variations (CNV) between CHO cell lines**

236   The changes in absolute gene copy numbers between the CHO cell lines were calculated using

237   CHO-K1 as the reference from the normalized sequencing read depth of individual cell lines which

238   was reported previously (Kaas et al., 2015). A gene is considered to be deleted in a cell line if the

239   read depth is zero in that particular cell line but >0.95 in CHO-K1. Gene amplifications and

240   reductions were also calculated in a similar manner.

241

242   **3 RESULTS**

243   **3.1 CHO-K1 grows faster in suspension batch cultures than DG44 and DXB11**

244   The three CHO parental cell lines, i.e. K1, DG44 and DXB11, were propagated in duplicates in

245   the same DMEM/F12-based protein free chemically defined medium (PFCDM) supplemented

246   with Soybean peptone and HT supplement, in order to reduce variability due to differences in cell

247   culture media. These cell culture experiments clearly showed that the K1 cells grew much faster

248   when compared to the other two cell lines, reaching the maximum cell densities first (**Figure 1b**).

249   The doubling time of K1 cells is 20±1 hr whereas DG44 and DXB11 correspond to 30±0.3 and

11

250   32±4, respectively. Notably, these results were highly consistent with an earlier study which

251   reported that the maximum viable cell densities achieved in batch culture following the order of

252   K1>>DG44>DXB11 (Sommeregger et al., 2013). While the cells were grown in commercial

253   media in the earlier study, the concordance between the two results highlight that the slow growth

254   of *Dhfr⁻* cells is mainly because of the genetic divergence. Here, it should be highlighted that

255   although all the three media contained HT, the purine supplements enabling the cells to overcome

256   deprivation of nucleoside precursors as a result of the *Dhfr* deficiency, we still observed much

257   lower growth rates than K1. This confirms that the loss of *Dhfr* gene activity could not be simply

258   compensated by supplementing HT as there could be additional pathways contributing to such

259   growth differences (Florin et al., 2011). In order to further investigate metabolic traits of CHO host

260   cells, we also profiled the major nutrients and byproducts in the cell culture. Glucose was rapidly

261   consumed in the K1 compared to others (**Figure 1c**). Similarly, a sharper decrease in the glutamine

262   concentrations was also observed in K1 cell cultures with a concomitant higher accumulation of

263   ammonia and lactate than DXB11 and DG44 cell cultures (**Figure 1d**), possibly indicating that K1

264   cells undergo higher carbon and nitrogen catabolism.

265

266   **3.2 Global transcriptome and proteome profiling of CHO parental cell lines confirms Known**

267   **genetic heterogeneity**

268   To systematically analyze the underlying molecular mechanisms leading to the observed growth

269   phenotypes across CHO host cells, samples were harvested during the exponential phase for

270   transcriptome and proteome profiling. Transcriptomics analysis was performed using a proprietary

271   microarray encompassing CHO probe sets that correspond to 13,358 unique mouse genes

272   (**Supplementary file**). As expected, the expression of *Dhfr* gene in K1 was more than twice the

273   intensity of DXB11 while a very low detectable signal was present in DG44, possibly due to noise,

274   thus confirming the known genetic backgrounds of CHO cells. MS-based proteomic profiling

275   using multiplexed isobaric tags (iTRAQ) identified and detected a total of 1998 unique proteins

276   corresponding to 1983 genes in more than one biological replicate (**Supplementary file**). Note

277   that 1857 of these 1983 genes were also present in the microarray, thus providing relevant

278   information at both the transcript and protein levels (**Figure 2a**). After transcriptome and proteome

279   profiling, we initially analyzed the global mRNA and protein expression patterns via two different

280   methods: 1) principal component analysis (PCA) and 2) hierarchical clustering of Pearson

281   correlation coefficients. PCA revealed that the expression patterns of all the cell lines were

282   markedly different at both transcript and protein levels (**Figure 2b and 2c**), highlighting that the

283   genetic divergence of the CHO parental cell lines indeed influences the phenotype via

284   transcriptional and translational regulations. The hierarchical clustering of transcriptome and

285   proteome also showed similar trends, where the global expression patterns of K1 are closer to

286   DXB11 than DG44 cells at both transcript and protein levels (**Figure 2d and 2e**), recapitulating

287   the known genetic divergence of the CHO parental cell lines (**Figure 1a**).

288

289   **3.3 Integrative transcriptome and proteome analysis highlights that cellular metabolism and**

290   **cell cycle progression are negatively regulated in *Dhfr*-deficient cells**

291   To pinpoint relevant cellular functions attaining the observed phenotypic differences based on the

292   transcriptomic and proteomic features, we first identified the differentially expressed genes at both

293   levels using K1 as a reference. In the K1 vs DXB11 comparisons, we identified 2231 mRNAs and

294   239 proteins to be differentially expressed, with a large fraction being down-regulated in DXB11

295   (**Supplementary file**). A total of 2875 mRNAs and 304 proteins were also differentially expressed

13

296   in K1 vs DG44 comparisons where some are down-regulated in DG44. Together, the larger fraction

297   of up-regulated compared to down-regulated genes in K1 suggests more active transcription and

298   translation compared to DG44 and DXB11. Subsequent gene ontology-based enrichment analysis

299   of the differentially expressed transcripts and proteins revealed that the cellular metabolic process

300   including lipid and energy metabolism, DNA, RNA and protein metabolic processes, cell cycle,

301   growth, and cell proliferation and death present the major expression differences among the three

302   cell lines (**Figure 3a**). Particularly, both DG44 and DXB11 presented a negative regulation of cell

303   cycle progression, DNA replication and the biosynthesis of nucleotides. Several genes which

304   commit the cells in cell cycle from G2 phase to M phase and the next mitotic phase were all down-

305   regulated in DG44 and DXB11. Further, the down-regulation of cyclins including *Ccny*, *Ccnl2*,

306   *Ccnc* and *Ccnh*, and serine/Threonine Kinase, *Akt1*, indicates that the cell cycle progression and

307   DNA replication are negatively regulated by PI3K-Akt and MAPK signaling pathways (Duronio

308   and Xiong, 2013; Rhind and Russell, 2012). In addition to the cell cycle and cellular organization

309   related genes, we also found that the purine nucleotide metabolic processes (as expected, due to

310   the *Dhfr* gene copy number differences) and sterol metabolic process to be significantly enriched

311   in the differentially expressed genes. Remarkably, the expression patterns of glycosylated protein

312   processing processes such as N-glycosylation, Golgi organization, vesicle-mediated protein

313   transport and protein folding were also significantly altered across cell lines.

314   We next overlaid the transcriptomic and proteomic expression patterns to compare their

315   concordance in each comparison (**Figure 3b**). A set of 55 proteins commonly downregulated in

316   DG44 and DXB11 at both levels were identified, including *Pacsin2* and *Ezr*, both of which are

317   substrates of protein kinases associated with the regulation of actin cytoskeleton dynamics

318   (D'Angelo et al., 2007; Goh et al., 2012; de Kreuk et al., 2011), altering the cellular structure as

14

319   well as cell migration and macromolecular assembly. The current analysis also identified 110 and

320   113 up- and down-regulated proteins in DXB11 vs K1 and DG44 vs K1 comparisons, respectively,

321   which do not show any expression differences in their corresponding mRNAs (Supplementary file).

322   Their gene enrichment analysis allowed us to understand that central metabolic pathways,

323   particularly energy metabolism including the ATP biosynthesis (e.g., *Atp5j*, *Atp5o* and *Atpif1*) via

324   cellular respiration and ROS detoxifying glutathione metabolism, are post-transcriptionally

325   regulated. In addition, some of the genes from nucleotide metabolism and glycosyl compound

326   biosynthetic pathways were also potentially post-transcriptionally regulated. Such examples

327   include *Ada*, *Neu2*, *Galnt2* and *Rrm2*.

328   The fold-changes of 1857 proteins that are detected at both protein and transcript levels

329   showed a moderate correlation with corresponding transcript fold-changes in both DXB11 vs K1

330   (Spearman's $r = 0.6017$; Pearson's $r = 0.5949$; **Figure 3c**) and DG44 vs K1 (Spearman's $r = 0.7524$;

331   Pearson's $r = 0.7425$; **Figure 3d**). Previously, several studies have reported similar correlations

332   (moderate or poor) which are attributable to possibly the changes in translation and transcription

333   rates, protein and transcript degradation rates and other regulatory factors (Liu et al., 2016).

334   However, despite the moderate correlation, we were able to identify a list of 109 genes from such

335   comparisons as potential targets for cellular engineering in CHO cells since they showed a very

336   good correlation in expression between transcript and protein levels ($R^2 > 0.9$; **Supplementary**

337   **file**).

338

339   **3.4 Hard-wired changes of gene expression in signaling and nucleotide metabolic pathways**

340   **are observed in DXB11 and DG44 cell lines**

341   In order to understand whether significant differences in the transcriptional and translational

342   patterns across the cell lines are caused by the genetic heterogeneity, we examined their gene copy

343   number changes. Interestingly, this analysis showed that 9% of the total 2231 differentially

344   expressed mRNAs have a hard wiring of copy number changes in the DXB11 genome when

345   compared to K1 (**Supplementary file**; **Figure 3e**). 51 up-regulated genes have increased copy

346   numbers including two serine/threonine-protein kinases (*Wnk2* and *Mknk2*) that are known to

347   inhibit cell proliferation by negatively regulating the ERK pathway (Moniz et al., 2007); 154

348   down-regulated genes showed a decrease in gene copy numbers, including homozygous deletion

349   of two genes, *Tmem14a* and *Plgrkt*. Notably, *Tmem14a* is a transmembrane protein which inhibits

350   apoptosis by negatively regulating the mitochondrial membrane permeabilization (Woo et al.,

351   2011). Therefore, its complete deletion in DXB11 could possibly result in dysregulation of

352   apoptosis inhibition pathway, contributing to the cell death. In addition, certain cell cycle

353   associated genes such as *Cdk11b*, *Taf4* and *Dscc1* were also heterozygously deleted, presumably

354   leading to the negative regulation of cell cycle progression as mentioned earlier.

355   When compared to DXB11 cells, even larger fraction of differentially expressed mRNAs

356   (18%) are associated with the copy number changes in the DG44 genome (**Figure 3f**). Interestingly,

357   DG44 cells showed homozygous deletion of 12 genes including the known removal of *Dhfr*. In

358   addition, passenger deletion of neighboring genes (*Fam15b*, *Msh3* and *Zfyve16*) and other cell

359   cycle related genes (e.g., *Cdk11b*, *Taf4* and *Dscc1*) were observed. Clearly, these results suggest

360   that DG44 cell lines have the inherent limitation in cell cycle progression and nuclear division as

361   reflected in the slow growth phenotype. Several up-regulated genes in DG44 also have

362   corresponding augmentation in gene copy numbers where some genes were amplified up to 4 times.

363   Such examples include vesicle-mediated transport regulating genes, *Axl* and *Synj2bp*, which are

364   related to the vascular transport of proteins. The complete list of differentially expressed genes

16

365 with corresponding copy number differences are provided in Supplementary file.

366

3.5 Dysregulation of nucleotide and sterol metabolism may lead to slow growth in *Dhfr*-

deficient cell lines

369 Since majority of the differentially expressed genes were from cellular metabolism, we integrated

370 the transcriptome and proteome data with the CHO genome-scale metabolic network (Hefzi et al.,

371 2016) to unravel the transcriptional signatures of the CHO metabolism. The transcriptome and

372 proteome data were mapped onto the metabolic network using the available gene-protein-reaction

373 (GPR) rules. The microarray-acquired transcriptome data maps to 1287 of 1764 gene loci in the

374 model while the MS-acquired proteome data is linked to 548 genes. Following the transcriptome

375 and proteome data mapping, we assessed the enrichment of various metabolic pathways among

376 the up-/down-regulated transcripts and proteins (see **Methods**; **Figure 4a**). As expected, the purine

377 nucleotide metabolism is grossly dysregulated in both DG44 and DXB11 when compared to K1

378 due to the absence of *Dhfr* gene. Particularly, adenylosuccinate lyase (*Adsl*), an enzyme which

379 converts the succinylaminoimidazole carboxamide ribotide (SAICAR) into key AMP precursor,

380 aminoimidazole carboxamide ribotide (AICAR), was downregulated by more than 3.5 times in

381 both cell lines (**Figure 4b**). Similarly, nicotinamide nucleotide adenylyltransferase 2 (*Nmnat2*)

382 which catalyzes an essential step in NAD biosynthetic pathway was also downregulated by more

383 than 3 times in both cell lines. Apart from nucleotide metabolism, interestingly, several genes in

384 the cholesterol biosynthesis including *Hmgcs1*, *Fdps*, *Idi2*, *Lss*, *Ebp* and *Hsd17b7* were up-

385 regulated in DG44 and DXB11 (**Figure 4b**). Since endogenous cholesterol levels modulate the

386 epidermal growth factor receptor (EGFR) signaling and control meiosis (Gabitova et al., 2014),

387 elevated sterol levels in *Dhfr*-deficient cell lines could have a potential role in negatively regulating

17

388 cell cycle progression although it is unclear how exactly sterols control the cell cycle. ROS-

389 detoxifying glutathione metabolism was observed to be upregulated, possibly suggesting that these

390 cell lines have better ability for protein processing as an earlier study reported that higher levels of

391 glutathione were found in high-producer when compared to low-producer CHO cells (Orellana et

392 al., 2015).

393   To further examine whether such observed expression differences in the transcripts and

394 proteins of metabolic pathways ultimately affect the endogenous metabolite levels, we performed

395 LC-MS based metabolite profiling of DG44 and K1 cells. Noticeably, this analysis showed several

396 metabolites in central metabolism, e.g., glyceraldehyde-3-phosphate and lactate, and energy

397 generation (NAD and FAD), are much abundant in K1 (**Figure 4c**). On the other hand, the levels

398 of metabolites that are associated with cellular stress and growth inhibition such as oxidized

399 glutathione and CDP-choline were observed to be higher in DG44 (**Supplementary file**).

400

401 **3.6 N- and O-glycosylation of host cell proteins from parental CHO cell lines is discernably**

402 **different**

403 The considerable genetic heterogeneity of various parental CHO cells is also thought to influence

404 the quality attributes such as N- and O-glycosylation, charge variants and sequence variants of the

405 recombinant proteins produced from them. Therefore, in order to analyze the possible variations

406 in quality attributes of the protein products, we profiled the N- and O-glycosylation of the host cell

407 proteins. Corresponding glycomic analyses successfully detected 38 different N-glycan species

408 which are usually processed in the ER, cis-Golgi, medial-Golgi and trans-Golgi compartments

409 (See Methods). Comprehensive analyses of the resulting N-glycan species and their relative

410 abundances revealed several interesting aspects. The initial processing in ER is not likely to be

18

411   different among the cell lines; the high mannose structures were found to be relatively similar.

412   However, there exists an appreciable difference in downstream processing of N-glycans. The

413   terminal GlcNAc containing hybrid structures processed at the cis-Golgi were slightly higher in

414   DXB11 than other two cell lines while the complex tri-antennary structures which are processed

415   in medial- and trans-Golgi were higher in DG44 and K1 (**Figure 5a**). Notably, K1 cells had much

416   higher final capping of glycan species, i.e. sialylation, beyond which no further sugar moieties can

417   be added. Apart from N-glycans, we also measured the relative abundances of O-glycans including

418   three major structures, T antigen, sialyl-T antigen and diasialyl-T antigen. Further, the sialylated

419   structures were found to be more abundant in K1 than the other two cell lines even in O-glycans

420   (**Figure 5b**).

421   We subsequently correlated the variations in N-glycan species with the transcriptomic and

422   proteomic data of the N-glycosylation associated pathways (**Figure 5c**). Although there was no

423   appreciable difference in the expression of genes in early processing of N-glycans, one notable

424   exception is that the family of N-acetylglucosaminyltransferase genes (*Mgat2*, *Mgat4b* and *Mgat5*)

425   were up-regulated in DXB11. Consistent with this observation, the terminal GlcNAc structures

426   were slightly abundant in DXB11. In addition, the high levels of terminal galactose structures were

427   correlated well with the up-regulation of two beta-1,4-galactosyltransferase isoforms (*B4galt4* and

428   *B4galt5*), the enzymes which add galactose to the GlcNAc structures, and the UDP-galactose

429   transporter (*Slc35b1*). Similarly, four isoforms of beta-galactoside alpha-2,3-sialyltransferase

430   (*St3gal1*, *St3gal2*, *St3gal5* and *St3gal6*) were all up-regulated in K1, supporting the high levels of

431   CMP-sialic acid structures (**Figure 4c**).

432

19

4 DISCUSSION

Currently, 7 out of the top 10 drugs are industrially produced from CHO cells where various companies use different CHO host cell lines empirically based on their experience and regulatory track records. While it has been argued that the possible differences in the CHO host cells could present appreciable variations in therapeutic protein production which could be further complicated by the clonality, still the literature available on comparing various CHO hosts is relatively scarce. Therefore, in order to evaluate the phenotypic traits and bioprocessing potentials of CHO host cell lines and provide a basis for rational cell line selection, we compared the three most widely used CHO cell lines via multi-omics profiling for the first time ever in this study. As such, our analysis uncovered that the $Dhfr^-$ cell lines, i.e. DG44 and DXB11, grow much slowly due to the mutations in purine nucleotide biosynthesis pathways, which retarded the cell growth directly as well as indirectly by interfering with DNA replication and cell cycle progression. Interestingly, this study also showed that the passenger deletions in both $Dhfr^-$ cell lines have critical impact on cell cycle progression and growth factor signaling as the loss or gain of gene functions associated with these pathways reflected well at the transcriptomic and proteomic levels. Moreover, the N- and O-glycomic profiling of host cell proteins in all the three cell lines revealed notable differences in the resulting glycosylation, thus highlighting the need to select a relevant cell line for the target protein of interest based on the preferred glycosylation patterns.

Consistent to earlier reports (Florin et al., 2011; Sommeregger et al., 2013), our study confirmed the ability of K1 cells to grow faster than the DG44 and DXB11 under suspension cultures. The transcriptomic and proteomic profiling uncovered that this physiological difference could be a consequence of two scenarios: 1) the lack of $Dhfr$ gene leading to severe limitation for nucleotide synthesis and/or 2) the negative regulation of cell cycle via multiple signaling pathways

20

456   including PI3K-Akt, MAPK and ERK pathways, and a possible pathway activated by endogenous

457   sterol levels (Gabitova et al., 2014). In this sense, overexpression of the several down-regulated

458   cyclins such as *Ccny, Ccnl2, Ccnc and Ccnh*, and serine/threonine kinase, *Akt*, could be an

459   interesting strategy for improving the cellular growth in *Dhfr⁻* cell lines as it has been shown

460   earlier in several mammalian cells (Jaluria et al., 2007).

461   In addition to the inhibition of cell cycle progression, several genes associated with cellular

462   macromolecular assembly and component biogenesis were down-regulated in both DG44 and

463   DXB11. Here, it should be highlighted that such observations of varied cellular structure is in good

464   agreement with earlier studies which reported the size of ER and mitochondrial mass in K1 is

465   higher than that of DXB11 (Hu et al., 2013). From a bioprocessing point of view, a larger ER

466   provides significant advantage for K1 cells as it could handle much more proteins for their proper

467   folding and further processing than the other two cell lines. The limitations in the ER size could

468   also be a possible reason for growth inhibition in *Dhfr⁻* cells; since the ER size is limited, as the

469   unprocessed proteins accumulate, the unfolded protein response pathways could be activated,

470   which in turn could trigger the ROS stress pathways, thus ultimately invoking the apoptosis

471   pathway (Hu et al., 2013). The multi-omics data generated in this study supports this hypothesis

472   at several levels: both transcriptome and proteome showed the up-/down-regulation of several

473   apoptosis activators and inhibitors, respectively. Subsequent metabolome analyses confirmed the

474   high levels of oxidized glutathione, a ROS stress marker, possibly highlighting the relevance of

475   elevated ROS stress in DG44 than K1. However, whether all such pathways operate in tandem has

476   to be confirmed with further experiments.

477   Our study pinpointed that the N- and O-glycosylation patterns of protein products could

478   have cell line specific variation due to the inherent characteristics of CHO hosts. For example,

479   DXB11 have high amounts of less processed structures, particularly terminal GlcNAc glycans, and

480   K1 show much more sialylated structures in both N- and O-glycosylation profiles, which is

481   possibly attributed to high expression of relevant biosynthetic enzymes as well as high levels of

482   precursor availability. Notably, these observations are consistent with an earlier study which

483   compared the glycosylation profile of an antibody produced from DG44 and K1 cells and observed

484   that K1 produced much higher sialylated structures (Reinhart et al., 2018). While we observed no

485   gene copy number difference in the glycosylation biosynthetic pathway, all the cell lines have their

486   own expression preferences in each step of the glycan processing as different isoforms were up-

487   /down-regulated. Such variation in isoforms expression could possibly be a result of differences in

488   SNPs (Lewis et al., 2013) or due to changes in epigenetic control which await experimental

489   verification. Although the variations of glycosylation patterns can be modulated to a certain degree

490   by modifying the media and feeding strategies, the selection of an appropriate host depending on

491   the product of interest would nonetheless be relevant to achieve consistent quality. To this end, K1

492   and DG44 are probably more suitable for producing erythropoietin (EPO) or interferon beta 1

493   (IFNB1) as these molecules have complex tetra-/tri-antennary N-glycan species whereas DXB11

494   could provide a minor advantage for producing monoclonal antibodies since they have higher

495   terminal GlcNAc structures that are usually bi-antennary.

496   In summary, our work exploited the wealth of multi-omics technologies to systematically

497   link the possible phenotypic differences across CHO host cell lines with their genotypes. It also

498   clearly highlighted that the genetic heterogeneity of CHO cells indeed affects their growth

499   phenotype as well as N- and O-glycosylation profiles which could impact the bioprocessing steps.

500   Although the integration of transgene could further alter the genomic landscape of host cells

501   compared here, particularly DG44 and DXB11 due to the additional selection step, we believe that

22

502 most results unraveled in this study should hold good since the transgene targets only fragile

503 regions of the genome and importantly the glycosylation profiles remain unaltered (Yusufi et al.,

504 2017). Therefore, the current study could well serve as a good starting point for the rational

505 selection of relevant CHO host cells as well as to establish cell line engineering targets.

506

Acknowledgements

511

Conflict of interest

513 The authors declare that there are no conflicts of interest.

514

References

516 Butler M, Spearman M. 2014. The choice of mammalian cell host and possibilities for

517   glycosylation engineering. *Curr. Opin. Biotechnol.* **30**:107–12.

518 D'Angelo R, Aresta S, Blangy A, Del Maestro L, Louvard D, Arpin M. 2007. Interaction of ezrin

519   with the novel guanine nucleotide exchange factor PLEKHG6 promotes RhoG-dependent

520   apical cytoskeleton rearrangements in epithelial cells. *Mol. Biol. Cell* **18**:4780–93.

521 Duronio RJ, Xiong Y. 2013. Signaling pathways that control cell proliferation. *Cold Spring Harb.*

522   *Perspect. Biol.* **5**:a008904.

523 Florin L, Lipske C, Becker E, Kaufmann H. 2011. Supplementation of serum free media with HT

524   is not sufficient to restore growth properties of DHFR-/- cells in fed-batch processes -

525   Implications for designing novel CHO-based expression platforms. *J. Biotechnol.* **152**:189–

526   93.

527   Gabitova L, Gorin A, Astsaturov I. 2014. Molecular pathways: sterols and receptor signaling in

528   cancer. *Clin. Cancer Res.* **20**:28–34.

529   Goh SL, Wang Q, Byrnes LJ, Sondermann H. 2012. Versatile Membrane Deformation Potential of

530   Activated Pacsin. *PLoS One* **7**:e51628.

531   Golabgir A, Gutierrez JM, Hefzi H, Li S, Palsson BO, Herwig C, Lewis NE. 2016. Quantitative

532   feature extraction from the Chinese hamster ovary bioprocess bibliome using a novel meta-

533   analysis workflow. *Biotechnol. Adv.* **34**:621–33.

534   Hefzi H, Ang KS, Hanscho M, Bordbar A, Ruckerbauer D, Lakshmanan M, Orellana CA, Baycin-

535   Hizal D, Huang Y, Ley D, Martinez VS, Kyriakopoulos S, Jimenez NE, Zielinski DC, Quek

536   LE, Wulff T, Arnsdorf J, Li S, Lee JS, Paglia G, Loira N, Spahn PN, Pedersen LE, Gutierrez

537   JM, King ZA, Lund AM, Nagarajan H, Thomas A, Abdel-Haleem AM, Zanghellini J,

538   Kildegaard HF, Voldborg BG, Gerdtzen ZP, Betenbaugh MJ, Palsson BO, Andersen MR,

539   Nielsen LK, Borth N, Lee DY, Lewis NE. 2016. A Consensus Genome-scale Reconstruction

540   of Chinese Hamster Ovary Cell Metabolism. *Cell Syst.* **3**:434–43.e8.

541   Hong JK, Lakshmanan M, Lee D-Y. 2018. Towards next generation CHO cell line development

542   and engineering by systems approaches. *Curr. Opin. Chem. Eng.* **22**:1–10.

543   Hu Z, Guo D, Yip SSM, Zhan D, Misaghi S, Joly JC, Snedecor BR, Shen AY. 2013. Chinese

544   hamster ovary K1 host cell enables stable cell line development for antibody molecules which

545   are difficult to express in DUXB11-derived dihydrofolate reductase deficient host cell.

546   *Biotechnol. Prog.* **29**:980–85.

547   Jaluria P, Betenbaugh M, Konstantopoulos K, Shiloach J. 2007. Enhancement of cell proliferation

548   in various mammalian cell lines by gene insertion of a cyclin-dependent kinase homolog.

549   *BMC Biotechnol.* **7**:71.

550   Kaas C, Kristensen C, Betenbaugh MJ, Andersen M. 2015. Sequencing the CHO DXB11 genome

551   reveals regional variations in genomic stability and haploidy. *BMC Genomics* **16**:160.

552   Kao FT, Puck TT. 1968. Genetics of somatic mammalian cells, VII. Induction and isolation of

553   nutritional mutants in Chinese hamster cells. *Proc Natl Acad Sci U S A* **60**:1275–81.

554   Kaufman RJ, Wasley LC, Spiliotes a J, Gossels SD, Latt S a, Larsen GR, Kay RM. 1985.

555   Coamplification and coexpression of human tissue-type plasminogen activator and murine

556   dihydrofolate reductase sequences in Chinese hamster ovary cells. *Mol. Cell. Biol.* **5**:1750–

557   59.

558   Könitzer JD, Müller MM, Leparc G, Pauers M, Bechmann J, Schulz P, Schaub J, Enenkel B,

559   Hildebrandt T, Hampel M, Tolstrup AB. 2015. A global RNA-seq-driven analysis of CHO

560   host and production cell lines reveals distinct differential expression patterns of genes

561   contributing to recombinant antibody glycosylation. *Biotechnol. J.* **10**:1412–23.

562   de Kreuk B-J, Nethe M, Fernandez-Borja M, Anthony EC, Hensbergen PJ, Deelder AM, Plomann

563   M, Hordijk PL. 2011. The F-BAR domain protein PACSIN2 associates with Rac1 and

564   regulates cell spreading and migration. *J. Cell Sci.* **124**:2375–88.

565   Lewis NE, Liu X, Li Y, Nagarajan H, Yerganian G, O'Brien E, Bordbar A, Roth AM, Rosenbloom

566   J, Bian C, Xie M, Chen W, Li N, Baycin-Hizal D, Latif H, Forster J, Betenbaugh MJ, Famili

567   I, Xu X, Wang J, Palsson BO. 2013. Genomic landscapes of Chinese hamster ovary cell lines

568   as revealed by the Cricetulus griseus draft genome. *Nat. Biotechnol.* **31**:759–67.

569   Liu Y, Beyer A, Aebersold R. 2016. On the Dependency of Cellular Protein Levels on mRNA

570   Abundance. *Cell.* **165**(3):535-50

571 Moniz S, Veríssimo F, Matos P, Brazão R, Silva E, Kotevelets L, Chastre E, Gespach C, Jordan P.

572   2007. Protein kinase WNK2 inhibits cell proliferation by negatively modulating the activation

573   of MEK1/ERK1/2. *Oncogene* **26**:6071–81.

574 North SJ, Huang H-H, Sundaram S, Jang-Lee J, Etienne AT, Trollope A, Chalabi S, Dell A, Stanley

575   P, Haslam SM. 2010. Glycomics profiling of Chinese hamster ovary cell glycosylation

576   mutants reveals N-glycans of a novel size and complexity. *J. Biol. Chem.* **285**:5759–75.

577 Orellana CA, Marcellin E, Schulz BL, Nouwens AS, Gray PP, Nielsen LK. 2015. High-antibody-

578   producing Chinese hamster ovary cells up-regulate intracellular protein transport and

579   glutathione synthesis. *J. Proteome Res.* **14**:609–18.

580 Reinhart D, Damjanovic L, Kaisermayer C, Sommeregger W, Gili A, Gasselhuber B, Castan A,

581   Mayrhofer P, Grünwald-Gruber C, Kunert R. 2018. Bioprocessing of recombinant CHO-K1,

582   CHO-DG44 and CHO-S: CHO expression hosts favor either mAb production or biomass

583   synthesis. *Biotechnol. J.* **1700686**:1700686.

584 Rhind N, Russell P. 2012. Signaling pathways that regulate cell division. *Cold Spring Harb.*

585   *Perspect. Biol.* **4(10)**:pii: a005942.

586 Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. 2006. XCMS: processing mass

587   spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and

588   identification. *Anal. Chem.* **78**:779–87.

589 Sommeregger W, Gili A, Sterovsky T, Casanova E, Kunert R. 2013. Powerful expression in

590   Chinese Hamster Ovary cells using bacterial artificial chromosomes: parameters influencing

591   productivity. *BMC Proc.* **7(Suppl 6)**:P25.

592 Urlaub G, Chasin LA. 1980. Isolation of Chinese hamster cell mutants deficient in dihydrofolate

593   reductase activity. *Proc Natl Acad Sci U S A* **77**:4216–20.

594   Urlaub G, Käs E, Carothers AM, Chasin LA. 1983. Deletion of the diploid dihydrofolate reductase

595   locus from cultured mammalian cells. *Cell* **33**:405–12.

596   Walsh G. 2018. Biopharmaceutical benchmarks 2018. *Nat. Biotechnol.* **36**:1136–45.

597   Wettenhall JM, Smyth GK. 2004. limmaGUI: a graphical user interface for linear modeling of

598   microarray data. *Bioinformatics* **20**:3705–6.

599   Wishart DS, Feunang YD, Marcu A, Guo AC, Liang K, Vázquez-Fresno R, Sajed T, Johnson D,

600   Li C, Karu N, Sayeeda Z, Lo E, Assempour N, Berjanskii M, Singhal S, Arndt D, Liang Y,

601   Badran H, Grant J, Serra-Cayuela A, Liu Y, Mandal R, Neveu V, Pon A, Knox C, Wilson M,

602   Manach C, Scalbert A. 2018. HMDB 4.0: The human metabolome database for 2018. *Nucleic*

603   *Acids Res.* **46(D1)**:D608–D617.

604   Woo IS, Jin H, Kang ES, Kim HJ, Lee JH, Chang KC, Park JY, Choi WS, Seo HG. 2011.

605   TMEM14A inhibits N-(4-hydroxyphenyl)retinamide-induced apoptosis through the

606   stabilization of mitochondrial membrane potential. *Cancer Lett.* **309**:190–8.

607   Wurm F. 2013. CHO Quasispecies—Implications for Manufacturing Processes. *Processes* **1**:296–

608   311.

609   Wurm FM, Hacker D. 2011. First CHO genome. *Nat. Biotechnol.* **29**:718–720.

610   Wurm FM. 2004. Production of recombinant protein therapeutics in cultivated mammalian cells.

611   *Nat. Biotechnol.* **22**:1393–8.

612   Xu N, Ma C, Ou J, Sun WW, Zhou L, Hu H, Liu XM. 2017. Comparative Proteomic Analysis of

613   Three Chinese Hamster Ovary (CHO) Host Cells. *Biochem Eng J* **124**:122–9.

614   Xu X, Nagarajan H, Lewis NE, Pan S, Cai Z, Liu X, Chen W, Xie M, Wang W, Hammond S,

615   Andersen MR, Neff N, Passarelli B, Koh W, Fan HC, Wang J, Gui Y, Lee KH, Betenbaugh

616   MJ, Quake SR, Famili I, Palsson BO. 2011. The genomic sequence of the Chinese hamster

617   ovary (CHO)-K1 cell line. *Nat. Biotechnol.* **29**:735–41.

618   Yusufi FNK, Lakshmanan M, Ho YS, Loo BLW, Ariyaratne P, Yang Y, Ng SK, Tan TRM, Yeo HC,

619   Lim HL, Ng SW, Hiu AP, Chow CP, Wan C, Chen S, Teo G, Song G, Chin JX, Ruan X, Sung

620   KWK, Hu WS, Yap MGS, Bardor M, Nagarajan N, Lee DY. 2017. Mammalian Systems

621   Biotechnology Reveals Global Cellular Adaptations in a Recombinant CHO Cell Line. *Cell*

622   *Syst.* **4**:530–42.e6.

623

624

Figure Legends

Figure 1. **Genetic and phenotypic differences of CHO parental cell lines**. a) The known genetic background of the three CHO cell lines studied, i.e. CHO-K1, DG44 and DXB11 (figure re-drawn based on Wurm and Hacker, 2011), b) viable cell densities and % viabilities of three parental cell lines grown in suspension cultures, c) residual concentrations of glucose and lactate, and d) concentration of glutamine and ammonia in the three cell cultures. The black, blue and red lines in each graph represent the CHO-K1, DG44 and DXB11 cell lines, respectively.
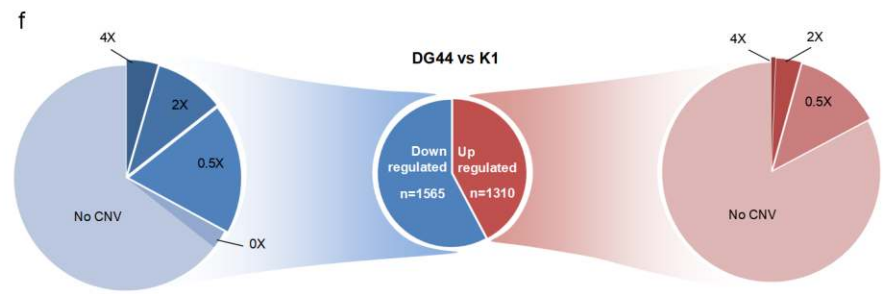
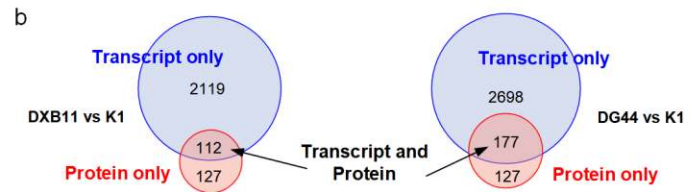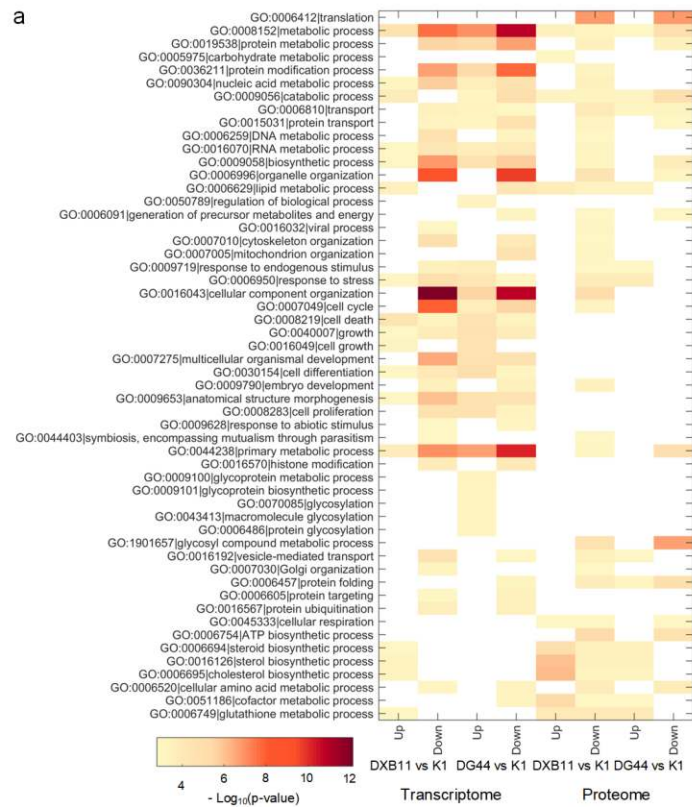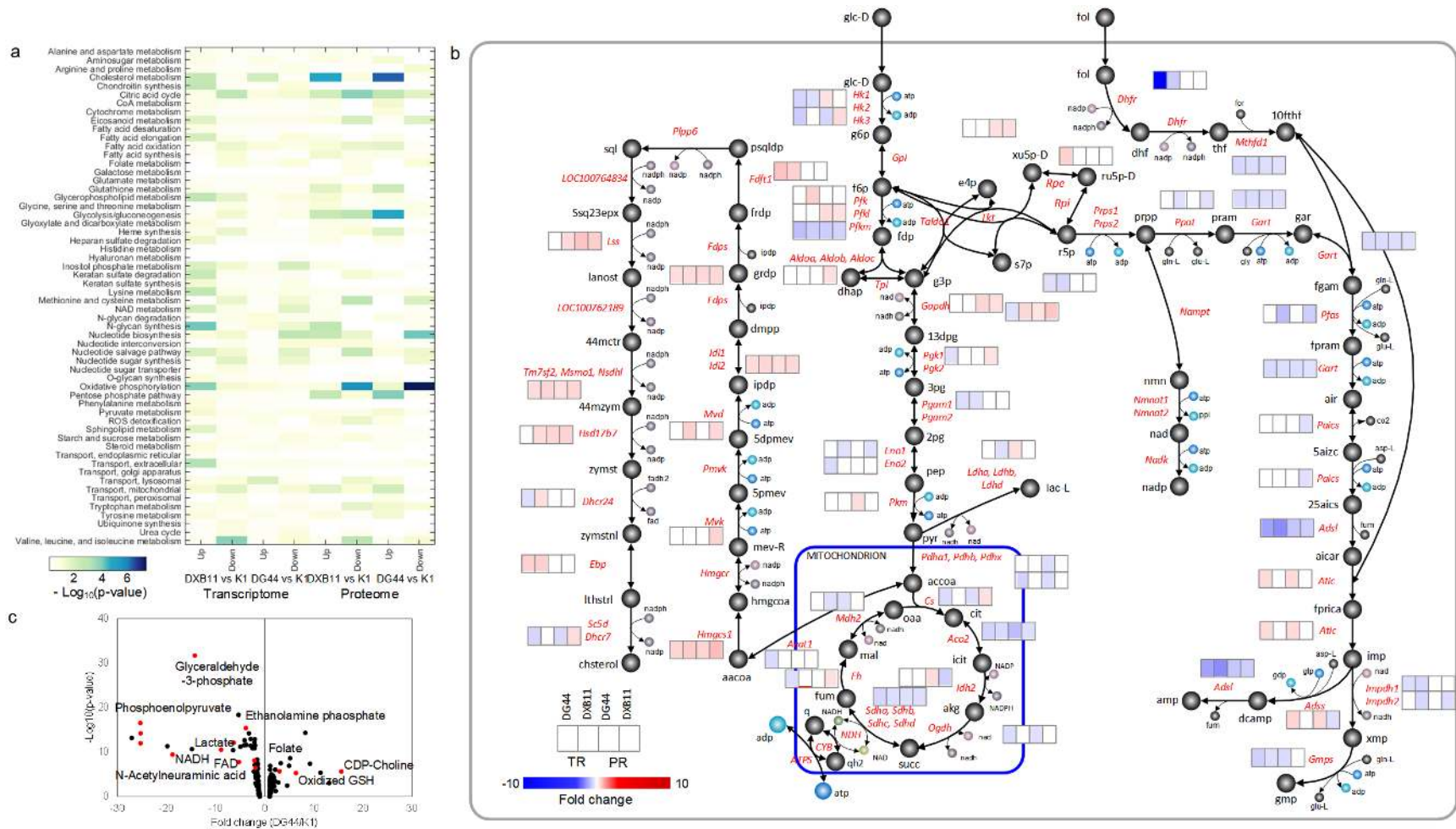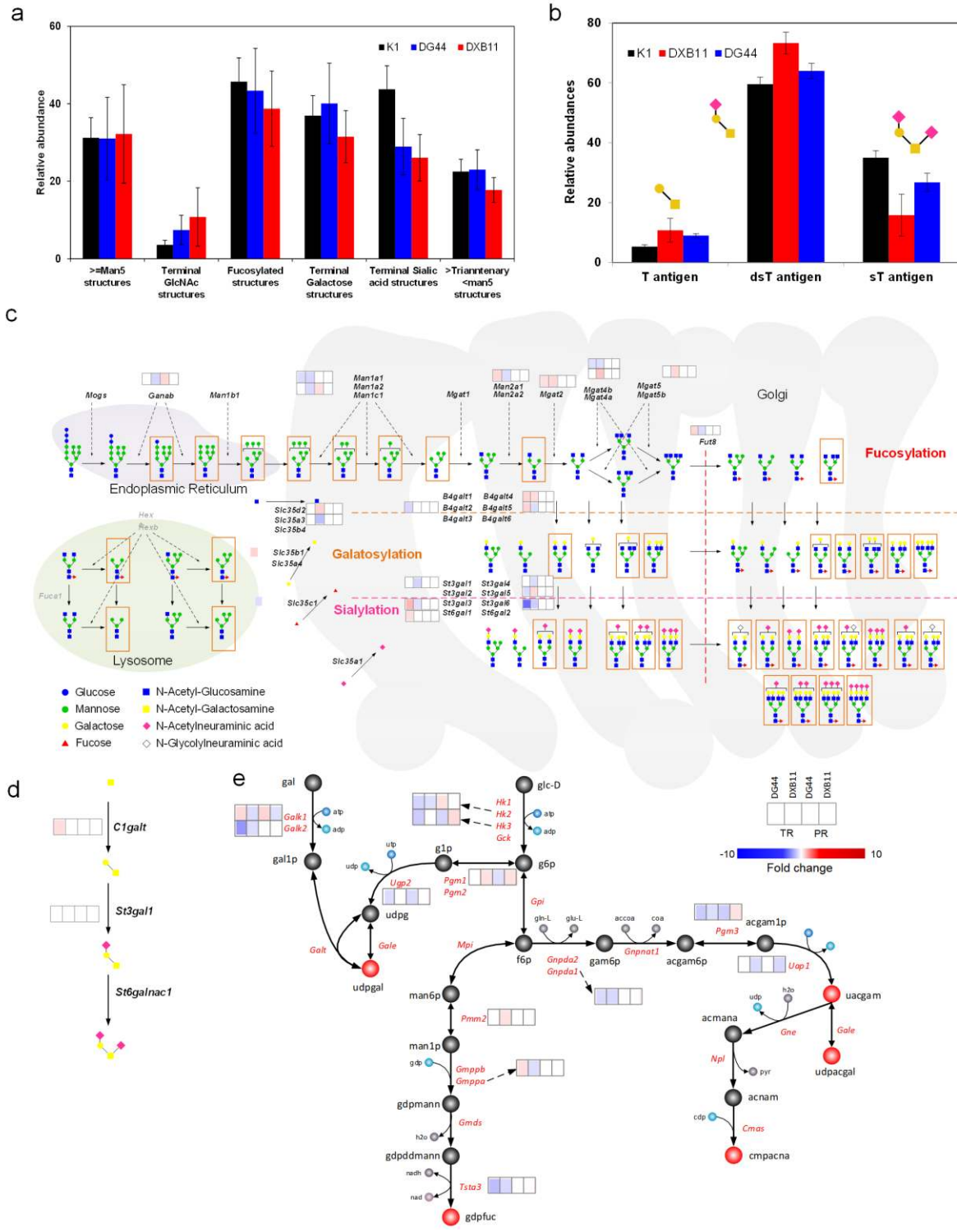**Figure 2**. **Global mRNA and protein expression patterns of CHO parental cell lines**. a) Global coverage of transcriptome and proteome data, b) PCA of the transcriptome, c) PCA of the proteome, d) hierarchical clustering of transcriptome and e) hierarchical clustering of proteome. Clustering was based on Euclidean distance of Pearson correlation coefficients.

Figure 3. **Differential expression signatures of CHO parental cell lines**. a) GO enrichment analysis of differentially expressed genes at both transcript and protein level , b) number of differentially expressed genes at transcript and protein level in DXB11 and DG44 w.r.t. K1, c) correlation between transcript and protein expression ratios in DXB11 vs K1, d) correlation between transcript and protein expression ratios in DG44 vs K1, e) the effect of genomic variations in differentially expressed genes at the transcript level in DXB11 vs K1 comparison and f) the effect of genomic variations in differentially expressed genes at the transcript level in DG44 vs K1 comparison.

Figure 4. **Differential expression signatures of metabolic genes in CHO parental cell lines**. a) The enrichment of various metabolic pathways in the CHO genome-scale network at both transcript and protein level and b) the differential expression of various genes from central metabolism, de novo purine biosynthesis and sterol metabolism at transcript and protein levels are shown in the metabolic map, and c) comparison of intracellular metabolite levels in DG44 and K1.

Figure 5. **Glycosylation differences between the CHO parental cell lines**. a) N-glycosylation differences, b) O-glycosylation differences, c) the differentially expressed genes and corresponding glycan species in N-glycosylation pathway, d) differentially expressed genes in O-

656   glycan pathway and e) the biosynthetic pathway of nucleotide-sugar donors. The glycans enclosed

657   in boxes in (c) were detected in our experiments.

Figure 1. Genetic and phenotypic differences of CHO parental cell lines. a) The known genetic background of the three CHO cell lines studied, i.e. CHO-K1, DG44 and DXB11 (figure re-drawn based on Wurm and Hacker, 2011), b) viable cell densities and % viabilities of three parental cell lines grown in suspension cultures, c) residual concentrations of glucose and lactate, and d) concentration of glutamine and ammonia in the three cell cultures. The black, blue and red lines in each graph represent the CHO-K1, DG44 and DXB11 cell lines, respectively.

Figure 2. Global mRNA and protein expression patterns of CHO parental cell lines. a) Global coverage of transcriptome and proteome data, b) PCA of the transcriptome, c) PCA of the proteome, d) hierarchical clustering of transcriptome and e) hierarchical clustering of proteome. Clustering was based on Euclidean distance of Pearson correlation coefficients.

32

Figure 3. Differential expression signatures of CHO parental cell lines. a) GO enrichment analysis of differentially expressed genes at both transcript and protein level , b) number of differentially expressed genes at transcript and protein level in DXB11 and DG44 w.r.t. K1, c) correlation between transcript and protein expression ratios in DXB11 vs K1, d) correlation between transcript and protein expression ratios in DG44 vs K1, e) the effect of genomic variations in differentially expressed genes at the transcript level in DXB11 vs K1 comparison and f) the effect of genomic variations in differentially expressed genes at the transcript level in DG44 vs K1 comparison.

33

678
679 **Figure 4. Differential expression signatures of metabolic genes in CHO parental cell lines.** a) The enrichment of various metabolic
680 pathways in the CHO genome-scale network at both transcript and protein level and b) the differential expression of various genes from
681 central metabolism, *de novo* purine biosynthesis and sterol metabolism at transcript and protein levels are shown in the metabolic map,
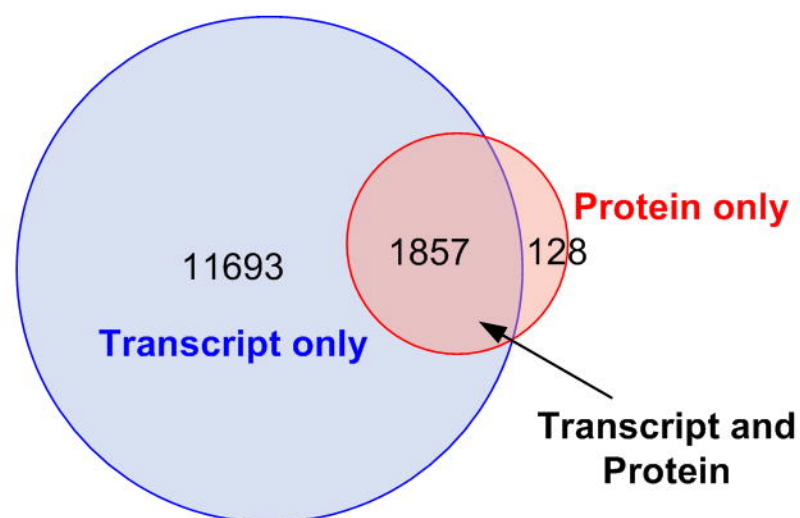682 and c) comparison of intracellular metabolite levels in DG44 and K1.

34

683

Figure 5. Glycosylation differences between the CHO parental cell lines. a) N-glycosylation differences, b) O-glycosylation differences, c) the differentially expressed genes and corresponding glycan species in N-glycosylation pathway, d) differentially expressed genes in O-glycan pathway and e) the biosynthetic pathway of nucleotide-sugar donors. The glycans enclosed in boxes 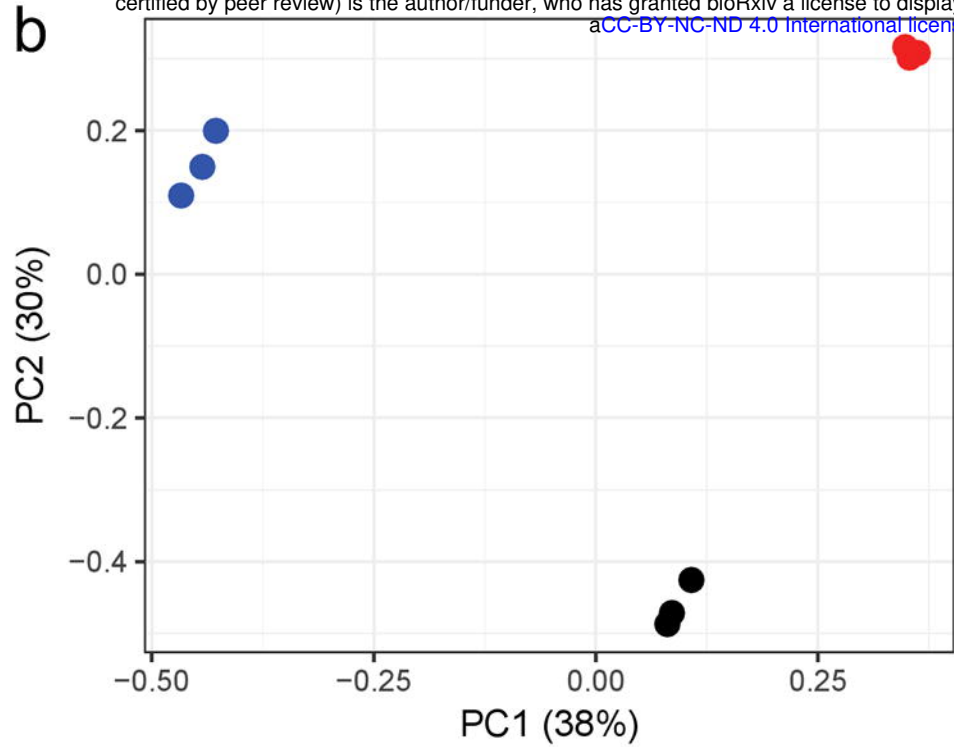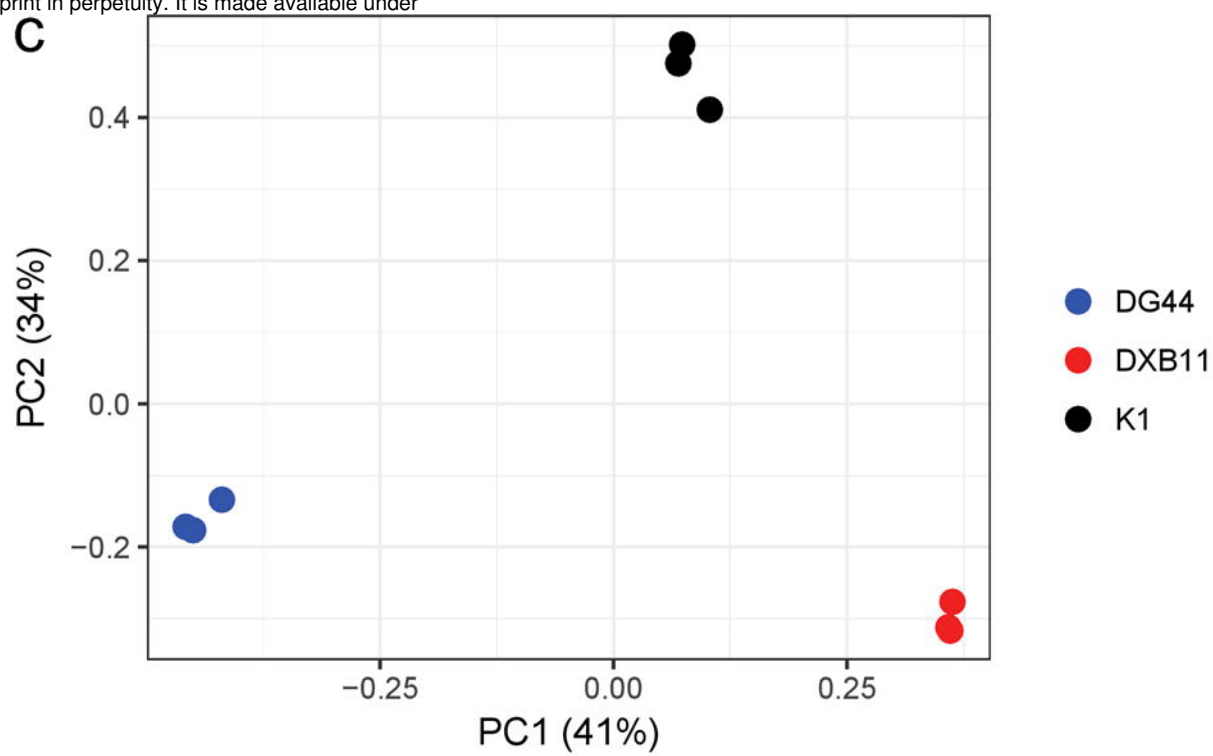in (c) were detected in our experiments.