Clemson University
TigerPrints

All Dissertations | Dissertations

5-2023

Assesment of Structure, Function, and Microevolutionary Dynamics of Extrachromosomal Circular DNA in Chinese Hamster Ovary Cells

Dylan Chitwood
dchitwo@clemson.edu

Follow this and additional works at: https://tigerprints.clemson.edu/all_dissertations

 Part of the Bioinformatics Commons, Biotechnology Commons, Genetics Commons, Genomics Commons, and the Molecular, Cellular, and Tissue Engineering Commons

Recommended Citation

Chitwood, Dylan, "Assesment of Structure, Function, and Microevolutionary Dynamics of Extrachromosomal Circular DNA in Chinese Hamster Ovary Cells" (2023). *All Dissertations*. 3296.
https://tigerprints.clemson.edu/all_dissertations/3296

This Dissertation is brought to you for free and open access by the Dissertations at TigerPrints. It has been accepted for inclusion in All Dissertations by an authorized administrator of TigerPrints. For more information, please contact kokeefe@clemson.edu.

ASSESMENT OF STRUCTURE, FUNCTION, AND MICROEVOLUTIONARY DYNAMICS OF EXTRACHROMOSOMAL CIRCULAR DNA IN CHINESE HAMSTER OVARY CELLS

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Bioengineering

---

by
Dylan G. Chitwood
May 2023

Accepted by:
Dr. Sarah W. Harcum, Co-committee Chair
Dr. Christopher A. Saski, Co-committee Chair
Dr. Mark Blenner
Dr. Angela Alexander-Bryant

ABSTRACT

Chinese hamster ovary (CHO) cell lines are among the most popular expression hosts used in biopharmaceutical manufacturing due to relative ease of culture, capacity to perform human-like post-translational modifications, and non-susceptibility to viruses. However, the intrinsic plasticity of the CHO genome can lead to undesired genetic rearrangements, phenotypic shifts, reduced product quality, and early culture termination that prevents continuous biomanufacturing. A characteristic of plastic and unstable genomes that is poorly understood in CHO cells is extrachromosomal circular DNA (eccDNA). EccDNAs are focal amplifications of the genome that reside in the extranuclear space. These plasmid-like entities are structurally complex and are capable of contributing to a wide variety of biological functions including gene overexpression, regulation of nuclear-encoded genes, immunostimulation, and adaptive stress responses.

The objective of this work is to establish the foundational knowledge of eccDNA structure, function, and microevolutionary dynamics in CHO cells under various conditions. This work characterizes eccDNA content in CHO cells grown in bioreactors for two weeks under control and lactate-stressed conditions, two CHO K-1-derived cell lines of different ages, and CHO cells gradually adapted to high extracellular lactate levels. More than 2,000 genes were observed to be encoded on eccDNAs and summaries of gene function are presented using Gene Ontology and KEGG pathway analyses. RNA-seq data is utilized to identify potential changes in gene expression mediated by

eccDNAs. Furthermore, the study presents a broad characterization of eccDNA sequence structures and biogenesis sites that may be used as targets in future work.

DEDICATION

This work is dedicated to my mother, Bobbie Jean Chitwood, as well as my grandparents, Bob and Patricia Chitwood. Most of my success is attributed to their unwavering love and support. I would not be where I am today without their encouragement.

ACKNOWLEDGMENTS

I am very grateful to have had two co-advisors as a graduate student, Dr. Sarah Harcum and Dr. Chris Saski. Working under the two of them has provided me with a very unique opportunity to develop a broad set of skills as a scientist. Their guidance and belief in me were instrumental in my success as a graduate student. Further, I would like to acknowledge my committee members for their feedback and guidance.

I would also like to thank current and former members of the Harcum and Saski research groups for their camaraderie and support through graduate school. Further, I would like to acknowledge current and past members of Clemson's Research and Education in Disease Diagnosis and Intervention (REDDI) Lab for providing a sense of community during the COVID-19 pandemic. Special thanks is due to the co-authors of this work who helped to make it the best it could be. I'd like to recognize the organizations that have funded me and my research over the years: the National Science Foundation, the Advanced Mammalian Biomanufacturing Innovation Center, the Clemson REDDI Lab, the National Institute for Innovation in Manufacturing Biopharmaceuticals, and the Clemson University Vice President of Research.

Finally, I want to thank my friends, family and colleagues that have encouraged and supported me through my academic career. Being a first-generation student has come with many challenges, however, the guidance and support I have received are unmatched. I could not have done this without the incredible community around me.

ABBREVIATIONS

Abbreviations	Full Form
α-KG	α-ketoglutarate
A-CoA	Acetyl-CoA
Ala	Alanine
BP	Biological process
CC	Cellular component
CHO	Chinese hamster ovary
CIDER-seq	Circular DNA enrichment sequencing
CMV	Cytomegalovirus
CQAs	Critical quality attributes
DM	Double minute
DMSO	Dimethylsulfoxide
DO	Dissolved oxygen
DSB	Double-strand break
eccDNA	Extrachromosomal circular DNA
ERC	Extrachromosomal ribosomal DNA
FDA	Food and Drug Administration
gDNA	Genomic DNA
Glc	Glucose
Gln	Glutamine
Glu	Glutamate
GMP	Good Manufacturing Practices
GO	Gene ontology
HCP	Host cell protein
HDR	Homology-directed repair
HiFi	High fidelity
i-loops	Internal loops
IGV	Integrated Genome Viewer
INDEL	insertion/deletion

Abbreviations	Full Form
Lac	Lactate
LINE	Long interspaced nuclear element
LTC	Long-term cultured cells
LTR	Long terminal repeat
mAb	Monoclonal antibody
MF	Molecular function
MMR	Mismatch repair
MSI	Microsatellite Instability
mtDNA	Mitochondrial DNA
NCBI	National Center for Biotechnology Information
ncRNA	non-coding RNA
NHEJ	Non-homologous end joining
NIH	National Institutes of Health
ORI	Origin of replication
PBS	Phosphate buffered saline
PDL	Population doubling level
PID	Proportional Integral Derivative
PMID	PubMed ID
Pyr	Pyruvate
qp	Cell-specific productivity
RCA	Rolling circle amplification
rDNA	Ribosomal DNA
ROS	Reactive oxygen species
rRNA	Ribosomal RNA
SINE	Short interspaced nuclear element
SMRT	Single-molecule Real Time
SNP	single nucleotide polymorphism
spcDNA	Small polydispersed circular DNA
t-circles	Telomeric circles

Abbreviations	Full Form
TMM	Trimmed Means of M
tRNA	Transfer RNA
VCD	Viable cell density
WGS	Whole genome sequencing

TABLE OF CONTENTS

Page

Table of Contents (Continued) Page

LIST OF TABLES

LIST OF FIGURES

List of Figures (Continued)

List of Figures (Continued)

CHAPTER 1

INTRODUCTION

1.1 Motivation

Global sales of biopharmaceuticals reached a new high of $343 billion in 2021, a 44% increase from 2017. Notably, monoclonal antibody (mAb) therapies accounted for $217 billion, or 63% of all sales[1]. Undoubtedly, the COVID-19 pandemic was a driving force in innovation and economic development in the biopharmaceutical industry. Chinese hamster ovary (CHO) cell lines are the most widely used mammalian cell culture system for biomanufacturing; used to produce 89% of products manufactured in a mammalian system[1]. CHO cell lines are broadly used because of the relative ease of culture, availability of genetic engineering tools, and capability to produce high titers of recombinant protein; however, the plasticity of the CHO genome is a driver of cell line instability through chromosome rearrangement, transgene exclusion, and accumulation of somatic variants in critical DNA repair genes[2]. A variety of efforts in both process and genetic engineering have been attempted to improve genome instability with varying results.

Recently, advancements in library preparation protocols and sequencing technology have enabled the study of extrachromosomal circular DNAs (eccDNAs). EccDNAs have been described as hallmarks of unstable genomes[3] and have been broadly identified across multiple species including yeast[4], plants[5], and humans in both normal and

diseased states[6]. CHO cells have been previously observed to harbor eccDNAs, however, the most recent study of eccDNAs in CHO was published in 1986[7]. Recent research has described multiple functions of eccDNA including gene overexpression, immunostimulation, and DNA mobilization[8]. The broad prevalence of eccDNAs likely indicates a conserved biological function. EccDNAs have been showed to confer genetic heterogeneity that can be acted upon as a latent stress response or mechanism of adaptation[9]. In biopharmaceutical manufacturing, maintaining homogeneity of the cell phenotype ensures consistent productivity and quality of biopharmaceuticals in accordance with the FDA's Good Manufacturing Practices (GMPs).

Currently, little is understood about the composition, function, and evolution of eccDNAs in CHO cell lines used for biopharmaceutical manufacturing. In this dissertation, the novel circular DNA enrichment sequencing (CIDER-seq) protocol[10] was used to prepare eccDNA libraries for PacBio single-molecule real time (SMRT) sequencing. EccDNA and RNA-seq data sets were integrated with cell culture data to identify phenotypic shifts that may be attributed to eccDNA-mediated gene expression. These analyses were conducted in three contexts: an industry-standard, two-week fed-batch culture under control and stressed conditions, young and aged cells, and in cells gradually adapted to high levels of extracellular lactate. Key components of eccDNA structure, including repeat content, genes, and origins of replication, were annotated to identify shifts in composition in response to the tested conditions. Gene expression data was collected and integrated to identify transcriptome level responses to advanced cell age and lactate adaptation. The findings of this work establish a baseline understanding of

how eccDNA content and dynamics influence cell line performance and potentially identify genetic engineering targets for cell line improvement. Furthermore, eccDNA sequences generated from this study represent the first publicly available dataset of its kind in CHO cells. Lastly, RNA-seq data also provides additional public datasets in response to lactate stress, increasing cell age, and lactate adaptation for further study by the CHO community.

1.2 Organization

This dissertation is divided into seven chapters. Chapter One describes the motivation behind this work. Chapter Two presents a review of relevant literature regarding Chinese hamster ovary (CHO) cell lines, and extrachromosomal circular DNA (eccDNA). This includes a background on the origins and applications of CHO, cell line instability issues, common sources of culture stress, and a brief history of tools used for studying CHO 'omics. A thorough background on genome instability and the role of eccDNA is also presented. Specifically, eccDNA discovery, structure, and functions are covered in detail. Chapter Three consists of a recently published research paper entitled "*Microevolutionary dynamics of eccDNA in Chinese hamster ovary cells grown in fed-batch cultures under control and lactate-stressed conditions*"**.** This paper examines the structure and dynamics of eccDNAs grown under control and lactate-stressed conditions in fed-batch ambr®250 bioreactors. I was the primary author of this paper; co-authors are Dr. Qinghua Wang, Stephanie R. Klaubert, Kiana Green, Dr. Cathy H. Wu, Dr. Sarah W. Harcum, and Dr. Christopher A. Saski[11]. I was partially responsible for study design and conception, and fully responsible for conducting experiments, analyzing and interpreting sequencing data,

primary drafting of the manuscript, and addressing reviewer comments. Chapter Four characterizes eccDNA content and transcriptome shifts in two cell lines, VRC01 and CHOZN GS23, across varying ages. It is intended that this chapter will be submitted to a peer reviewed journal in the near future under a working title of "*Comparison of eccDNA evolution, gene expression changes, and genome structure variation between two reference Chinese hamster ovary cell lines in young and aged cells*". I will be the primary author with Dr. Qinghua Wang, Dr. Lauren Cordova, Dr. Wanfang Fu, Dr. Kelvin Lee, Dr. Sarah W. Harcum, and Dr. Christopher A. Saski as co-authors. Chapter Five describes eccDNA structure and dynamics alongside gene expression changes in VRC01 cells adapted to grow in high levels of extracellular lactate. We intend to submit this chapter to a peer-reviewed journal in the future under the working title "*Dynamics of amino acid metabolism, gene expression, and circulomics in a recombinant Chinese hamster ovary cell line adapted to moderate and high levels of extracellular lactate*" where I am the primary author and co-authors are Lisa Uy, Dr. Wanfang Fu, Stephanie R. Klaubert,Dr. Sarah W. Harcum, and Dr. Christopher A. Saski. Chapter Six is a research paper published in 2021 entitled "*Characterization of Metabolic Responses, Genetic Variations, and Microsatellite Instability in Ammonia-stressed CHO cells grown in Fed-batch Cultures*". This paper identified microsatellite loci that exhibited unfaithful replication in a dose-dependent response to elevated levels of ammonia. I am the primary author of this work and co-authors are Dr. Qinghua Wang, Dr. Kathryn Elliott, Aiyana Bullock, Dwon Jordana, Dr. Zhigang Li, Dr. Cathy H. Wu, Dr. Sarah W. Harcum, and Dr. Christopher A. Saski[12]. My contribution to this work was analyzing microsatellite data, drafting the manuscript, and responding to reviewer comments. While this research was published

first, it is appended to the end of the dissertation to avoid interrupting the presentation of eccDNA-related data. Finally, Chapter Seven outlines conclusions and recommended future directions for study and potential engineering of eccDNAs in CHO cell lines.

CHAPTER 2

LITERATURE REVIEW

2.1 Chinese hamster ovary (CHO) cells

2.1.1 Origins and industrial applications

The origin of Chinese hamster ovary (CHO) cell lines can be traced back to a spontaneous immortalization event in the late 1950's[13]. The CHO-ori cell line was used heavily in research due to its utility in studying the structure of the mammalian genome; of particular interest to many researchers were the observable chromosomal rearrangements[14]. Throughout the years, cell populations with desirable phenotypes were selected as a result of mutagenesis or spontaneous variation; however these selection processes were poorly documented[15]. CHO cells were adapted independently by many research groups to grow in suspension culture, which allowed for better scalability and higher viable cell densities (VCDs)[16-18]. Multiple factors including ease of culture[19], capability to perform post-translational modifications[20], and availability of tools for genetic modification[21], have strengthened CHO cells as a host for biopharmaceutical production. As of 2018, 84% of all monoclonal antibody (mAb) products were produced in a CHO cell line[22]. Typically, CHO cell cultures are grown in fed-batch cultures over a period of a few weeks[23,24]. A fed-batch culture normally maintains high cell viability by providing supplemental nutrients on a regular basis. Termination of fed-batch cultures usually occurs at 2 weeks when cell viability drops below 70% due to waste accumulation[25] and/or cell line instability[26].

2.1.2 Cell line instability and continuous biomanufacturing

Continuous manufacturing is currently the goal for the biomanufacturing industry; however, it is limited by cell line instability and shortcomings in process development[2]. The inherent plasticity of the CHO cell genome allows for a wide range of variations that make the cells more adaptable[2]. In nature, adaptability can be beneficial[27]; however, in biomanufacturing, it can cause unwanted instability[28-30]. Specifically, genome instability has been demonstrated to impact central carbon metabolism[31], exclude transgenes[32], and interrupt coding sequences of critical genes such as those responsible for DNA repair[33]. In addition to losing productivity, transgene exclusion can lead to culture decline by losing the selectable marker gene. Often, CHO hosts are deficient in genes such as glutamine synthetase[34] or dihydrofolate reductase[35]. Disruptions within DNA repair genes can further worsen genome instability issues as replication errors or mutations accumulate globally throughout the genome as a result[33,36]. This genome instability causes phenotypic shifts that make the cell line less productive[37]. Cell line instability, also called phenotypic drift, can also cause improper post-translational modification[38], increase cell lysis which leads to higher host cell proteins (HCPs) that need to be removed in downstream purification[39], partial or isoform products[37], and/or premature culture termination[2]. Improving culture stability has been attempted through numerous process and genetic engineering approaches. Genetic engineering approaches include promoter optimization[40], site-specific integration of the transgene[41], and insertion or deletion of beneficial or detrimental genes, respectively. Process engineering efforts to enhance the viability of long-term cell culture techniques, such as perfusion, have greatly extended

culture lifespan[42,43]. Thus, creating a need for a better understanding of cell line instability in extended cultures.

2.1.3 Culture stresses

2.1.3.1 Process-induced stress

Modern, computer-controlled bioreactors offer tight control of culture parameters such as pH, dissolved oxygen (DO), stir speed, gas sparging, and nutrient addition[44]. In bioreactor systems, pH is typically controlled using a base such as sodium bicarbonate or CO_2 gas[45]. Culturing cells at an improper pH can impede cell growth[46], unbalance redox metabolism[47], and negatively impact critical quality attributes (CQAs)[48,49]. Proper tuning of PID (proportional integral derivative) control loops is necessary to ensure the pH setpoint is maintained without unintended consequences[50]. Over addition of base or CO_2 can cause increased osmotic pressure that can negatively impact cell physiology[51,52]. DO PID control loops ensure proper oxygenated[53] by balancing agitation rates[47] with too high DO setpoints that would favor the formation of reactive oxygen species (ROS)[54].

2.1.3.2 Metabolic stress

Accumulation of metabolic waste products is another source of culture stress[25]. Two specific metabolic wastes, ammonia and lactate, are among the most studied CHO culture stressors. Ammonia is generated through multiple metabolic pathways including deamination of alanine, serine, glutamine, glutamate, and asparagine[55]. Elevated levels of ammonia have been shown to reduce intracellular pH[56], decrease cell growth and productivity[57], modify glycosylation patterns[56,58], induce gene expression shifts[58], and

weaken genome stability[12]. Lactate is primarily formed via conversion of pyruvate to lactate; however, parallel pathways can affect the rate of lactate production and consumption such as deamination of alanine[58] and glutamine depletion[59]. CHO cultures often have an initial accumulation of lactate and, under the right conditions can switch to lactate consumption[60-62]. The exact mechanism of this shift is unclear; however some studies suggest nutrient depletion[59,63-65], perturbation of pH and/or temperature[66,67], and increased oxidative capacity[68]. The switch to lactate consumption is critical as lactate accumulation inhibits cell growth[69] and acidifies uncontrolled cultures[70]. Ammonia and lactate metabolic pathways have both been the targets of gene knockdown to mitigate formation[71]. Further process control methods have examined adaptation of cells to media containing waste products[72], pH-based delivery of glucose[73], limiting addition of certain amino acids[74], and feeding lactate to cultures[75]. Multiple stress response mechanisms have been observed within CHO cells at the substrate[58], transcriptomic[76], and epigenetic[77] levels to mitigate environmental and metabolic challenges, but less has been reported on the relationship of lactate and ammonia to CHO cell genome instability[12].

2.1.4 CHO Reference Genome

Studying the CHO genome became more accessible in 2011 when the first reference genome for the CHO K-1 cell line was published[78]. This first publication provided large contigs, or scaffolds, comprising 2.45 Gb were used to develop the 21 chromosome-scale scaffolds and more than 24,000 genes were annotated. Despite inconsistencies in the number of chromosomes reported in previous CHO K-1 karyotypes, karyotypes performed to validate the reference genome confirmed the existence of 21

non-overlapping, chromosome-scale scaffolds. Many smaller scaffolds were still unplaced in the chromosome-scale scaffolds. The current standard for the CHO reference genome, known as PICRH from the Chinese hamster, was released in 2019 and contains 97% of the genome in 11 chromosomes or mega-scaffolds[79]. Unplaced scaffolds in this construction were reduced from 1,830 to 647. Further advancements in sequencing technology, such as the development of single molecule real time (SMRT) sequencing, would allow for more in depth characterization of 'omics datasets.

2.2 Genome instability and extrachromosomal circular DNA

2.2.1 Causes and implications of genome instability

Genome instability is the accumulation of variants throughout a genome[28,80]. Instability can refer to single-nucleotide changes such as single nucleotide polymorphisms (SNPs), insertions/deletions (INDELs)[81], and larger structural changes such as chromosome rearrangement[82,83], or changes in ploidy[84]. Variation in the genome can occur through a variety of mechanisms such as replication stalling[85] or replication fork collapse[86,87], genotoxic compounds in the environment[88-90], double-strand breaks[91], regulatory errors[92], and chromothripsis[93]. A genome become more unstable when variants or rearrangements arise within coding or regulatory regions of genes responsible for maintaining genome fidelity; specifically, genes responsible for mismatch repair[33,36,94-97], cell cycle regulation[98-100], and DNA damage repair[101-103]. Damage to these regions allows mutations to perpetuate in daughter cells and permeate into the cell population[104]. In multicellular organisms, genome instability is often the cause of carcinogenesis[105].

Biomarkers of genome instability have become an important tool in cancer diagnosis. One of the most well-known biomarkers of genome instability is the Bethesda Panel: a set of 13 microsatellite regions found to be unfaithfully replicated in instable human genomes[106]. Microsatellites are repetitive sequences that contain four or more repeats of a short nucleotide motif[107]. Due to the structure of microsatellite loci, sequences are often replicated incorrectly[108]. As such, assessing the length of well-studied microsatellites in humans can correlate with the degree of genome instability[109-111]. This technique has shown effectiveness in diagnosing multiple cancer types such as colorectal[110,112], gastric[113], and endometrial[114]. Recently, evidence of a new potential biomarker for genome instability-linked cancers has emerged in the form of extrachromosomal circular DNA (eccDNA)[8,115]. EccDNAs are considered hallmarks of an unstable genome[3,116]. EccDNAs are currently poorly understood in CHO cells and will be further studied in this work.

2.2.2 EccDNA discovery, prevalence, and classification

EccDNAs were first identified in boar sperm cells in 1964[117]. Since the initial observation, eccDNAs have been observed in multiple species across kingdoms including plants[118-120], yeast[4,121], drosophila[122], mice[123], HeLa cells[124], both somatic and cancerous human tissue[6,125-129], and CHO cells[7,130]. It is hypothesized that the broad prevalence of eccDNA could be indicative of a conserved biological function. EccDNA is a broad term that refers to all circular DNA within a cell. Well known subcategories of eccDNA include small poly-dispersed circular DNA (spcDNA), telomeric circles (t-circles), microDNA, extrachromosomal ribosomal DNA (ERCs), and double minute chromosomes (DMs); all

have varying size and functionality such as telomere maintenance[131], regulation of microRNA[132,133], ribosomal RNA templates[134,135], and gene amplification[119,136-138]. Recently, eccDNA functions identified have expanded to include intracellular communication[139] and innate immune responses[8,140].

2.2.3 EccDNA structure

2.2.3.1 Size and biogenesis

While structurally similar to plasmids, a distinguishing characteristic is that eccDNAs originate from the host genome[141]. EccDNAs come from a variety of error-prone biogenesis pathways including but not limited to recombination between tandem repeat regions[122,142], replication errors[143,144], chromothripsis[138], VDJ-like recombination[137], and episome aggregation[145]. EccDNAs can range in size from tens of bases to millions of bases[146]. Some eccDNAs have been observed to have a higher-order chromatin structure[147]. While eccDNAs exist in the extranuclear space, eccDNAs are capable of reintegration into the genome due to the sequence structure[9,145] and may act as mobile elements. Due to the repetitive structure of eccDNA, some hypothesize that this enables eccDNAs to recombine with multiple sequences from distal regions of the genome to form larger eccDNAs, though further study is required to support this hypothesis.

2.2.3.2 Repeat motifs

Many repetitive regions of the genome are highly susceptible to eccDNA biogenesis[122,123,134,142] which results in many eccDNAs harboring some form of repetitive motif. In *Drosophila*, eccDNAs were found to account for up to 10% of all repeat content

in the genome; these repetitive sequence structures were shown to initiate eccDNA biogenesis via intramolecular homologous recombination[122]. EccDNAs in *Amaranthus palmeri,* or pigweed, have been observed to contain simple repeats, long terminal repeats (LTRs), and other low complexity regions[119]. Furthermore, geographically distal populations of *A. palmeri* were observed to have some variation in distribution of repetitive motifs on eccDNAs[119]. Long interspaced nuclear elements (LINEs) were also identified in *A. palmeri*. In humans, LINEs tend to be transcriptionally silent, but are capable of retrotranscription and may cause disruptions in coding regions[148]. Other notable repetitive elements identified in *A. palmeri* eccDNAs are short interspaced nuclear elements (SINEs). SINEs are also retrotransposons and originate from transfer RNA (tRNA) sequences[149].

2.2.3.3 Origins of replication and centromeres

EccDNAs have been observed to contain sequence motifs consistent with known origins of replication[118]. In yeast, up to 80% of studied eccDNAs contain sequences that enable autonomous replication[4]. Autonomous replication of eccDNAs increases the impact of eccDNA on a cell's phenotype. Further, centromeres have yet to be identified on eccDNA as of this writing[9,138,150]. A lack of centromeres prevents equal segregation of eccDNAs between daughter cells. Uneven segregation can worsen heterogeneity in cell populations as some daughter cells may receive a high copy number of an eccDNAs while the other daughter cells could receive none.

2.2.4 EccDNA functions

2.2.4.1 Gene amplification and overexpression

Gene overexpression is the best characterized function of eccDNA. Excessive gene expression leads to phenotypic heterogeneity that can allow a subpopulation of organisms or cells to gain a selective advantage. An eccDNA amplified in germline cells can result in an entire organism with a selective advantage as seen in *A. palmeri*[5]. Specifically, eccDNA-mediated overexpression of the gene *Epsps* was shown to enable rapid metabolism of glyphosate[5], the active ingredient in commercial herbicides. This eccDNA resulted in geographically distal populations of *A. palmeri* with eccDNA-mediated glyphosate resistance[119]. Further, eccDNA amplified in a single cell can lead to mosaicism in the local tissue as seen in human cancers[9,151]. Overexpression of known oncogenes, such as *Myc*[152], via eccDNA has been observed in many cancer types, particularly those with poorer prognoses[153]. Additionally, gene amplification, such as the epidermal growth factor receptor (*Egfr*), can confer resistance to common therapeutics[154]. Mobility of eccDNAs allows for transposition and recombination between sequences to create large eccDNAs harboring multiple genes[5].

2.2.4.2 Transposable elements

Circular transposable elements, or helitrons, were first observed in *Arabidopsis*[155]. Helitrons are amplified via rolling circle amplification using a single-stranded DNA template[156,157] and go through a circular DNA intermediate when moving around the genome[146]. Functionally, helitrons have been observed to interrupt genes through insertion[158], and amplify portions of genes to create truncated proteins[159]. The

distinguishing features of helitrons are the 5’ motif T (C/T) and the 3’ motif CTRR. Furthermore, helitrons carry large, palindromic motifs ranging from 16 to 20 bp and are capable of forming a hairpin near the 3’ terminus[157]. By this convention, all helitrons are eccDNAs, but not all eccDNAs are helitrons.

2.2.4.3 Protein turnover

Previous studies in *Arabidopsis* and *A. palmeri* have identified a significant number of transfer RNA (tRNA) genes encoded on eccDNAs[119,120]. It is speculated that additional tRNA genes allows for more rapid adaptation by modulating protein turnover[160,161]. In *Saccharomyces cerevisiae*, when tRNA genes were intentionally removed, the deleted gene would eventually be restored by mutating another tRNA gene; this demonstrated that tRNA gene modifications are an essential adaptation mechanism[160]. Another study in *S. cerevisiae* showed that the tRNA pool is highly dynamic and tRNA abundance was hypothesized to be modulated by the cell to affect translation of proteins related to a diverse spectrum of external stresses[161].

2.2.4.4 Potential biomarker

EccDNAs are present in somatic and oncogenic human tissue[6,115,129]; however, are more prevalent in cancer tissue. This segregation is likely due to erroneous DNA repair and/or recombination genes in cancer cells[3]. Human plasma has been shown to harbor eccDNA as apoptosis and cell lysis leads to the release of eccDNAs into circulation[6]. Further, releasing eccDNAs in circulation has been shown to trigger an immune response in humans[140]. With technology now enabling better detection of

eccDNAs, more evidence is emerging that suggests eccDNAs could be reliable biomarkers for cancer[162]. The increasing importance of eccDNA in cancer diagnostics and treatment have led to the development of a shared database, eccDNAdb, which contains information on structure and gene content of eccDNAs found across 57 different cancer types[163]. A similar reference circulome or database of eccDNAs in CHO cells under different stresses across multiple cell lines would aid in potential identification of an eccDNA marker of genome instability.

2.2.4.5 Adaptation and two-speed genomes

The high frequency of eccDNAs originating from repeat-rich regions of the genome supports the hypothesis of two-speed genomes which was originally described in eukaryotic plant pathogens. Two-speed genomes refer to two subgenomes: a slower, gene-rich subgenome and a faster, repeat-rich subgenome[164]. Gene-rich regions tend to maintain stable “housekeeping” genes, those necessary for day-to-day functions within the cell, while repeat-rich regions contained genes that would benefit from variation, such as virulent effectors. A study of the *Fusarium graminearum* genome identified 350 genes in the faster, repeat-rich subgenome. These genes had lower GC content, were shorter in length, and had higher variation in exon counts when compared to the slower genome. Further, genes within the faster subgenome were found to be linked with pathogenicity; chromatin of the faster subgenome was found to be open during an infectious cycle[165]. EccDNAs have been characterized in a similar filamentous plant pathogen, *Magnaporthe oryzae. M. oryzae* eccDNAs were shown to have a broad diversity of transposable elements, genes, and virulent effectors[166]. Research on two-speed genomes is still

relatively new and has primarily focused on plant pathogens as of this writing. However, the susceptibility of repeat-rich regions of the genome to eccDNA biogenesis could imply a link between eccDNA and adaptation.

2.2.5 Future directions of eccDNA work

2.2.5.1 Detection methods

Advancements in sequencing technology and library preparation techniques have greatly enabled study of eccDNA in recent years; however, some optimization is still desired. Purifying eccDNA from linear genomic DNA is often overcome by rolling circle overamplification of circular DNA as described in CIDER-seq[10]. In this scenario, linear DNA is greatly underrepresented compared to circular DNA in the sample; however, this method is nonquantitative as smaller eccDNAs are capable of more rapid synthesis. Another source of contamination in eccDNA library preparation is mitochondrial DNA (mtDNA)[167]. In humans, mtDNA is a double-stranded, circular sequence of approximately 16,500 bp harboring 37 genes[168]. Recent work has demonstrated *in vivo* removal of mtRNA in mouse and human cell lines using specific CRISPR-Cas9 constructs to improve efficiency of eccDNA library preparation[169]. Quantifiable methods of eccDNA library preparation are highly desired to capture an accurate eccDNA profile in samples without using a prohibitive amount of starting material.

2.2.5.2 Trait engineering

Artificial eccDNA and ring chromosomes have been constructed using a CRISPR-Cas9 expression cassette[170]. Using HEK-293T cells as a model cell line, this work

showed that multiple eccDNA of varying sizes can be generated from multiple parts of the genome. While using eccDNA as an expression cassette would be a desirable goal for trait engineering, the variability and instability of eccDNA remains a challenge. As of this writing, little is known about successful optimization of native eccDNA, or synthesizing artificial eccDNA, to act as expression cassettes[171]. Arguably, the largest barrier to eccDNA engineering is the unequal segregation of eccDNAs between daughter cells as this creates unwanted phenotypic heterogeneity and inconsistent results[150]. Designing a reliable sequence motif to enable equal segregation of eccDNAs would be an ideal initial step in constructing an expression cassette. In nature, many viral genomes, such as Epstein Barr, have such genome persistence mechanisms[172], yet no such reliable sequence motif has been identified on eccDNA.

CHAPTER THREE

MICROEVOLUTIONARY DYNAMICS OF ECCDNA IN CHINESE HAMSTER OVARY CELLS GROWN IN FED-BATCH CULTURES UNDER CONTROL AND LACTATE-STRESSED CONDITIONS

3.1 Abstract

Chinese hamster ovary (CHO) cell lines are widely used to manufacture biopharmaceuticals. However, CHO cells are not an optimal expression host due to the intrinsic plasticity of the CHO genome. Genome plasticity can lead to chromosomal rearrangements, transgene exclusion, and phenotypic drift. A poorly understood genomic element of CHO cell line instability is extrachromosomal circular DNA (eccDNA) in gene expression and regulation. EccDNA can facilitate ultra-high gene expression and are found within many eukaryotes including humans, yeast, and plants. EccDNA confers genetic heterogeneity, providing selective advantages to individual cells in response to dynamic environments. In CHO cell cultures, maintaining genetic homogeneity is critical to ensuring consistent productivity and product quality. Understanding eccDNA structure, function, and microevolutionary dynamics under various culture conditions could reveal potential engineering targets for cell line engineering. In this study, eccDNA sequences were investigated at the beginning and end of two-week fed-batch cultures in an ambr®250 bioreactor under control and lactate-stressed conditions. This work characterized structure and function of eccDNA in a CHO-K1 clone. Gene annotation identified 1,551 unique eccDNA genes including cancer driver genes and genes involved

in protein production. Furthermore, RNA-seq data is integrated to identify transcriptionally active eccDNA genes.

3.2 Introduction

Chinese hamster ovary (CHO) cell lines are broadly used in the manufacturing of biopharmaceuticals due to ease of culture, adaptability to manufacturing processes, and tolerance to genetic manipulation[2,22]. While CHO cell lines are immortalized and capable of indefinite culture, the adaptability of CHO cell lines can lead to unintended phenotypic drift, referred to as cell line instability. For example, the most common biopharmaceutical products produced by CHO cells, monoclonal antibodies (mAbs), are metabolically challenging to produce. Exclusion of the transgene is a common mechanism to alleviate the cell's metabolic burden at the cost of losing productivity[32]. This loss of culture productivity is one of the barriers to continuous biomanufacturing[31]. The most common culture method in biomanufacturing is fed-batch cultures where the bioreactor is periodically supplemented with nutrients; however, these additions contribute to the accumulation of metabolic waste products such as ammonia and lactate, that impart a stressful environment on the cells which can lead to genome instability and culture termination[12,58,70].

The clonability of CHO cells is partially due to the inherent plasticity of the CHO genome[27]. This plasticity can lead to rearrangements and variant accumulation within critical regions of the genome, such as DNA repair mechanisms, that leads to genome instability[30,173]. Most recombinant CHO cell lines exhibit genome instability after a short

time due to the inherent plasticity of the CHO genome [2,21,31]. Genome instability can have multiple detrimental effects on cultures such as decreased productivity, poor product quality, and decreased cell viability [31,174]. Various engineering attempts to maintain genome stability have been explored such as site-directed transgene integration [41], promoter engineering [40,175], and waste product reduction [72,73] with varying levels of success: Site-directed integration allowed for more consistent generation of clones with transgenes inserted into stable safe harbors, modifications of the cytomegalovirus (CMV) promoter prevented reduction of productivity in some clones, and alternative feeding strategies, such as pH-mediated delivery of glucose, reduced the accumulation of lactate.

A poorly understood genomic entity that contributes to gene expression alterations, chromatin maintenance, and genetic heterogeneity that may also have a role in cell line instability and phenotypic drift in CHO cells is extrachromosomal circular DNA (eccDNA). EccDNA is a hallmark of genome plasticity[119,162,176-181] and has been identified within many eukaryotes such as yeast, plants, and drosophila[4,5,117,127,129]. In humans, eccDNA has been observed to contain amplified oncogenes and drug-resistant genes in cancers[126,138,150,151]; and in blood plasma[6,138,150,151]. The broad prevalence of eccDNA across kingdoms as well as in both diseased and normal tissue[6] likely indicate a conserved biological function. The eccDNA content in an organism seems to be dynamic and change as cells age in terms of abundance, size, sequence composition, and structural peculiarities[136,178,179,182]. These circularized, focal amplifications of small segmental chromosomal DNA look and function similar to episomes and constitute a rapidly accessible pool of genetic heterogeneity for the cell to utilize as the environment

changes[136]. EccDNA are often found in high copy numbers, which imparts ultra-high levels of gene expression [5,151,154]. Gene overexpression can serve as a rapid stress response mechanism[183], which can lead to genetic mosaicism and phenotypic drift[150]. Historically, eccDNA were first observed in CHO cells by Stanfield et al. in 1984 where they reported the presence of circular DNA with high homology to repetitive sequences[130]. Further sequencing studies confirmed eccDNA are partially composed of satellite DNA and show evidence of homologous recombination during biogenesis[7]; however, neither study identified genes encoded on eccDNA in CHO cells.

This study aims to characterize the sequence structure, function, and microevolutionary dynamics of eccDNA within a monoclonal antibody-producing CHO cell line grown in tightly controlled fed-batch cultures. Samples were collected at the beginning and end of cultures for sequencing. A lactate stress was added to duplicate bioreactors to understand the impact of culture stress on eccDNAs. EccDNAs were discovered and annotated for genes and structural features such as repeat motifs, transfer RNA (tRNA) content, and replication origins. The identified genes were mapped to the respective human orthologs for functional profiling in gene ontology (GO) and KEGG pathway analyses. Transcriptome data was also obtained and intersected with eccDNA data to identify potentially transcriptionally active eccDNA genes. Characterizing the dynamics of eccDNA content, or the circulome, in recombinant CHO cells under control and lactate-stressed conditions will improve our understanding of genome plasticity, cell line instability, stress response mechanisms, and implications in biopharmaceutical manufacturing.

3.3 Materials & Methods

3.3.1 Cell culture

Clone A11, a recombinant CHO-K1 cell line expressing an anti-HIV monoclonal antibody (VRC01) that was generated and donated by the NIH, was scaled up through four passages at three-day intervals post thaw from 1 mL working cell banks stored in liquid nitrogen. The inoculum train was expanded in 250 mL shake flasks with a 70 mL working volume and maintained at 37ºC, 5% CO_2, and 180 rpm. Bioreactors used in this study were ambr®250 vessels (Sartorius Stedim, Gottingen, Germany) with two pitched blade impellers and an open pipe sparger (vessel part number: 001-5G25). Bioreactors were inoculated with a target seeding density of $0.4x10^6$ cells/mL and a working volume of 210 mL in ActiPro media (Cytiva) supplemented with 6 mM of glutamine. Feeding began on Day 3 and followed a pyramid feeding scheme (3%/0.3% v/v Days 3-4, 4%/0.4% v/v Days 5-6, 5%/0.5% v/v Days 7-8, 4%/0.4% v/v Days 9-10, 3%/0.3% v/v Day 11 and beyond) with Cell Boost 7a/b (Cytiva), respectively.

Temperature and pH were controlled to 36.5ºC and 6.9 +/- 0.1, respectively. The pH was maintained using CO_2 and sodium bicarbonate; dissolved oxygen (DO) was maintained at 50%. The PID settings have been previously reported as the results of the third tuning in Harcum et. Al 2022[50]. A 10% antifoam solution (Cytiva) was added via a control loop as needed. To induce a lactate stress, a highly concentrated (1.338 M) sodium lactate solution was added at 12, 24, and 36 hours post-inoculation to duplicate cultures to increase the lactate concentration in 10 mM increments for a total 30 mM addition. Bioreactors were sampled daily for cell density (Vi-Cell, Beckman Coulter),

metabolite concentrations (Cedex Bio Analyzer, Roche), and to collect cell pellets for eccDNA and RNA analysis. Cell pellets were obtained by centrifuging culture broth at approximately 10,000x g for 10 minutes at 4°C, treated with RNAlater, and stored at -20°C until needed for nucleic acid extraction.

3.3.2 Library preparation

Cell pellets were split for RNA and gDNA extraction. Extractions were conducted with RNeasy midi kits (Qiagen, 74004) and DNeasy Blood and Tissue kits (Qiagen, 69504), respectively, per the manufacturer's instructions. Extracted RNA was quantified using a NanoDrop spectrophotometer and treated with DNase before sequencing. The gDNA was quantified using a Qubit Fluorometer (Thermo) prior to circular DNA enrichment. EccDNA was randomly amplified per the Circular DNA Enrichment sequencing (CIDER-Seq) protocol [10]. CIDER-Seq uses a Phi29 polymerase and exo-resistant random primers to randomly amplify circular DNA via rolling-circle amplification. This reaction was performed at room temperature over an 18-hour incubation. After amplification, circular sequences were debranched and the branches released and repaired to improve yield. EccDNA was then isolated using magnetic bead purification (KAPA Pure Beads, Roche, KK8000) prior to sequencing. SMRTbell barcodes were adapted to the samples by the sequencing vendor prior to sequencing using a PacBio Sequel II with HiFi reads.

3.3.3 Bioinformatic pipeline

The DeConcat algorithm was used to process the raw sequence data obtained from PacBio Sequel sequencing[10]. The confirmed eccDNA sequences for each replicate were compiled into a singular file per experimental condition and clustered to a 90% similarity threshold using CD-HIT[184,185]. The clustered sequences were then screened for repeat sequences using RepeatMasker v4.1.1 (Smit, AFA, Hubley, R & Green, P. *RepeatMasker Open-4.0*.2013-2015 <http://www.repeatmasker.org>) to characterize and mask repetitive motifs before annotating gene content using Maker[186]*.* The confirmed eccDNA sequences were mapped to respective chromosomal origins and intersected with 500 kbp genome windows to characterize biogenesis locations[187,188]. The content of tRNA was summarized using tRNAscan-SE 2.0[189]. Functional profiling of eccDNA genes was conducted using gene ontology (GO) and KEGG pathway analyses (detailed below)[190] with ClusterProfiler[191,192]. Furthermore, to annotate potential origins of replication, databases of known mammalian origins of replication and autonomous replication motifs were compiled from NCBI (retrieved on 07/14/2022, **Supplementary Table S3.1**) and used to BLAST search against a confirmed eccDNA sequence database[193]. The raw RNA-seq data were cleaned of sequencing adapters and low-quality bases with the Trimmomatic[194] software and quality checked with FastQC, respectively. Clean sequence data was aligned to the reference transcriptome using Bowtie2 read aligner[195], transcript abundance calculated using RSEM[196], and differentially expressed genes identified using edgeR[197] ($p \leq 0.001$, FDR ≤ 0.05). EccDNA and RNA sequences were called for variants against the Chinese hamster PICRH and CHO-K1 reference genomes using Varscan[198]. Transcripts containing SNPs that suggest

an eccDNA template were then visually analyzed using Integrative Genome Viewer (IGV)[199].

3.3.4 Gene function analysis and literature mining for eccDNA and genome instability-linked genes

Human eccDNA-relevant genes from literature were identified by Entrez Gene IDs from PubTator gene annotations[200]. PubMed Medline was queried with "extrachromosomal DNA" (ecDNA) and "extrachromosomal circular DNA" (eccDNA) to obtain available full-text article PMCIDs. The PMCIDs were used to retrieve the BioC xml files accessible from PubTator Central, which provides gene annotations in the full text articles [201]. The tool ezTag[202] was used to display the BioC xml files to allow for efficient manual curation of the gene entities (namely, to remove the non-relevant genes mentioned only in the reference sessions, and validate gene annotations provided by PubTator). The curated eccDNA-relevant human genes (N=431) were collected and hereinafter referred as eccDNA-relevant genes known from literature (**Supplementary Table S3.2**). Chinese hamster genes with human orthologs were identified based on NCBI ortholog assignment (ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/gene_orthologs.gz release of 02/23/2022, **Supplementary Table S3.3**). ClusterProfiler was used for GO and KEGG pathway enrichment analysis[192]. Human genes linked to genomic instability (N=2,897) from literature[12] were also used to intersect with the genes on CHO eccDNAs detected in this study.

3.4 Results

3.4.1 Phenotypic cell culture data

To characterize eccDNA dynamics within CHO, cells expressing VRC01 were cultured in an ambr®250 bioreactor system under fed-batch conditions in duplicate cultures with and without a lactate stress. Samples for eccDNA analysis were taken immediately after cells were inoculated into the bioreactor (Day 0), which resulted in four replicate samples (N=4). The duplicate control and duplicate lactate-stressed samples for eccDNA analysis were collected on Day 12 (Control Day 12, Lactate-stressed Day 12 respectively; N=2 each). The lactate stress resulted in lower growth rates, yet did not negatively impact cell viability (**Figure 3.1a**). Lactate-stressed cultures began lactate consumption on Day 2, while control cultures switched to lactate consumption on Day 4 (**Figure 3.1b**). Recombinant protein titers were lower for the lactate-stressed cultures (**Figure 3.1c**). The cell specific productivity was also lower for lactate-stressed cultures (**Figure 3.1d**).

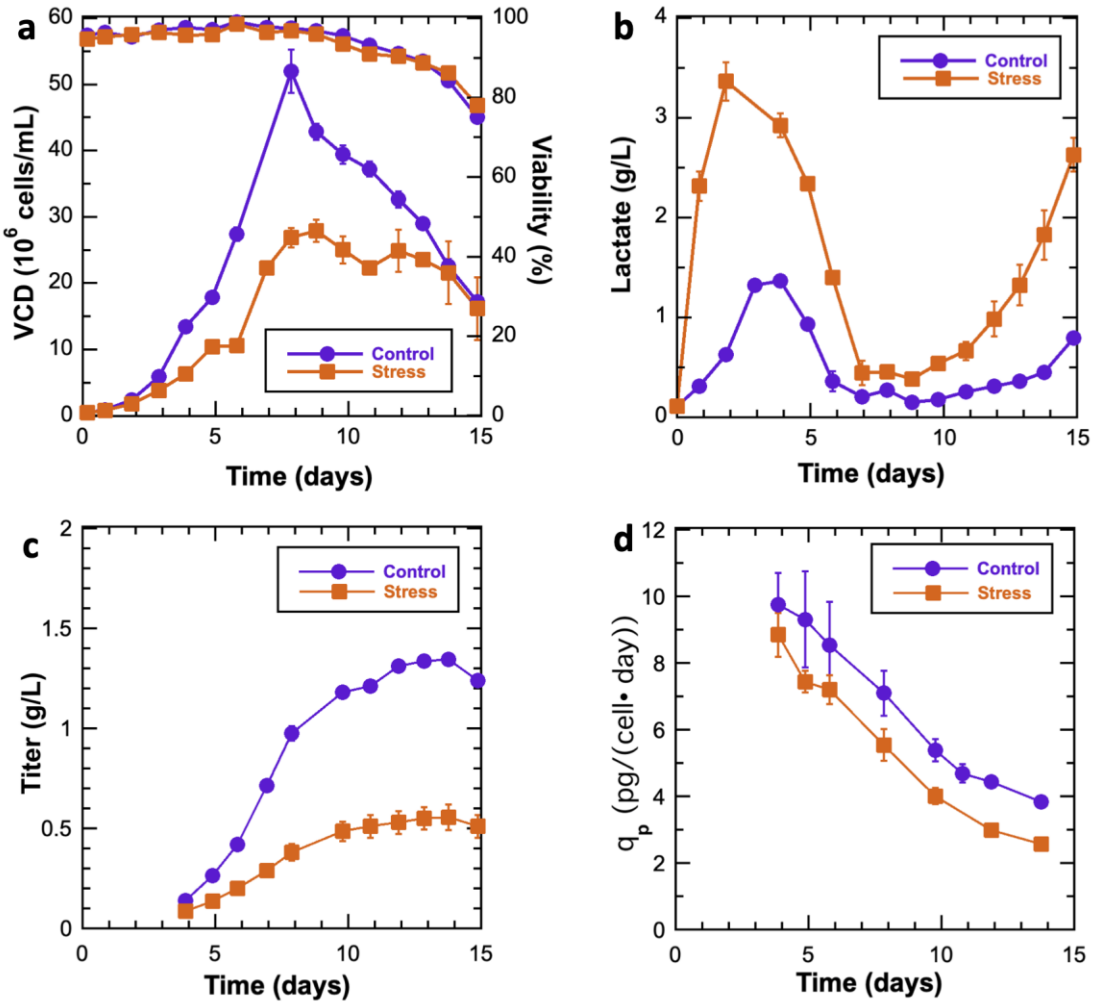


Figure 3.1: Growth characteristics for control and lactate-stressed CHO cell cultures. a) Viable cell density (VCD) (10^6 cells/mL) and cell viability (%), **b)** Lactate (g/L), **c)** Titer (g/L), and **d)** Cell-specific productivity (pg/cell•day). Control cultures (purple circles) and stressed cultures (orange squares) N=2. Error bars represent standard deviations.

3.4.2 Characterization of eccDNA sequence structure and gene content

EccDNAs were captured, sequenced, and verified following the CIDER-Seq pipeline (See Methods). The eccDNA predict algorithm identified 95,517 sequences across the three experimental conditions. Clustering of the eccDNAs at a similarity threshold of 90% collapsed similar sequences together to account for sequencing errors and short reads to a total of 76,317. Sequence length ranged from 21 bp to 24,309 bp. Mean sequence length was 4,063 bp in Day 0 samples and was a bit lower for the Day 12 samples, 3,579 bp and 3,534 bp for control and lactate-stressed Day 12 samples, respectively. Approximately 37% of bases were identified as repetitive motifs and masked. Long interspaced nuclear elements (LINEs) were the most abundant repeats identified in all three experimental groups followed by long terminal repeat (LTR) elements and short interspaced nuclear elements (SINEs). Distribution of repetitive motifs were mostly consistent across conditions, though lactate-stressed Day 12 samples had more LINEs (16.0%) than control Day 12 (14.3%) or Day 0 (13.3%) samples. tRNA motifs were predicted and found to be in relatively high abundance among the observed eccDNAs. For the Day 0 samples, there were 4,520 sequences (9.81%) that contained one or more tRNA motifs while the Control Day 12 and Lactate-stressed Day 12 samples had 1,182 (7.86%) and 1,151 (7.56%) tRNA motifs, respectively. The full tRNA annotation data can be found in **Supplementary Tables S3.4-3.6.** A database of known mammalian origins of replication was queried to identify sequences harboring motifs associated with autonomous replication. In the Day 0 samples, 4,639 sequences (10.1%) were observed to have an origin of replication motif with 95% or greater homology to a known mammalian origin of replication. For the Day

12 samples, only 134 (0.89%) and 12 (0.078%) origin of replication motifs were found in the Control and Lactate-stressed samples, respectively. The full origin of replication results can be found in **Supplementary Tables S3.7-3.9** and a detailed summary of sequence composition can be found in **Table 3.1.**

Table 3.1: Sequence characteristics for eccDNA sequences from control and lactate-stressed cultures including repetitive motif content, gene content, tRNA content, and potential origins of replication. Sequences were grouped by condition and clustered for similarity (>90%). The Day 0 samples include all four bioreactors, while Day 12 samples include duplicate bioreactors for control and lactate-stressed respectively.

Condition	Sequences	Sequences Clustered	Max Length (bp)	Average Length (bp)	Total Bases	Repeat bases masked	GC (%)	eccDNA with genes	eccDNA with tRNA	ORI (>95%)
Day 0	56,460	46,053	24,113	4,063	187,135,316 bp	67,898,424 bp (36.28 %)	40.85%	1,622 (3.52%)	4,520 (9.81%)	4,639
Control Day 12	19,354	15,047	24,309	3,579	53,852,662 bp	19,774,911 (36.72%)	40.92%	486 (3.23%)	1,182 (7.86%)	134
Stressed Day 12	19,703	15,217	21,482	3,534	53,776,705 bp	20,494,955 bp (38.11 %)	40.51%	451 (2.96%)	1,151 (7.56%)	12

		Day 0			Control Day 12			Lactate-stressed Day 12		
Repeat Structure	Subcategory	Number of Sequences	Base pairs (bp)	Percent of total bases	Number of Sequences	Base pairs (bp)	Percent of total bases	Number of Sequences	Base pairs (bp)	Percent of total bases
SINEs:		117,690	15,191,341	8.12	31,575	4,032,279	7.49	30,448	3,883,277	7.22
	Alu/B1	49,213	5,792,592	3.10	13,346	1,563,504	2.9	12,844	1,497,972	2.79
	MIRs	6,756	759,480	0.41	1,764	195,634	0.36	1,643	186,465	0.35
LINEs:		57,169	24,875,542	13.30	16,901	7,686,694	14.3	18,410	8,590,588	16
	LINE1	52,788	562,110	12.90	15,764	7,508,682	13.9	17,240	8,407,560	15.6
	LINE2	3,575	83,775	0.3	918	144,521	0.27	932	149,331	0.28
	L3/CR1	567	83,775	0.04	160	24,152	0.04	166	23,420	0.04
	RTE	213	34,675	0.02	54	8,805	0.02	63	9,279	0.02
LTR elements:		62,652	17,620,142	9.42	18,266	5,017,165	9.32	18,007	4,909,086	9.13
	ERVL	5,964	1,634,524	0.87	1,700	461,014	0.86	1,650	420,104	0.78
	ERVL-MaLRs	28,941	7,395,572	3.95	8,234	2,054,567	3.82	8,092	1,991,508	3.7
	ERR_class I	5,591	1,136,525	0.61	1,709	330,819	0.61	1,645	326,866	0.61
	ERV_class II	21,634	7,209,443	3.85	6,450	2,089,133	3.88	6,451	2,103,571	3.91
DNA elements:		11,960	2,185,216	1.17	3,351	588,742	1.09	3,250	567,968	1.06
	hAT-Charlie	7,181	1,238,630	0.66	2,060	340,327	0.63	2,002	329,155	0.61
	TcMar-Tigger	3,066	635,180	0.34	808	161,611	0.3	803	162,637	0.3
Unclassified:		2,445	921,187	0.49	684	255,392	0.47	768	322,187	0.6
Total Interspaced Repeats:		-	60,793,428	32.5	-	17.580,272	32.7	-	18,273,106	34
Small RNA:		3,405	267,780	0.14	1,071	87,090	0.16	929	71,975	0.13
Satellites		6,319	2,860,952	1.53	1,994	937,053	1.74	2,066	1,001,624	1.86
Simple Repeats		74,529	3,580,819	1.91	21,490	1,054,443	1.96	20,814	1,033,385	1.92
Low Complexity		9,021	456,155	0.24	2,547	130,081	0.24	2,518	130,232	0.24

Next, the 76,317 eccDNAs were analyzed for coding sequences (protein-coding genes). A total of 2,559 sequences (3.35%) were found to harbor one or more genes or gene fragments. All gene annotations are listed in **Supplementary Tables S3.10-3.12**. The distribution of gene content was observed to be relatively consistent across the three conditions (3.52% of Day 0, 3.23% of Control Day 12, and 2.96% of Lactate-stressed Day 12 sequences). These sequences contained 1,551 unique genes across the three conditions. The majority of these genes, 1,364 (88.0%), were only observed in one of the three conditions. However, 143 genes (9.21%) were observed in two conditions and 44 genes (2.83%) were observed in all three conditions. Interestingly, the gene content distribution was biased towards the Day 0 samples which included four bioreactors versus only two bioreactors for the control and lactate-stressed cultures, as the Day 0 samples were biologically identical at this timepoint. The distribution of genes across the conditions is summarized in **Figure 3.2.**

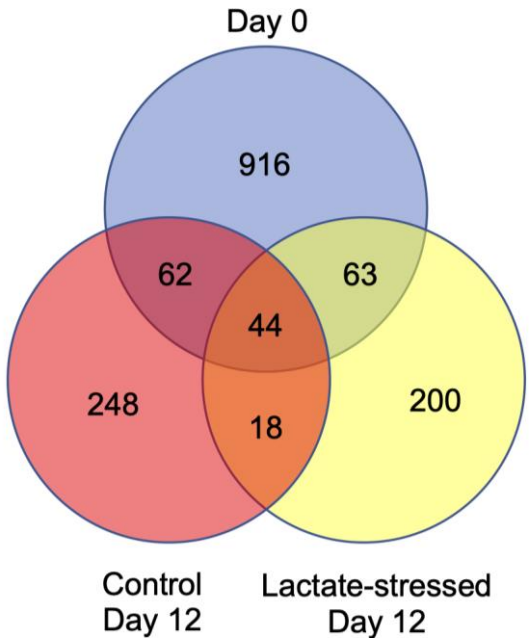


Figure 3.2: Venn diagram of the genes identified on eccDNA using CIDER-Seq. Day 0 is blue, Control Day 12 is red, and Lactate-stressed Day 12 is yellow.

3.4.3 EccDNA-encoded gene functional enrichment and text-mining analysis

A survey of biological function of the detected genes in CHO culture eccDNA sequences was performed using the enrichment analysis of GO hierarchy[203] and KEGG pathways. Gene lists for each culture condition were analyzed for enriched functions. Multiple GO biological process terms were found to pertain to translation, such as cytoplasmic translation, non-coding RNA (ncRNA) processing, and ribosome assembly. A network plot for the Day 0 GO biological process terms is shown in **Figure 3.3**; network diagram for GO terms observed in Lactate-stressed and Control samples on Day 12 are shown in **Supplementary Figs S3.1 and S3.2**. KEGG pathway analysis also showed significant enrichment in Ribosome and Coronavirus Disease COVID-19 pathways; Lactate-stressed Day 12 samples also had significant enrichment in the Folate biosynthesis pathway (**Supplementary Figure S3.3).**

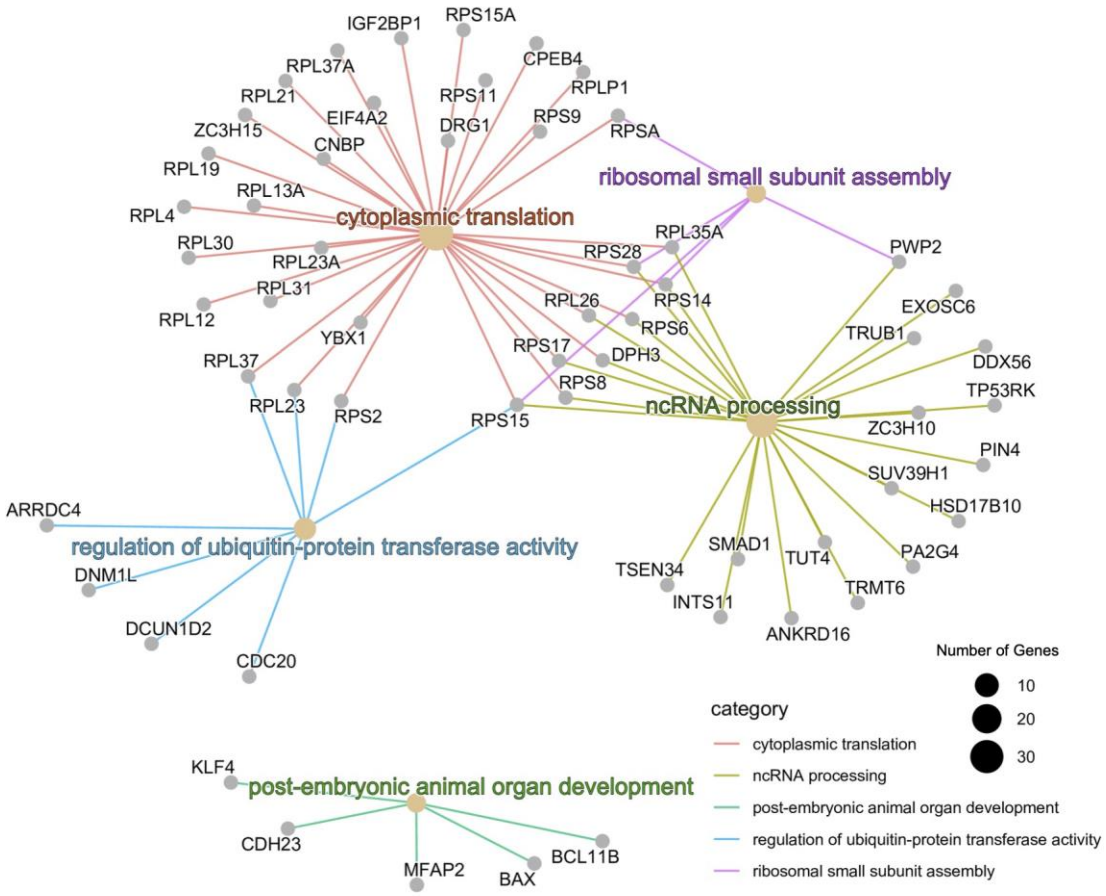
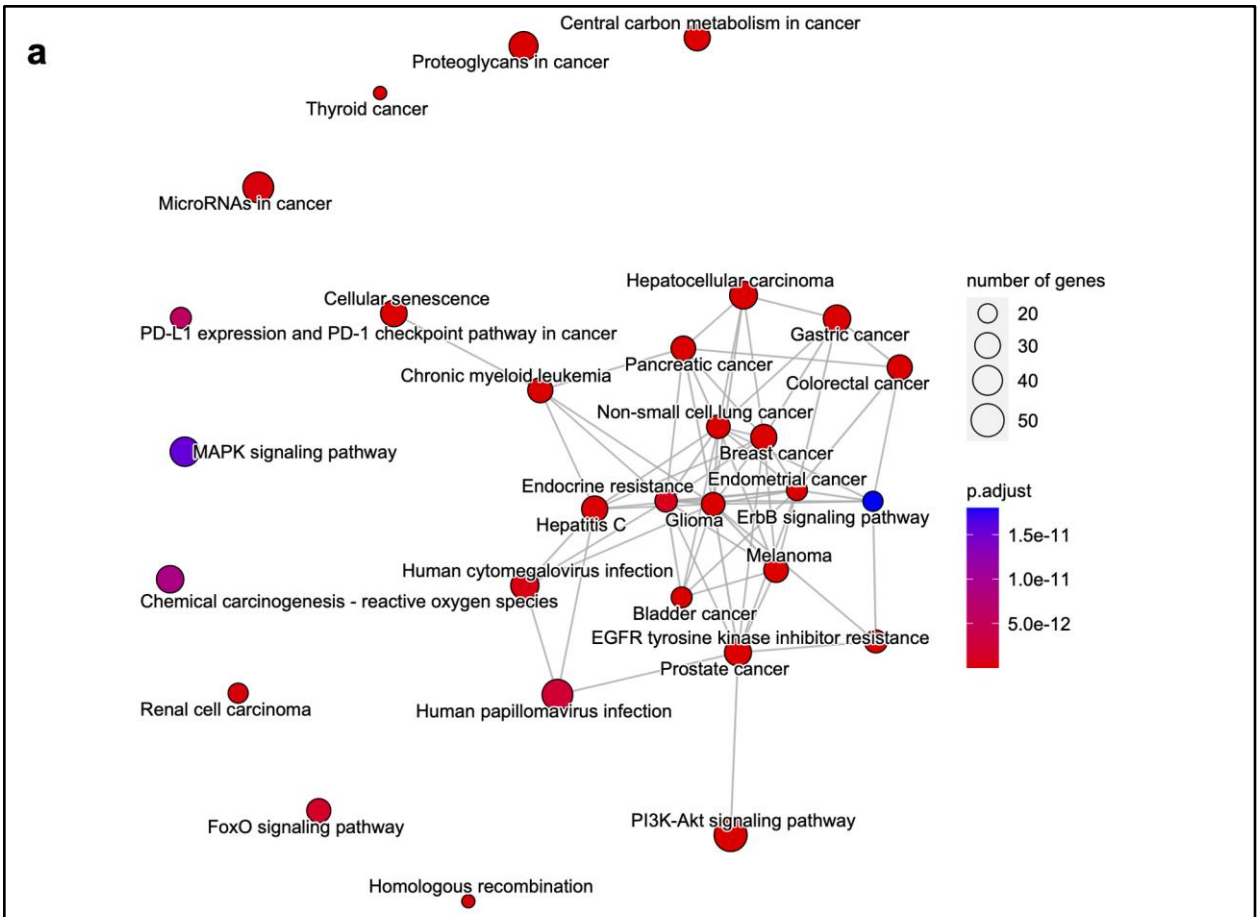


Figure 3.3: Network diagram of significantly enriched GO biological process terms (adjusted *p*-value ≤ 0.1) for the human orthologs of Chinese hamster genes detected in Day 0 samples. The small gray nodes show individual genes and larger beige nodes indicate GO terms. The size of the beige nodes is proportional to the number of genes with that GO term and the colored lines indicate the GO category for which a gene belongs to.

A total of 566 unique genes were manually curated from eccDNA-relevant literature available from PubTator Central (see Methods), which included 431 human genes. For these eccDNA-relevant human genes from literature, the enrichment analysis found 151 significantly enriched KEGG pathways (p-value ≤ 0.05) (**Supplementary Table S3.13**). Notably, many of the enriched pathways pertained to cancer **(Figure 3.4a)**. Multiple eccDNA-relevant genes associated with cancer progression in humans were also observed on eccDNA sequences in CHO cells (**Figure 3.4b**). Several cancer driver genes are amplified via eccDNA-mediated gene duplications in various tumor types[126]. Specifically, *Rac1* was observed in both the Lactate-stressed Day 12 and Day 0 samples; *Eef1a1* was observed in Day 0 and Control Day 12 samples; *Eif1ax*, *Gna11*, *Idh2* and *Ppp2r1a* were observed in Day 0 samples. Additionally, 2 genes were identified in all three conditions: *Gapdh* (glyceraldehyde-3-phosphate dehydrogenase) and *Dhfr* (dihydrofolate reductase). The presence of *Dhfr* on eccDNA sequences in CHO is notable as it is a common selectable marker gene used in CHO cell line development and was the selectable marker for the clone used in this study[35]. Further, CHO eccDNA genes were queried for relation to genome instability genes identified via literature mining, and 117 genes in the Day 0, 31 genes in the Control Day 12, and 29 genes in the Lactate-stressed Day 12 samples were identified (**Supplementary Table S3.14**). Functional profiling of genes associated with genome instability revealed significantly enriched GO biological processes involved in response to oxidative stress and toxic substances for the Day 0 and Lactate-stressed Day 12 samples (**Figure 3.4c)**. Notably, eccDNA gene ratios were higher in the lactate-stressed samples compared to the Day 0 samples, despite the

substantial different sizes of the gene lists. No significantly enriched terms were identified for the Control Day 12 samples (**Figure 3.4c).**



b

Samples	Number in Group	Gene ID	Gene Name
D0, C12, & ST12	2	2597	*Gapdh*
		1719	*Dhfr*
D0 & C12	2	23480	*Sec61g*
		1915	*Eef1a1*
D0 & ST12	1	5879	*Rac1*
D0	15	328	*Ape1*
		64072	*Cdh23*
		3312	*Hsc70, hspa8*
		1964	*Rif1ax*
		3418	*Idh2*
		2767	*Gna11*
		9582	*Apobec3b*
		10169	*Smn2*
		5518	*Ppp2r1a*
		11201	*Poli*
		6566	*Mct1*
		23435	*Tardbp, Tdp-43*
		3320	*Hsp90aa1*
		6510	*Asct2*
		5551	*Prf1*
		1676	*Icad*
C12	4	4507	*Mtap*
		79728	*Palb2*
		3014	*H2ax*
		10621	*Polr3f*

c

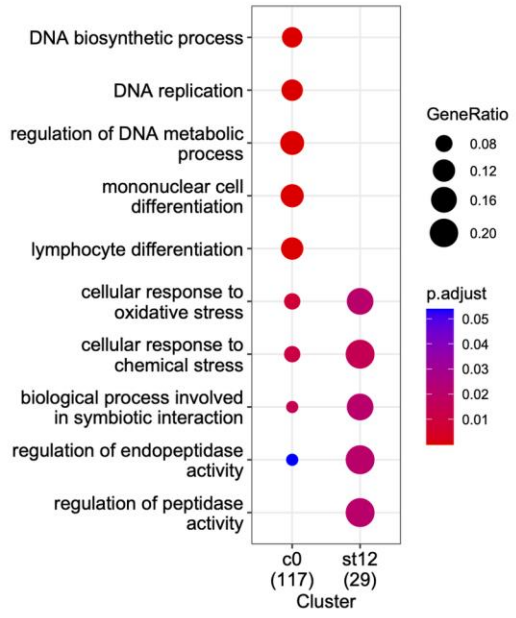
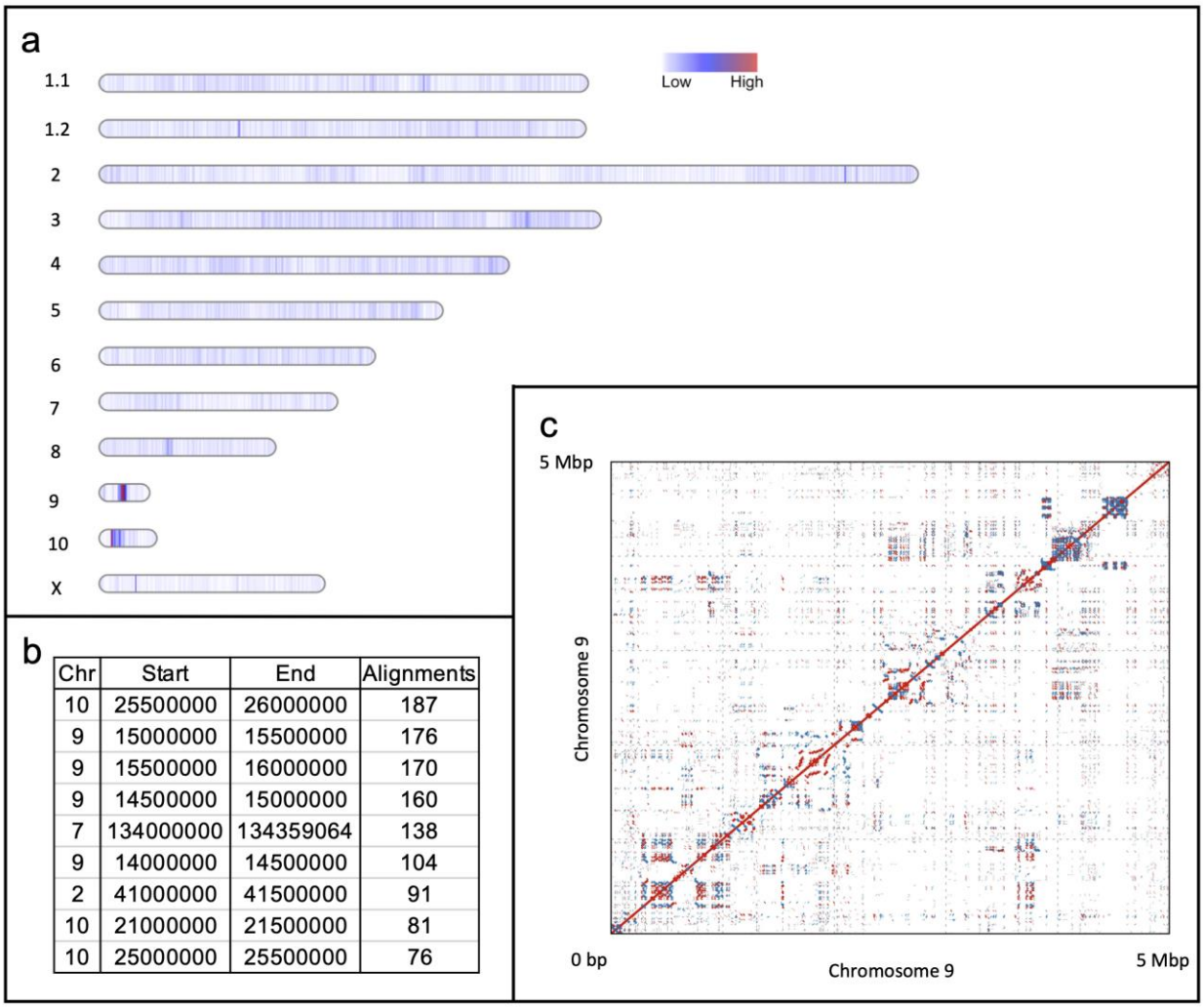


Figure 3.4: Summary of literature mining results for eccDNA-relevant genes and genome instability linked genes known from literature: KEGG pathway enrichment and CHO culture eccDNA genes. a) Top 30 of the 151 enriched KEGG pathways for the eccDNA-relevant genes texted mined from literature. The node size is proportional to the number of genes found in the pathway, while the node color indicates the pathway's statistical significance. Pathways with significantly overlapping genes are connected by grey lines. **b)** EccDNA-relevant genes identified via literature mining that were also found in CHO culture eccDNA sequences from one or more culture conditions: Day 0 (D0), Control Day 12 (C12), Lactate-stressed Day 12 (ST12). **c)** GO biological process enrichment analysis of genome instability linked genes observed on CHO cell eccDNA. Gene ratio reflects the fraction of genes pertaining to a GO term and the color indicates the statistical significance.

3.4.4 Characteristics of eccDNA biogenesis sites

There is very little knowledge regarding genomic origins of eccDNA as of this writing, thus, to gain a better understanding of this process, linearized eccDNA sequences were aligned to the CHO PICRH reference assembly, binned into 4,602 non-overlapping 500 kb windows, and counted to identify the genomic distribution of biogenesis sites and possible hotspot regions. If an eccDNA is mapped to two windows, it is assigned to the leftmost window, however this occurrence is highly infrequent. EccDNA biogenesis sites were identified throughout the genome with several regions that had higher eccDNA mapping rates, areas considered to be hotspots. Interestingly, only 22 genomic windows (500kb) out of the total 4,602 500 kbp windows (< 0.5% of the genome) had no eccDNA

mapped to these windows within the Day 0 samples (**Supplementary Table S3.15**). For the Day 0 samples, there were 44,402 unique eccDNAs that mapped to the reference genome (**Figure 3.5a**). The mean number of eccDNA mapped to a window was 9.64 (standard deviation of 8.02). To identify windows with the highest frequency of eccDNA mapping, windows were assigned Z-scores based on eccDNA mapping frequency; 58 windows had a Z-score $\geq$ 2 corresponding to 26 or more instances of mapping, and six windows had $\geq$ 100 mapped eccDNA sequences. One window on chromosome 10 had the highest number of unique alignments with 187 eccDNA sequences. Chromosome 9 had a region spanning ~2Mbp that harbored 610 eccDNAs (**Figure 3.5b**). This region on chromosome 9 also contained 30 unnamed genes, seven of which are described as chromatin target of *Prmt1* protein-like, five are zinc-finger proteins, and one related to a growth inhibitor protein. A self-alignment of this region on chromosome 9 revealed a repetitive structure with dispersed direct and inverted repeats, tandem repetitive arrays, palindromic sequences, and regions with the potential for intramolecular recombination (**Figure 3.5c**). Maps for the control Day 12 and lactate-stressed Day 12 biogenesis sites can be found in **Supplementary Figures S3.4 and S3.5** respectively, and summaries of biogenesis the frequencies can be found in **Supplementary Tables S3.16 and S3.17.**



Chr	Start	End	Alignments
10	25500000	26000000	187
9	15000000	15500000	176
9	15500000	16000000	170
9	14500000	15000000	160
7	134000000	134359064	138
9	14000000	14500000	104
2	41000000	41500000	91
10	21000000	21500000	81
10	25000000	25500000	76

Figure 3.5: Observed eccDNA biogenesis sites and characteristics. **a)** Chromosome-scale heatmap of eccDNA sequences observed for the Day 0 samples. Frequency of observed eccDNA is shown in color; low (white) to high (red). Figure made with minimap2 [188] and bedtools[187]. **b)** 500 kbp windows with the highest eccDNA mapping frequencies (Top 10 from Day 0 only). **c)** Self-alignment of chromosome 9 biogenesis hot spot. Red lines indicate a direct repeat; blue lines indicate an inverted repeat. Plot made using Mummer[204].

The biogenesis analysis of the Day 12 samples identified similar patterns to the Day 0 samples. Control Day 12 samples had 14,457 unique alignments while Lactate-stressed Day 12 samples had 11,028 unique alignments. The mean biogenesis frequencies for Day 12 conditions were 3.14 and 2.39 for Control and Lactate-stressed samples, respectively, with respective standard deviations of 3.25 and 2.4. Control Day 12 samples had 97 windows with Z-scores $\geq$ 2 while the Lactate-stressed Day 12 samples had 125 windows with a Z-score $\geq$ 2. Some variation in biogenesis frequency was observed between the three conditions; however, the 2 Mbp region on chromosome 9 and the 500kbp window on chromosome 10 were ranked in the top 10 biogenesis sites for all three conditions. Detailed Information on biogenesis frequency and Z-scores for all conditions can be found in **Supplementary Tables S3.15-S3.17.**

3.4.5 Identification of transcriptionally active eccDNA

To identify eccDNA genes that may be transcriptionally active, RNA-seq data was intersected with eccDNA data. Observed eccDNA genes that were unique to the Day 0 samples (916 genes) and genes unique to Day 12 samples (448 genes), were intersected with genes found to have a $\leq$ -2 or $\geq$ 2 $\log_2$ fold change respectively to correlate eccDNA gene loss or gain with corresponding transcriptome differences. Of the 916 genes only observed in Day 0 samples, 13 genes correlated with reduced transcript abundance. Of the 248 genes found only in Control Day 12 samples, 2 were correlated with increased expression; and for the 200 genes found in the Lactate-stressed Day 12 samples, 4 followed this pattern (**Figure 3.6)**. For context, 996 genes were found to be downregulated in one or both Day 12 conditions while 1,002 were found to be upregulated

in one or both Day 12 conditions. This implies that eccDNA had minimal impact in global gene expression shifts. The RNA-seq gene expression data for the 19 eccDNA genes is summarized in **Supplementary Table S3.18** and global gene expression shifts are shown in **Supplementary Table S3.19**

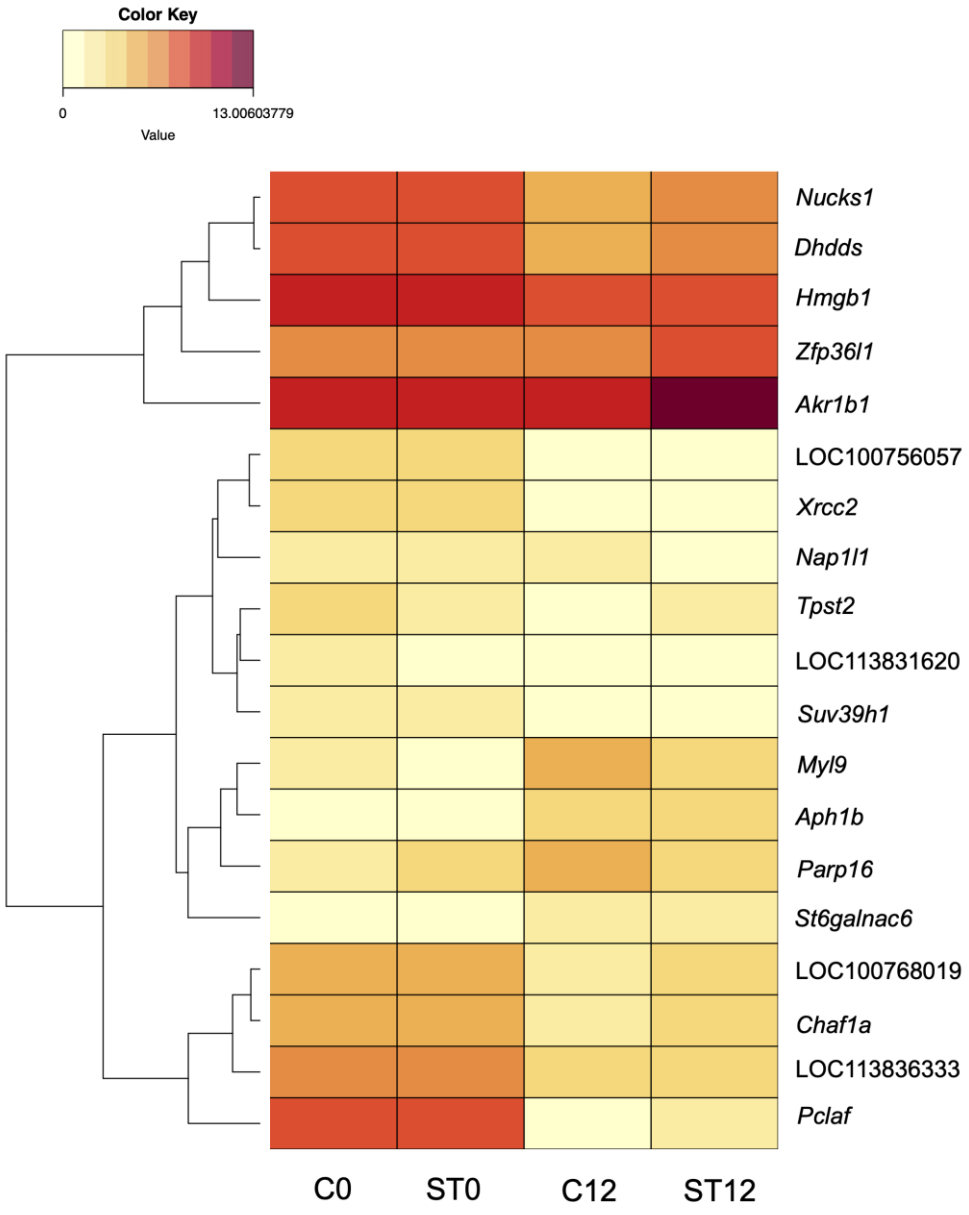


Figure 3.6: Heat map of differentially expressed genes from RNA-seq analysis that were only observed in one culture condition. Control is abbreviated as C and Lactate-stress is abbreviated as ST. Numbers in abbreviations are indicative of the Day samples were taken. Cells with light levels of shading indicate low levels of transcript abundance while darker cells correspond to higher levels of transcript abundance.

It is likely that genes encoded on eccDNA are under alternate selective pressure than chromosomal genes. To assign RNA transcripts directly to eccDNA, the RNA transcripts and eccDNA were called for variants against both the CHO-K1[78] and the Chinese hamster PICRH[79] reference genomes to identify single-nucleotide polymorphisms (SNPs) or insertions/deletions (INDELs) specific to eccDNA. For example, RNA transcripts containing a SNP relative to the CHO-K1 reference genome that is also within the consensus eccDNA sequence may have originated from the eccDNA template. Using this approach, homozygous SNPs were identified on an eccDNA and corresponding transcripts on chromosome 9 at base 14,641,267. For the CHO-K1 reference genome, 89% of the reads (63) at this location are adenine (A), with the remaining reads (8) being guanine (G). The consensus eccDNA and RNA transcripts at this locus are both guanine (G) in 100% of reads (**Figure 3.7)**.

Figure 3.7: **Example of transcripts with regions containing SNPs that may have originated from an eccDNA template on chromosome 9 at base 14,641,267.** Rows from top to bottom are: location of the sequence, coverage of the CHO-K1 reference assembly, the CHO-K1 reference genome, consensus eccDNA sequence for the Lactate-stressed Day 12 samples, coverage of the RNA data, RNA transcripts for the Lactate-stressed Day 12 cultures, and the Chinese hamster PICRH reference genome. The height of the gray bars in the coverage row reflects the amount of coverage for each nucleotide. Colored bars in the coverage row reflect SNPs relative to the reference genome. Nucleotides are shown as: adenine (A) - green, thymine (T) - red, guanine (G) - orange, and cytosine (C) - blue.

3.5 Discussion

Genome instability among CHO cell lines is a major contributor to declining productivity and product quality as cultures age; long-term cultured (LTC) cells have also been shown to have altered carbon metabolism due to genome instability [2,31]. The altered metabolism of LTC cells implies that genome instability has broader impacts beyond expression and glycosylation of recombinant protein products. Genome instability occurs through a variety of mechanisms such as chromatin condensation[205], DNA methylation[206], and variant accumulation[12]. Variant accumulation is often accelerated when DNA repair and or recombination mechanisms are compromised, which can be observed via biomarkers such as microsatellite instability[12]. Cell cultures are highly dynamic environments that constantly change in sometimes undesirable ways. Multiple factors, such as nutrient depletion, increased cell densities, and waste product accumulation, create stress within a culture that may elicit a stress response within the cells[25]. Cells have innate signaling and regulatory mechanisms that govern gene expression and, ultimately, the phenotype of the cell population. The cascades of these mechanisms that lead to adaptation have been understood to be encoded and maintained within the main chromatin body. Recent evidence has shown that eccDNA can harbor and express genes, influence gene expression of chromosomally encoded genes, and rapidly respond to cellular stress[5,8,125]. EccDNA are poorly understood in many systems and may function in the genetic coordination of traits in CHO under stress and homeostatic conditions.

Lactate was chosen as the stress in this study because it is a common waste metabolite; reducing or eliminating lactate is an intense area of research [25,61,62,72,73,75].

Elevated levels of lactate contribute to culture acidification; in controlled systems, such as the ambr®250, this can cause the system to add excessive base and/or increase the pCO_2 level, which can increase the osmotic pressure on the cells and retard growth [207,70,208]. Furthermore, lactate has also been shown to stunt cell growth and limit cell-specific productivity [57,69]. Previous work showed that lactate becomes detrimental at approximately 20 mM, and culture termination will occur in concentrations exceeding 40 mM[209]. The lactate stress was added incrementally in this study to avoid an overstressed environment while creating sufficient stress on the cultures. Preliminary shake flask experiments demonstrated that 10 mM doses of lactate allowed for better growth than a single 30 mM addition[210]. Further, as the osmolarity of the Control and Lactate-stressed cultures were not different, the slightly higher volume addition (< 2.2%) to supplement the lactate was considered an insignificant effector ($p \leq 0.05$). The reduced VCD and cell specific productivities of the stressed cultures show that a sufficient stress was achieved, while the tightly matched viability of all cultures demonstrates that the stressed cultures were not overwhelmed. Thus, the lactate-stress was the major effector, as well as culture time.

Capture and sequencing of eccDNA is a relatively new area of molecular biology that has been made more accessible by the rapid evolution of single-molecule sequencing technology such as Pacific Biosciences high fidelity (HiFi) reads. Circular DNA enrichment sequencing (CIDER-Seq) is an approach that circumvents the need for complex molecular protocols and computationally intensive analysis[10]. While the CIDER-Seq makes the identification of circular DNAs more robust, a caveat to the procedure is

that it is not quantitative due to the uncontrolled enrichment of eccDNA via rolling circle amplification (RCA). It is also critical to note that this technique is biased toward smaller sequences as these sequences are capable of much faster replication and hence accumulate more rapidly than larger sequences. CIDER-Seq is also limited to the read length offered by the sequencing instrument, but as sequencing technologies improve, it is anticipated that a parallel improvement in resolution of these elements will occur. It is possible that longer eccDNAs exist but were missed due to these current biases. Other methods that do not rely on Phi-29 amplification can be employed to estimate eccDNA abundance; however, these methods are much more costly and require a prohibitive amount of starting material. These cost and material requirements make characterizing the full distribution of discreet eccDNA size and abundance prohibitive for CHO cell culture experiments on this scale. Despite the limitations associated with the CIDER-Seq methodology, it still yields high quality data for sequences between approximately 20 and 25,000 bps.

Analysis of repetitive regions shows that the distribution of repeat structures within eccDNAs observed in this study were relatively equal across each condition. The similarity of repeat motifs across experimental groups suggests that biogenesis of eccDNAs due to repeat overlapping remains consistent in CHO cells when grown in both control and lactate-stressed conditions. The most identified repetitive element observed across all conditions was LINE1 (long interspaced nuclear element). In humans, LINE1 makes up approximately 17% of the genome[211]. While many LINE1s are transcriptionally silent in humans, some are capable of retro transcription, which can cause disruptions via

insertion, deletion, or rearrangement[148]. Another notable repeat motif, SINEs (short interspaced nuclear element), was also observed in relative abundance on eccDNA for all three conditions. SINEs are another type of retrotransposon that make up about 13% of the mammalian genome[212]. Structurally, SINEs have a conserved sequence structure as these transposable elements originate from tRNA sequences[149]. A relatively large portion of observed eccDNA in this study (8.98%) carried one or more tRNA motifs. EccDNA harboring tRNA has been previously described in Arabadopsis[120]. It is speculated that maintaining tRNA genes extrachromosomally may aid in stress response by facilitating rapid or high protein turnover required by a dynamic transcriptome load[160]. Other work has established that tRNA abundance is selectively modulated under stress conditions to regulate protein synthesis in yeast[161]. This could suggest an additional function of eccDNA within modulating protein production beyond gene expression.

When eccDNA sequences were annotated for genes, an average of ~3.35% were found to have one or more genes. Again, because the CIDER-Seq protocol is not quantitative, this is not indicative of the abundance of coding eccDNA, but rather a representation of eccDNA with predicted gene sequences. Most of the identified genes were only observed in one of the three conditions. This implies that gene content is highly dynamic across a 12-day fed-batch culture and between stress and control conditions. A functional enrichment analysis using gene ontology identified significantly enriched GO terms among each of the three culture conditions, many of which were linked to ribosomal assembly, cytoplasmic translation, and ncRNA processing. This enrichment could be influenced by the host cell line (CHO K-1) and/or the cell line development process. While

we cannot comment on eccDNA content of the CHO K-1 host, it is highly likely, if not certain, that eccDNA is present due to the ubiquitous nature of eccDNA in normal[4,127], disease[9,138], and stressed states[5].

Identifying multiple GO terms linked to translation could be attributed to selection of a clone with a high-producing phenotype during cell line selection as this is desirable for biomanufacturing. These eccDNA genes could be widely dispersed through cells in culture if present in the original clone; however, without selective pressure, these genes could be lost over time, which results in reduced productivity. Clones with high-producing phenotypes have been observed to lose the desired phenotype over time[213]. While there are multiple factors that could contribute to this, such as transgene exclusion and variant accumulation[2], eccDNA-mediated loss of productivity has not been studied in recombinant CHO cell lines. Further, the presence of *Dhfr* in all three conditions could be indicative of attempted transgene exclusion as *Dhfr* is the selectable marker used for the CHO K-1 cell line[35]. More ontology terms were significantly enriched in the Day 0 samples, however, this was likely due to the pooling of all four bioreactors for the Day 0 gene list as opposed to two bioreactors each for the Control and Lactate-stressed Day 12 gene lists. Yet, the Day 0 gene list had more than twice the number of genes compared to the Day 12 lists (1,622 genes in Day 0 versus 486 for Control Day 12 and 451 for Lactate-stressed Day 12). While the Lactate-stressed Day 12 and Control Day 12 had some variation in enriched GO biological process terms, none of the terms observed in the lactate-stressed Day 12 group were indicative of a stress response, but rather translation, ncRNA processing, and ribosome assembly; significantly enriched terms

observed in the Control Day 12 genes also pertained to protein production. Maintaining genes related to protein production on eccDNA likely aid the cells in facilitating protein turnover.

Biochemical pathway analysis of the observed genes that are associated with eccDNA in humans from the literature showed a significant enrichment in multiple cancer pathways. Linkages between eccDNA and cancer are clear as both eccDNA biogenesis, and cancer progression often rely on compromised DNA repair and or recombination mechanisms[126,127,150]. Additionally, some of these genes, such as *Poli* (error-prone polymerase involved in DNA repair), *Rac1* (cell growth regulator), and *Palb2* (tumor suppressor) were observed on eccDNA in CHO cells (**Figure 3.4b**). Overexpression of these genes could accelerate cell division, increase eccDNA biogenesis or recombination, and hasten the onset of cell line instability. Ontology of genome instability linked orthologs for the genes observed on eccDNA in CHO cells showed a notable increase in genes related to oxidative and toxic substance stress response in the Lactate-stressed Day 12 samples, which were not observed for the Control Day 12 samples. Furthermore, the fraction of cancer genes increased in the Lactate-stressed Day 12 samples compared to the Day 0 samples despite having a much smaller number of eccDNA genes observed.

EccDNA biogenesis has been shown to occur through multiple error-prone pathways such as non-homologous end joining (NHEJ), double-strand break repair, chromothripsis, and transcription; errors in DNA repair pathways are among the most

prominent biogenesis mechanisms[7,121,151,214]. Regions of the genome enriched with tandem repeats and other repetitive motifs have been observed to be more susceptible to eccDNA formation[122,142]. Due to the varied nature of eccDNA biogenesis mechanisms, eccDNA in humans appear to arise almost ubiquitously from the genome[8]. This allows some eccDNA to carry other functional sequences, such as autonomous replication sequences that enable gene copy number amplification and eccDNA permeation [5,118]. Centromeres have yet to be identified on eccDNA [9,138,150], thus eccDNAs typically display uneven segregation between daughter cells, which can increase population heterogeneity. When mapped to the genome, eccDNA biogenesis was found to occur globally throughout the genome; however, a 2 Mbp region near the center of chromosome 9 was found to have the highest frequency of biogenesis, likely due to the repetitive sequence structure of the region as shown in the chromosome self-alignment (**Figure 5c**). It was also observed that 3.35% of observed eccDNA contained one or more genes or gene fragments. This is an overrepresentation of genes when compared to humans, as less than 2% of the human genome coding genes[211]. EccDNA being biased toward coding regions of the genome supports previous work published by Hull, which correlated elevated levels of gene transcription with higher eccDNA abundance in yeast[121].

While genes may be amplified in eccDNA, additional regions, such as promoter and transcription factor binding sites, are required for transcription may be excluded or mutated. There is ambiguity when attempting to assign a transcript to an eccDNA or chromosomal template. The most direct approach to identifying eccDNA-derived transcripts is to leverage eccDNA-specific variants relative to the chromosome-encoded

gene. It can be assumed that an eccDNA encoded gene is under alternate selective pressures than chromosomal genes, hence accumulating variants at different rates. However, recent focal amplifications in the form of eccDNA may contain an exact copy of a nuclear gene, making it impossible to know which copy is functional. Mapping variant transcripts to the respective template is a straightforward way to identify sequence origin; however, this would not reflect transcripts from high-fidelity eccDNA that closely matches its genomic template. In addition to having proper transcription machinery, eccDNA need a replication origin to permeate through the population after selection. Yet, a single sequence does not need to have replication origins, promoters, and gene bodies upon biogenesis to permeate through the population, as recombination between eccDNA can allow for the accumulation of functional elements. While the timespan in this experiment was short (12 days), recombination events likely occurred in these sequences, but would be more prominent in longer cell cultures, such as perfusion.

Of the thirteen genes only identified in Day 0 eccDNAs that were found to be downregulated by Day 12, one is involved in maintaining genome stability (*Nap1l1*), three facilitate DNA repair (*Nucks1, Pclaf,* and *Xrcc2*) and three maintain or regulate chromatin structure (*Suv39h1, Chaf1a,* and *Nap1l1).* Two genes only observed on the Control Day 12 eccDNA were observed to have increased transcription, while four genes were observed on the Lactate-stressed Day 12 eccDNAs that were upregulated. The two upregulated genes observed on the Control Day 12 eccDNAs included two signal transduction genes (*Myl9, Parp16*). The four upregulated genes on the Lactate-stressed Day 12 eccDNAs include *St6galnac6,* a cell surface receptor gene*, Zfp36l1,* a zinc finger

protein*, Aph1b,* a transmembrane protein that is part of the gamma-secretase complex[215], and *Akr1b1*, an aldo-keto reductase that catalyzes NADPH-reduction of carbonyl-compounds into alcohols[215]. Overexpression of *Akr1b1* has been observed in multiple cancer types and is thought to increase Warburg effects by triggering the AKT/mTOR signaling pathway[216]. *Akr1b1* overexpression in the Lactate-stressed Day 12 cultures could have contributed to the elevated lactate production and subsequent accumulation observed starting on Day 8.

3.6 Conclusion

This work has demonstrated that the eccDNA gene content within CHO cells is highly dynamic even across the relatively short time span of a fed-batch culture. While tightly controlled bioreactor systems, such as the ambr®250, are lauded for the tight level of culture control, internal genetic elements can still drive heterogeneity that leads to phenotypic drift. These issues may be difficult to address with process engineering and likely present a new challenge in cell line development efforts to curb genetic heterogeneity. EccDNA in CHO cells may bias the clone selection process by harboring beneficial genes for protein expression, modification, and secretion; yet, during production, conditions may allow phenotypic drift through amplification of genes responsible for cancer phenotypes or loss of beneficial genes that are unevenly segregated or lost in cell division. Furthermore, this work highlights the importance of eccDNA microevolution due to environmental disturbances, such as waste metabolite accumulation. Thus, eccDNA may be considered as new targets for CHO cell line improvement and genetic process control.

CHAPTER FOUR

COMPARISON OF ECCDNA EVOLUTION, GENE EXPRESSION CHANGES, AND GENOME STRUCTURE VARIATION BETWEEN TWO REFERENCE CHINESE HAMSTER OVARY CELL LINES IN YOUNG AND AGED CELLS

4.1 Abstract

The utilization of Chinese hamster ovary (CHO) cell lines in biopharmaceutical manufacturing is widespread due to ease of use, non-susceptibility to viruses, and ability to perform human-like post-translational modifications to proteins. Yet, the intrinsic instability of the CHO genome remains a major obstacle to continuous biomanufacturing. Genome instability can lead to transgene exclusion, incorrect DNA repair pathways, and the disruption of essential biological processes. To address these challenges, two reference CHO cell lines, VRC01 and CHOZN GS23, have recently been publicly characterized in fed-batch cultures in an effort to advance our understanding of CHO genome instability and CHO-based bioprocessing innovation. Despite these efforts, the genetic underpinnings that drive CHO cell genome instability with age remain unclear. Extrachromosomal circular DNAs (eccDNAs) from the two reference CHO cell lines were sequenced. This provided insight to eccDNA evolution in long-term cell cultures and heterogeneity that can arise from eccDNAs in CHO cell lines from the same lineage. This work presents robust annotation of sequence structures, such as repeats, genes, rRNA templates, origins of replication, as well as corresponding transcriptome data. Additionally, the impact of the CHO genome architecture on eccDNA biogenesis was

investigated. These results identified 583 unique genes within eccDNAs across both cell lines, with 141 newly identified genes differentially expressed in aged CHO cells. These findings contribute to a greater understanding of factors that contribute to long-term CHO cell culture stability and have the potential to inform future innovations in the field of biopharmaceutical manufacturing.

4.2 Introduction

The origin of all Chinese hamster ovary (CHO) cell lines can be traced to tissue isolated in the 1950s[13]. Over time, multiple cell lineages have been derived from the original tissue through a combination of induced mutagenesis and/or spontaneous adaptation, including the common lineages CHO-S, CHO-DG44, and CHO K-1[2,15]. The initial interest in CHO cells arose due to the observable genomic rearrangements[14]. Further, CHO cell lines have become the preferred expression hosts for monoclonal antibody manufacturing due to ease of cloning[1]. However, the intrinsic plasticity of the CHO genome leads to genomic rearrangements[217] and transgene exclusion or silencing[37,218], which can diminish cell health and productivity[2]. These phenomena collectively decrease cell line stability[219] and result in high variability in genetic and transcriptomic profiles in different batches of the same cell line, which has adverse outcomes in the context of biopharmaceutical manufacturing[220]. Cell line instability is a major limiting factor to the implementation of continuous biomanufacturing, as cell-specific productivity tends to decline with age; typically becoming uneconomical after approximately 65-70 generations or population doubling levels (PDLs)[2]. Multi ‘omics studies have shown that this decline is partially attributed to the downregulation of cell

cycle and DNA repair genes[221]. Thus, a greater understanding of CHO cell genome instability is desired.

Studies in recent years have documented the presence of extrachromosomal circular DNAs (eccDNAs) in oncogenic and healthy human tissue[6,126,222,223], as well as in organisms such as yeast[4,121], drosophila[122], and pigeons[224] in both healthy and disease-state tissue[6]. While eccDNAs were first described decades ago[117], these DNA structures have emergined as a critical biomarker for more aggressive cancers[3,162]. These plasmid-like sequences arise from an organism's genome through a variety of biogenesis pathways[8,136] and are characteristics of plastic and/or unstable genomes[3]. EccDNAs are highly heterogeneous in sequence structure and composition[115] and can serve as a mechanism for gene overexpression. Genes encoded on eccDNAs have been observed to confer selective advantages, such as oncogene amplification in cancer[222] or pesticide resistance in pigweed[5]. The persistence of eccDNA can also provide a sustained evolutionary advantage to the cell[8].

The accumulation of cells possessing advantageous eccDNAs can lead to gradual domination and result in genetic mosaicism within an organism[225]. Beyond gene overexpression, eccDNAs have been implicated in aging through a variety of mechanisms. Circularized telomere fragments, known as t-circles, have been observed and are thought to contribute to telomere maintenence[226,227]. Additionally, a significant role for eccDNAs in yeast aging has been demonstrated; while the content of eccDNAs in young yeast is diverse, it becomes homogenized with over 95% harboring ribosomal

DNA (rDNA) as the yeast ages[228,229]. Previous studies have characterized the dynamic nature of eccDNA content in a CHO K-1 cell line where it was found to vary greatly across a 12-day fed-batch culture[11]. An increased abundance of eccDNAs has been found to be positively correlated with rising genome instability[150], and may contribute to the decline in productivity and product quality as cultures age. In contexts where maintaining homogeneity is critical, such as biopharmaceutical manufacturing, understanding the evolution of eccDNAs throughout the duration of cell culture is imperative to fully comprehending the expression host.

Recently, two industry-relevant reference CHO cell lines, with respective medias and process platforms, have been characterized and made publicly available to accelerate bioprocess innovation[230]. These cell lines, VRC01 (NIH) and CHOZN GS23 (Millipore Sigma), both express monoclonal antibodies and have notably different phenotypes; VRC01 grows to higher viable cell densities (VCDs), but produces less recombinant protein per cell, while GS23 grows to much lower VCDs, but produces higher quantities of recombinant protein per cell[230]. Data for these reference cell lines has been published for industry-standard fed-batch cultures of approximately two weeks in several publications[11,50,58,210,230]. In this study, VRC01 and CHOZN GS23 cell lines were both characterized after long-term passages. Samples capturing early, middle, and late population doubling levels (PDLs) were examined. The structure and dynamics of eccDNAs were assessed . The effect of genome structure on eccDNA biogenesis was investigated as well as correlated with transcriptional changes via RNA-seq. These tools

were used to identify changes in eccDNA-mediated gene expression across both cell lines with age.

4.3 Materials & Methods

4.3.1 Cell Generation

A CHO K-1 derived cell line expressing the VRC01 monoclonal antibody (Clone A11) was donated by the National Institute of Health (NIH) and will be noted as "VRC01" throughout the text. The second cell line, noted as "CHOZN", was generated using the CHOZN® platform and was provided by MilliPoreSignma (reference: https://www.sigmaaldrich.cn/deepweb/assets/sigmaaldrich/marketing/global/documents/855/568/chozn-ucoe-white-paper-ms.pdf). Like VRC01, CHOZN is also a CHO K-1-derived cell line. Both cell lines were grown in commercially available media as recommended by the manufacturer; VRC01 in ActiPro (Cytiva) supplemented with 6 mM glutamine and CHOZN in EX-CELL Advanced CHO Fed-batch Media (MilliporeSigma) or EX-CELL CD CHO Fusion (MilliporeSigma).

Vials stored in liquid nitrogen were thawed and continuously passaged to generate cells at various ages. Cells were cultured at 37^{o}C with 5% CO_2 and 80% relative humidity in a shaking incubator (InforsHT) with a 25 mm throw diameter[210]. Cultures were maintained in 125 mL shake flasks with a working volume of 20-30 mL. The cultures were inoculated with a targeted seeding density of 0.4 x 10^6 cells/mL. Cell count and viability were measured using a Vi-Cell XR (Beckman Coulter). Cumulative PDL was calculated using initial and final VCD at each passage[231]. For the VRC01 cell line, cell banks were

generated at PDL 10, 24, 70 and 110 while the CHOZN cell banks were generated at PDL 20, 35, 60 and 90. The VRC01 cells were cultured in ActiPro (Cytiva) and the CHOZN cells were cultured in EX-CELL CD CHO fusion media (MilliporeSigma). Banked vials contained 1 x 10^7 cells in 1 mL media with 10% dimethyl sulfoxide (DMSO) (Sigma). Cell banks were stored in liquid nitrogen. For sequencing analysis, cells were thawed and expanded for 3 passages. For CHOZN, EX-CELL Advanced CHO Fed-batch media was used while VRC01 was expanded in ActiPro media (Cytiva). Between 2.5 and 4 million cells in duplicate 1 mL samples were collected for sequencing. Cell pellets were obtained by centrifugation at 150 g for 3 minutes. The cell pellets were resuspended in 200 µL RNA later and stored at -20^{o}C until sequencing preparation.

4.3.2 Library preparation and sequencing

Genomic DNA (gDNA) was extracted from cell pellets using a DNEasy kit (Qiagen, 69504) per the manufacturer's recommended protocol. The gDNA was measured with a Qubit fluorometer 2.0 (Invitrogen, Q32866) before circular DNA enrichment. EccDNAs were amplified and purified per the CIDER-seq protocol[10] and quantified with the Qubit fluorometer before sequencing. Samples were sequenced on a PacBio Sequel II for single molecule real time (SMRT) sequencing; barcode adapters were added to the sample by the sequencing vendor. RNA was extracted from cell pellets using an RNEasy midi kit (Qiagen, 74004) per the manufacturer's recommended protocol. Samples were treated with DNAse to remove gDNA contamination. Total RNA was analyzed for quality and integrity via UV spectroscopy (Nanodrop8000 ThermoFisher Scientific) and Bioanalyzer 2100 (Agilent), respectively. All samples had a minimum RNA integrity score

of 7.0. Total RNA was quantified using a double-stranded dye binding assay on the Qubit (ThermoFisher Scientific). Library preparation was conducted with the NEBNext Ultra II RNA Library Prep Kit for Illumina following the manufacturer's recommended procedures and pooled in equimolar ratios for sequencing. Paired-end reads for each sample (2x150bp) were collected on an Illumina NovaSeq 6000 S4 flow cell to an approximate depth of 20 million read pairs.

4.3.3 Bioinformatic pipeline

4.3.3.1 Raw data

Raw eccDNA data was initially processed using the CIDER-seq DeConcat algorithm[10] to identify sequences with a confirmed genomic origin. Experimental duplicates were combined and sequences were clustered using CD-Hit[185] to a 90% identity threshold. Trimmomatic[194] was used to remove sequencing adapters from raw RNA data before quality checking reads with FastQC. Bowtie2[195] aligned clean, high-quality reads to the reference transcriptome. RSEM[196] was used to calculate transcript abundance and edgeR[197] calculated differential expression.

4.3.3.2 EccDNA annotation

Repeat and low-complexity regions on eccDNAs were summarized and masked using RepeatMasker (Smit, AFA, Hubley, R & Green, P. *RepeatMasker Open-4.0*. 2013-2015 <http://www.repeatmasker.org>.) before annotating predicted genes using *Maker*[186]. Transfer RNA (tRNA) templates were annotated with tRNAscan-SE 2.0[189]. BLAST[193] was used to identify origins of replication (ORIs) and rDNA from custom databases. Known mammalian origins of replication and autonomous replication motifs

(N=118), (retrieved on 07/14/2022), and rDNA from human (N=804), mouse (N=81), rat (N=153), and Chinese hamster (N=25) (retrieved on 01/09/2023) were compiled from the National Center for Biotechnology Information (NCBI). BLAST searches were conducted with an e-value of 1e-50. Databases for ORI and rDNA mapping can be found in **Supplementary Tables S4.1 and S4.2,** respectively.

4.3.3.3 EccDNA genomic origins

Genomic eccDNA biogenesis sites were identified by dividing the chromosomes into 500 kbp windows generated using BEDTools[187]. EccDNAs were queried using BLAST[193] against the most recent Chinese hamster genome construction, PICRH[79], and intersected with the 500 kbp genome windows to count eccDNAs per window. In the unlikely event an eccDNA overlapped two windows, it was assigned to the leftmost window. BLAST parameters are the same as described for the ORI and rDNA annotations. Visualization of eccDNA biogenesis was conducted with minimap2[188]. Genome windows were assigned Z-scores to identify statistically significant regions with high biogenesis frequencies. To better understand the genomic features of eccDNA-encoded genes, the number of genes and repeats were quantified within the 200 kbp windows flanking each gene listed in the PICRH genome features file. Repeat content of the genome was annotated with *RepeatMasker*. Genes and repeats within each window were counted using BEDTools; A permutation test ($p \leq 0.05$, N=10,000 replicates) was conducted in JMP Pro 17 and compared the mean numbers of genes and repeats between the genes observed on eccDNAs and genes not observed on an eccDNA.

4.3.3.4 Gene functional enrichment

Chinese hamster genes with human orthologs were identified based on NCBI ortholog assignment (ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/gene_orthologs.gz release of 02/23/2022, **Supplementary Table S4.3**). ClusterProfiler was used for GO and KEGG pathway enrichment analysis[192].

4.4 Results

4.4.1 Long-term Cell Culture Phenotype

Cell line instability in bioprocessing typically refers to a cell line maintaining 70% or more of its maximum observed cell-specific productivity (q_p, pg/(cell•day))[2]. Cell-specific productivity can be negatively impacted by several environmental stressors, one of which is cell aging. For VRC01, the maximum cell-specific productivity is typically 7.73 pg/(cell•day) and has been observed to become unstable around PDL 24[230]. The maximum cell-specific productivity for CHOZN is typically 43.7 pg/(cell•day)[230] and has been observed to be unstable between PDLs 60-70 (**Figure 4.1**).

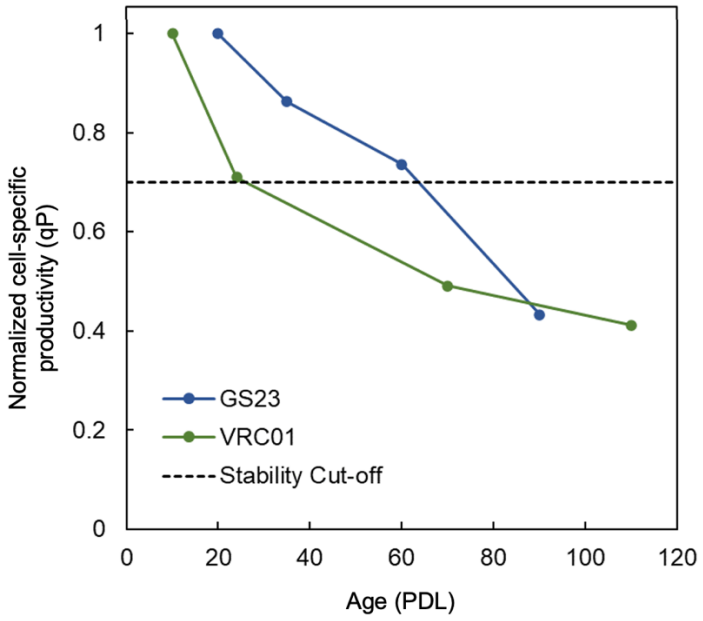


Figure 4.1: Normalized cell-specific productivity of VRC01 and CHOZN cell lines at varying PDLs. VRC01 (green), CHOZN GS23 (blue). The dashed line at q_p of 0.7 represents the stability cutoff.

4.4.2 EccDNA structure and origins

EccDNAs with lengths greater than 20 kbp were detected in both cell lines across all PDLs examined. The average observed eccDNA length of was greater for the VRC01 samples compared to CHOZN samples, with an average length of 5,400 bp and 4,500 bp, respectively. The GC content for the CHOZN samples (41.2%) was consistently higher compared to VRC01 samples (39.6%). The frequency of tRNA-harboring eccDNAs was showed small variation between cell lines and across all PDLs examined, (approximately 8.5% in VRC01 and approximately 7% in CHOZN). Full annotations of VRC01 and CHOZN eccDNA-encoded tRNAs are shown in **Supplemental Tables S4.4** and **S4.5**, respectively**.** The gene annotation contained some variation, with a higher percentage of genes for the VRC01 PDL 110 samples compared to the younger VRC01

PDL samples (5.6% and 4.9%, respectively). The gene content for CHOZN samples contained a variable content across the PDLs that increased from PDL 20 to PDL 35, (4.4% to 5.3%) then decreased from PDL 35 to PDL 90 (5.3% to 3.1%). Complete Maker annotations for VRC01 and CHOZN across all PDLs are available in **Supplemental Tables S5.6** and **S5.7**, respectively. A small number of eccDNAs with regions consistent with known mammalian origins of replication were observed for both cell lines across the PDLs. BLAST output for ORI analysis for VRC01 and CHOZN eccDNAs are provided in **Supplemental Tables S5.8** and **S5.9**, respectively. Genes encoding rRNA were also detected for both cell lines across all PDLs. Output of rDNA BLAST for VRC01 and CHOZN eccDNAs can be found in **Supplemental Tables S5.10** and **S5.11.** The distribution of repetitive elements was consistent for both cell lines across all PDLs. CHOZN had a lower proportion of LINEs compared to VRC01 (14.0% and 16.3%, respectively). Although subtle differences in structure of eccDNAs were observed, the distribution of fundamental sequence structures were relatively consistent in both cell lines across all PDLs. Detailed summaries of the sequence structures for VRC01 and CHOZN eccDNAs are shown in **Tables 4.1** and **4.2**, respectively.

Table 4.1: Sequence characteristics of VRC01 eccDNA sequences. Repetitive motif content, gene content, tRNA content, rDNA content, and potential origins of replication are included. Sequences were grouped by PDL and clustered for similarity (>90%).

Condition	Sequences	Sequences Clustered	Max Length (bp)	Average Length (bp)	Total Bases	Repeat bases masked	GC (%)	eccDNA with genes	eccDNA with tRNA	ORI (>95%)
PDL 10	5,205	5,140	22,800	5,291	28,066,199 bp	10,884,152 (38.78%)	39.84%	255 (4.9%)	443 (8.62%)	5
PDL 70	5,725	5,668	23,215	5,464	30,955,674 bp	12,140,094 (39.22%)	39.22%	280 (4.94%)	530 (9.35%)	10
PDL 110	5,981	5,913	26,554	5,552	32,674,191	12,688,431 (38.83%)	39.87%	333 (5.63%)	455 (7.69%)	4

		PDL 10			PDL 70			PDL 110		
Repeat Structure	Subcategory	Number of Sequences	Base pairs (bp)	Percent of total bases	Number of Sequences	Base pairs (bp)	Percent of total bases	Number of Sequences	Base pairs (bp)	Percent of total bases
SINEs:		14,417	1,877,201	6.69	15,649	2,041,591	6.6	16.159	2,106,219	6.45
	Alu/B1	6,251	742,057	2.64	6,785	810,326	2.62	7,119	852,516	2.61
	MIRs	783	90,537	0.32	808	90,639	0.29	912	104,860	0.32
LINEs:		9,041	4,545,475	16.20	9,998	5,143,304	16.62	10,604	5,336,770	16.33
	LINE1	8,506	4,453,935	15.87	9,426	5,051,161	16.32	10,024	5,243,856	16.05
	LINE2	432	74,591	0.27	450	71,201	0.23	478	76,964	0.24
	L3/CR1	68	11,111	0.04	87	12,875	0.04	76	11,063	0.03
	RTE	29	4,784	0.02	31	6,908	0.02	19	3,472	0.01
LTR elements:		8,817	2,744,323	9.78	9,752	3,023,901	9.77	10,448	3,251,924	9.95
	ERVL	771	215,630	0.77	866	240,648	0.78	924	262,899	0.8
	ERVL-MaLRs	3,639	977,465	3.48	4,139	1,116,795	3.61	4,271	1,142,468	3.5
	ERR_class I	849	178,153	0.63	862	171,746	0.55	962	210,732	0.64
	ERV_class II	3,489	1,329,033	4.74	3,822	1,458,574	4.71	4,187	1,566,010	4.79
DNA elements:		1,516	269,741	0.96	1,683	324,824	1.05	1,717	325,543	1
	hAT-Charlie	936	154,678	0.55	996	181,265	0.59	1,043	187,409	1.57
	TcMar-Tigger	363	75,340	0.27	443	97,026	0.31	441	93,564	0.29
Unclassified:		425	184,773	0.66	494	209,995	0.68	510	109,048	0.64
Total Interspaced Repeats:		-	9,621,513	34.28	-	10,743,615	34.71	-	11,229,504	34.37
Small RNA:		492	35,556	0.13	584	46,400	0.15	578	44,669	0.14
Satellites		1,231	573,735	2.04	1,413	671,933	2.17	1,461	668,980	2.05
Simple Repeats		11,384	591,133	2.11	12,297	613,184	1.98	12,845	675,865	2.07
Low Complexity		1,357	68,817	0.25	1,463	74,650	0.24	1,555	80,041	0.24

Table 4.2: Sequence characteristics of CHOZN eccDNA sequences. Repetitive motif content, gene content, tRNA content, rDNA content, and potential origins of replication are included. Sequences were grouped by PDL and clustered for similarity (>90%).

Condition	Sequences	Sequences Clustered	Max Length (bp)	Average Length (bp)	Total Bases	Repeat bases masked	GC (%)	eccDNA with genes	eccDNA with tRNA	ORI (>95%)
PDL 20	6,183	6,083	24,895	4,529	27,631,900 bp	10,538,217 (38.14%)	41.17%	266 (4.37%)	404 (6.64%)	8
PDL 35	3,659	3,575	22,890	4,730	16,973,623 bp	6,502,617 (38.31%)	41.08%	190 (5.31%)	314 (8.78%)	1
PDL 90	4,436	4,379	22,732	4,259	18,692,364 bp	7,034,082 (37.63%)	41.37%	162 (3.7%)	284 (6.49%)	5

		PDL 20			PDL 35			PDL 90		
Repeat Structure	Subcategory	Number of Sequences	Base pairs (bp)	Percent of total bases	Number of Sequences	Base pairs (bp)	Percent of total bases	Number of Sequences	Base pairs (bp)	Percent of total bases
SINEs:		15,358	1,977,433	7.16	9,082	1,177,081	6.93	10,580	1,364,177	7.3
	Alu/B1	6,675	789,997	2.86	4,058	482,767	2.84	4,657	549,664	2.94
	MIRs	831	92,489	0.33	380	42,144	0.25	547	61,174	0.33
LINEs:		8,070	3,900,988	14.12	4,733	2,379,657	14.02	5,346	2,471,400	13.22
	LINE1	7,533	3,815,850	13.81	4,427	2,332,447	13.74	4,987	2,414,921	12.92
	LINE2	432	69,028	0.25	236	37,507	0.22	284	44,924	0.24
	L3/CR1	72	10,937	0.04	51	8,025	0.05	48	7,456	0.04
	RTE	31	5,055	0.02	19	1,678	0.01	23	3,867	0.02
LTR elements:		9,103	2,684,863	9.72	5,382	1,623,416	9.56	6,424	1,921,379	10.28
	ERVL	763	207,964	0.75	444	122,112	0.72	565	146,318	0.78
	ERVL-MaLRs	3,702	953,116	3.45	2,123	547,496	3.23	2,526	647,013	3.46
	ERR_class I	904	192,573	0.7	510	103,855	0.61	688	137,522	0.74
	ERV_class II	3,652	1,295,460	4.69	2,261	825,629	4.86	2,592	963,033	5.15
DNA elements:		1,589	284,920	1.03	871	156,727	0.92	1,072	198,622	1.06
	hAT-Charlie	973	168,916	0.61	533	90,655	0.53	649	111,562	0.6
	TcMar-Tigger	380	76,556	0.28	201	41,915	0.25	272	58,178	0.31
Unclassified:		412	179,705	0.65	272	117,947	0.69	323	143,355	0.77
Total Interspaced Repeats:		-	9,027,909	32.67	-	5,454,828	32.14	-	6,098,933	32.63
Small RNA:		482	38,950	0.14	345	27,764	0.16	288	21,777	0.12
Satellites		1,232	550,009	1.99	778	313,966	1.85	887	358,543	1.92
Simple Repeats		12,638	859,498	3.11	7,745	664,080	3.91	8,241	508,843	2.72
Low Complexity		1,392	72,542	0.26	829	47,201	0.28	942	51,442	0.28

Regions with significant biogenesis were identified as genome windows with Z-scores greater than 2. Both VRC01 and CHOZN had many statistically significant windows detected across PDLs. For VRC01 PDL 10, 70, and 110, there were 58, 45, and 48 statistically significant windows identified, respectively. The mean biogenesis frequencies in these windows were 1.53 ± 4.75, 2.00 ± 6.66 and 2.02 ± 6.10 for PDLs 10, 70, and 110, respectively. In CHOZN PDL 20, 35, and 90, there were 74, 53, and 78 statistically significant windows identified, respectively. The mean biogenesis frequencies in these windows were 1.97 ± 4.42, 1.3 ± 4.81, and 1.36 ± 3.02 for PDLs 20, 35, and 90, respectively. The biogenesis frequencies for the 15 windows with the highest counts for both VRC01 and CHOZN are shown in **Table 4.3**. Genome-wide counts for each window are shown in **Supplemental Tables S4.12** and **S4.13** for VRC01 and CHOZN samples, respectively. Data visualization revealed that, in addition to the previously identified biogenesis hotspot on chromosome 9 from 14 Mbp to 16 Mbp in VRC01, CHOZN eccDNAs also had elevated biogenesis frequency on chromosome 5 that was not as apparent in VRC01 samples (**Figure 4.2**). Full genome maps for both cell lines at all PDLs are available in **Supplemental Figures S4.1** and **S4.2**.

Table 4.3: Summaries of 15 highest frequency eccDNA biogenesis windows observed for each PDL by cell line.
The start and stop positions for each window with the respective chromosome, eccDNA count, and Z-score.

VRC01 PDL 10					VRC01 PDL 70					VRC01 PDL 110				
Chr	Start	Stop	Count	Z-score	Chr	Start	Stop	Count	Z-score	Chr	Start	Stop	Count	Z-score
9	15000000	15500000	174	36.3	9	15000000	15500000	259	38.6	9	15000000	15500000	197	32.0
9	14500000	15000000	122	25.4	9	14500000	15000000	156	23.1	9	14500000	15000000	158	25.6
9	13500000	14000000	74	15.3	7	134000000	134359064	102	15.0	9	14000000	14500000	108	17.4
9	14000000	14500000	70	14.4	X	106500000	107000000	98	14.4	7	134000000	134359064	98	15.7
7	134000000	134359064	69	14.2	2	41000000	41500000	96	14.1	9	13500000	14000000	90	14.4
X	106500000	107000000	58	11.9	9	13500000	14000000	92	13.5	9	15500000	16000000	80	12.8
2	41000000	41500000	52	10.6	9	14000000	14500000	86	12.6	X	106500000	107000000	78	12.5
4	5000000	5500000	52	10.6	9	12500000	13000000	76	11.1	2	41000000	41500000	70	11.1
5	60500000	61000000	46	9.4	7	88000000	88500000	71	10.4	9	17500000	18000000	64	10.2
4	117000000	117500000	45	9.2	4	5000000	5500000	64	9.3	4	5000000	5500000	58	9.2
7	88000000	88500000	42	8.5	4	117000000	117500000	64	9.3	10	30000000	30500000	54	8.5
9	15500000	16000000	42	8.5	9	13000000	13500000	63	9.2	5	60500000	61000000	53	8.4
10	30000000	30500000	42	8.5	5	60500000	61000000	57	8.3	4	3000000	3500000	52	8.2
9	17500000	18000000	36	7.3	10	30000000	30500000	54	7.8	7	88000000	88500000	51	8.0
5	174000000	174500000	32	6.4	2	34500000	35000000	52	7.5	4	117000000	117500000	49	7.7
CHOZN PDL 20					CHOZN PDL 35					CHOZN PDL 90				
Chr	Start	Stop	Count	Z-score	Chr	Start	Stop	Count	Z-score	Chr	Start	Stop	Count	Z-score
9	15000000	15500000	109	24.2	X	126000000	126500000	144	29.7	9	15000000	15500000	77	25.0
9	16000000	16500000	93	20.6	9	15000000	15500000	94	19.3	9	16000000	16500000	52	16.8
X	126000000	126500000	93	20.6	9	14500000	15000000	84	17.2	9	17500000	18000000	51	16.4
9	14500000	15000000	78	17.2	7	134000000	134359064	79	16.2	9	14500000	15000000	48	15.4
9	16500000	17000000	73	16.1	X	106500000	107000000	66	13.5	9	15500000	16000000	48	15.4
1_1	228000000	228500000	61	13.3	1_1	228000000	228500000	63	12.8	X	126000000	126500000	41	13.1
9	17000000	17500000	50	10.9	2	41000000	41500000	62	12.6	9	16500000	17000000	39	12.5
7	134000000	134359064	47	10.2	5	172500000	173000000	62	12.6	3	135000000	135500000	36	11.5
2	34500000	35000000	42	9.1	4	5000000	5500000	55	11.2	9	14000000	14500000	36	11.5
9	12500000	13000000	42	9.1	8	59500000	60000000	52	10.5	9	17000000	17500000	34	10.8
9	14000000	14500000	42	9.1	10	30000000	30500000	52	10.5	1_1	228000000	228500000	32	10.1
9	15500000	16000000	42	9.1	5	60500000	61000000	50	10.1	9	13000000	13500000	32	10.1
2	41000000	41500000	41	8.8	4	117000000	117500000	49	9.9	2	34500000	35000000	30	9.5
9	17500000	18000000	40	8.6	2	34500000	35000000	47	9.5	9	13500000	14000000	30	9.5
4	5000000	5500000	39	8.4	9	14000000	14500000	45	9.1	2	266000000	266500000	26	8.2

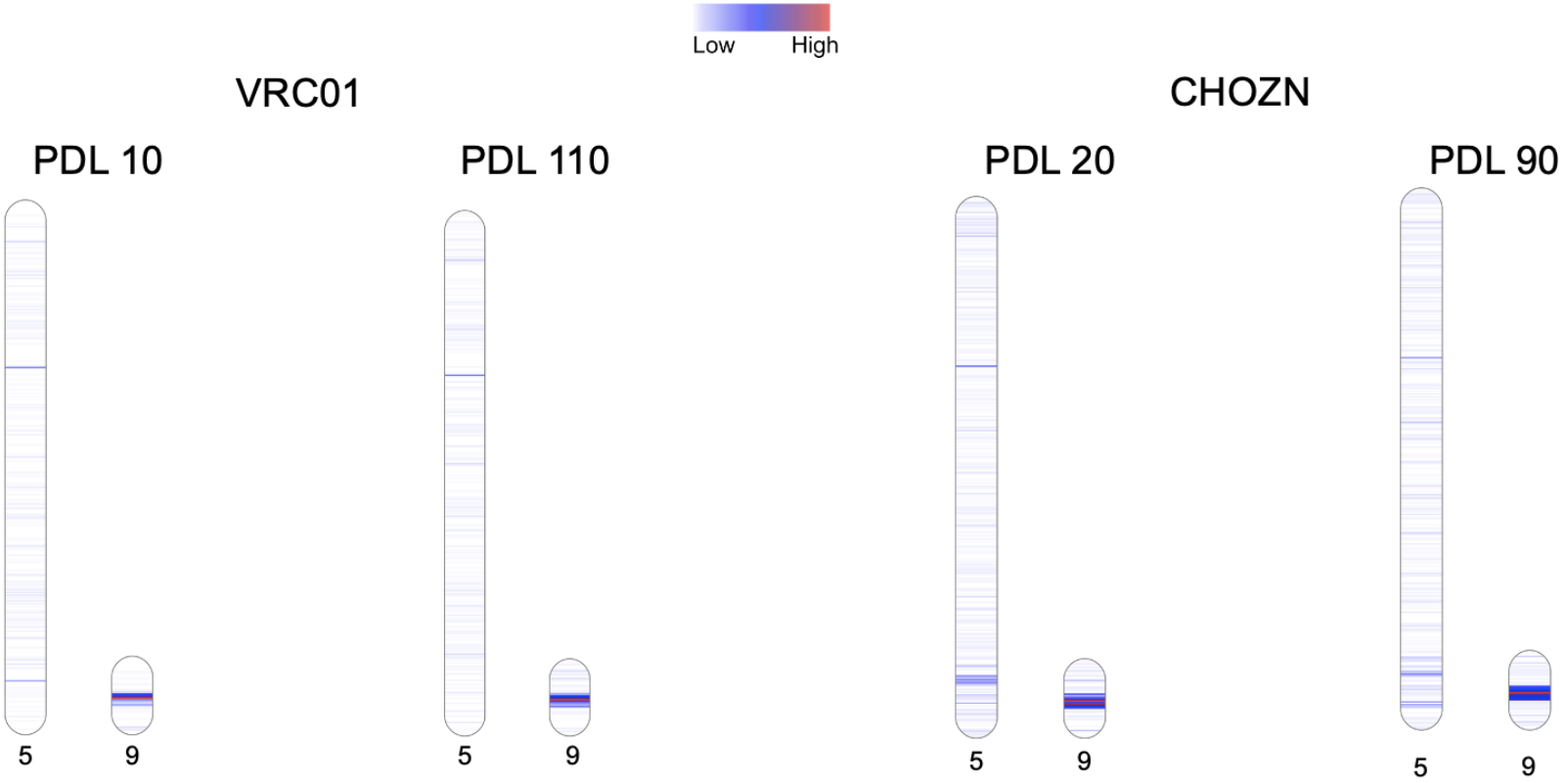


Figure 4.2: Chromosome-scale heatmaps of biogenesis frequencies of eccDNAs mapped to chromosomes 5 and 9 from VRC01 and CHOZN for youngest and oldest PDLs. Biogenesis frequency windows are 500 kbp.

Further analysis was performed to determine the relationship between gene content of eccDNAs and structure of eccDNA-originating regions of the genome. The mean number of repetitive elements flanking eccDNA-encoded genes was lower than the mean number of repetitive elements flanking non-eccDNA-encoded genes. The difference of means was 121 and and 78 in VRC01 and CHOZN samples, respectively and was found to be statistically significant ($p \leq 0.001$). The mean number of neighboring genes flanking all genes in the genome were also analyzed and found no significant difference of means for genes found encoded on eccDNAs compared to other genes (**Figure 4.3**). The differences of means were identified using a permutation test (N=10,000).

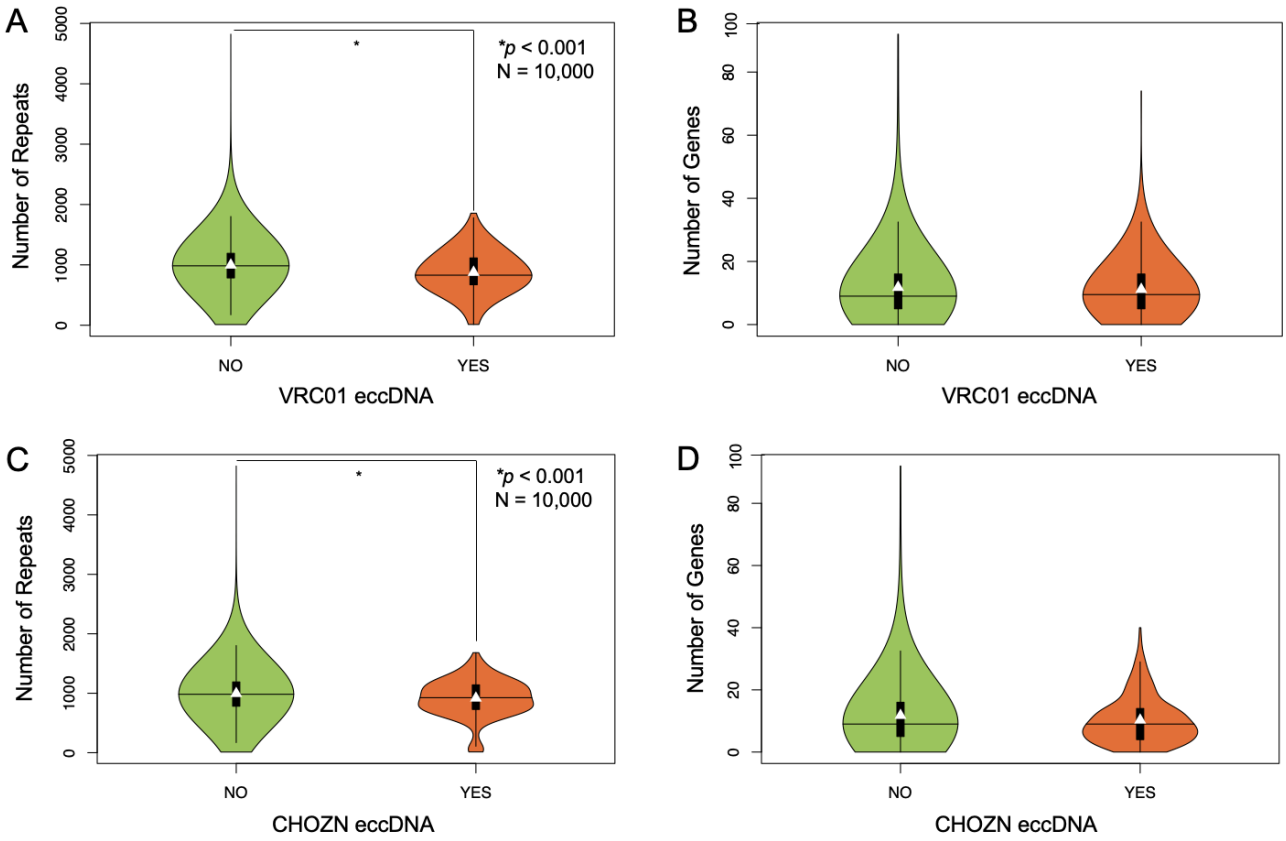


Figure 4.3: Violin plots for eccDNA content in genome structures for the VRC01 and CHOZN cell lines. The eccDNA-encoded genes compared to frequency of repeats in flanking 200 kbp windows for **A)** VRC01 and **C)** CHOZN. eccDNA-encoded genes compared to frequency of genes in flanking 200 kbp windows for **B)** VRC01 and **D)** CHOZN. Gene loci where eccDNAs were not observed (green) and where eccDNAs were observed (orange. White triangles indicate the mean, horizontal bars represent the median, and vertical bars indicate the 1.5x interquartile range of the mean. Data were pooled by cell line across PDLs.

4.4.3 EccDNA gene content and function

The observed eccDNAs harbored many unique genes, 583 genes for VRC01 and 370 genes for CHOZN spanning the three PDLs for each cell line. Of these identified genes, only 82 genes were observed in both cell lines (**Figure 4.4A)**. Further examination showed that over 87% of these genes were unique to a single PDL. **Figure 4.4** shows a Venn diagram of the identified genes by cell line and PDL, which highlights the uniqueness of the eccDNA genes observed. These data would imply that the eccDNA gene content is dynamic for both cell lines across the examined PDLs. A functional enrichment analysis of KEGG pathways and GO terms revealed significant enrichment for five KEGG pathways and two Molecular Function GO terms (adjusted *p*-value ≤ 0.05). The VRC01 PDL 70 samples had significant enrichment in the cytoplasmic translation term, while VRC01 PDL 110 samples had significant enrichment in five ribosome biogenesis related terms. No significantly enriched terms were identified in the VRC01 PDL 10 samples. The CHOZN PDL 20 samples had significant enrichment in genes associated with positive regulation of secretion by cell, while CHOZN PDL 35 samples had enrichment in azole transmembrane transport. The CHOZN PDL 90 samples had significant enrichment in genes related to protein transmembrane import and mitochondrial transmembrane transport (**Figure 4.5A**). A Cellular Component GO term for myelin sheath was observed as enriched for CHOZN PDL 20, and both CMG complex and DNA replication preinitiation complex were enriched for CHOZN PDL 90 (**Figure 4.5B**). The functional enrichment of terms with higher *p*-values and for an excluded condition, VRC01 PDL 24, are shown in **Supplemental Tables S4.14-S4.20**.

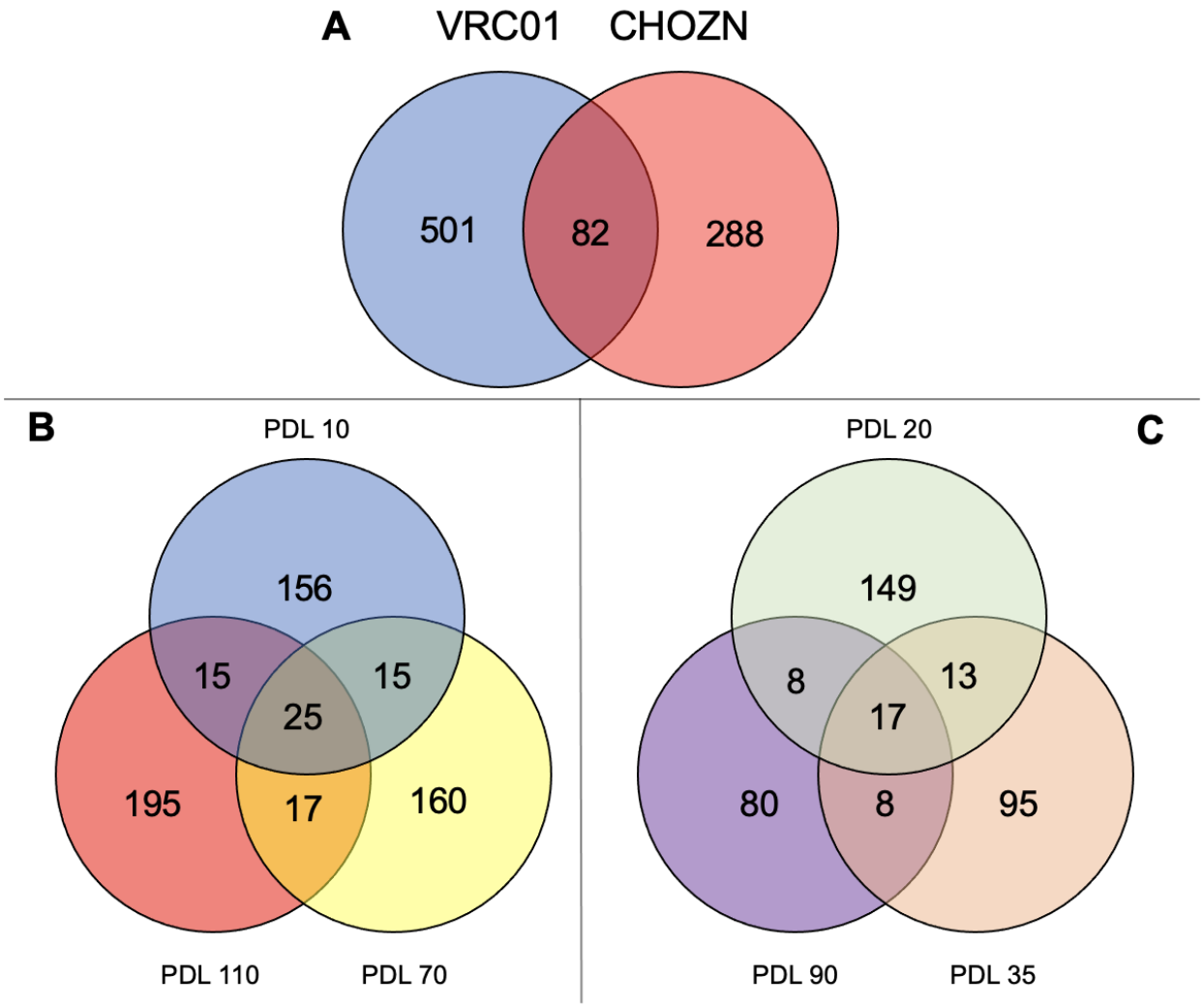


Figure 4.4: EccDNA gene distributions for VRC01 and CHOZN cell lines by PDLs. Venn diagrams of genes observed on eccDNAs for both cell lines for all PDLs **(A)**, for VRC01 by PDLs **(B)**, and CHOZN by PDLs **(C)**.

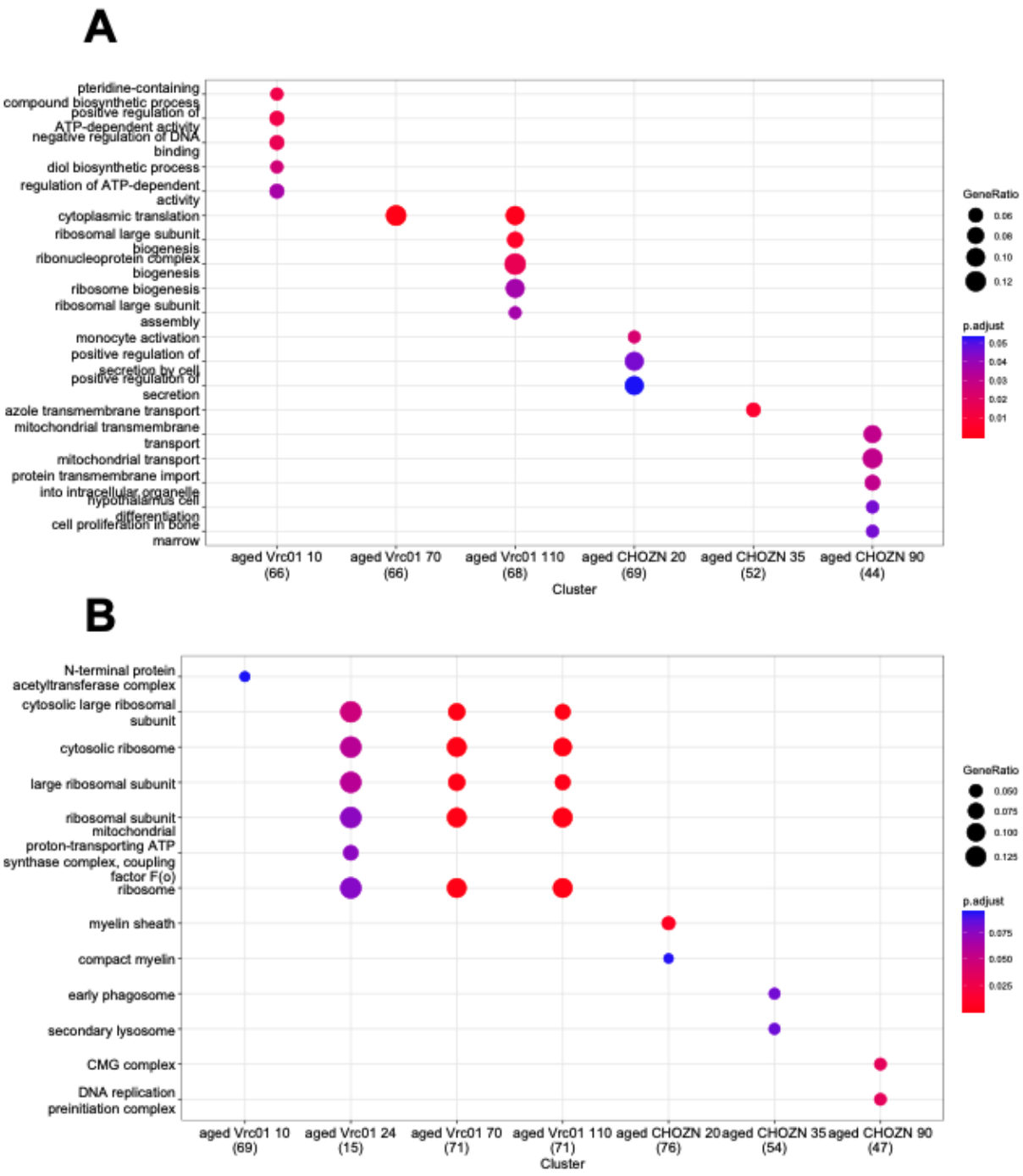


Figure 4.5: Functional enrichment of eccDNA genes observed in CHO cells by PDL. BP is Biological Process, MF is Molecular Function, and CC is Cellular Component. **A)** Comparison of significantly enriched Biological Process GO terms identified in each condition. Circle size reflects gene ratio, or fraction of genes observed per GO term and color of circles indicates statistical significance. **B)** Comparison of significantly enriched Cellular Component GO terms identified by PDL.

4.4.4 Transcriptome analysis

The transcriptomes for both cell lines at each of the three PDLs were obtained to understand gene expression changes with respect to increasing PDL. It was desired to correlate eccDNA gene gain or loss with differential expression to identify potential eccDNA-mediated gene expression. Several genes were observed to have differential gene expression between the young and old PDLs. For VRC01, 82 genes were identified between PDL 10 and PDL 110 with significantly different expression levels, where 43 genes had higher expression in PDL 110 relative to PDL 10 and 39 genes had lower expression in PDL 110 relative to PDL 10. For CHOZN, 64 genes were observed to have significant differential expression, where 28 genes had higher expression in PDL 90 relative to PDL 20 and 36 genes had lower expression in PDL 90 relative to PDL 20. The 10 genes with the greatest differential expression between PDLs are shown in **Tables 4.4** and **4.5** for VRC01 and CHOZN, respectively. Among the differentially expressed genes, two alternatively spliced proteins from *Fbxo9* were observed, where one splice variant was expressed more in older PDL samples and the other splice variant was expressed less in older PDL samples. Additionally, multiple genes associated with cancer and genome instability were observed to change in gene expression levels for the older PDLs for both cell lines. Interestingly, none of the 142 differentially expressed genes were observed to be encoded on the eccDNAs observed in this study. A full list of the differentially expressed genes for VRC01 and CHOZN can be found in **Supplemental Tables S4.21** and **S4.22,** respectively. A summary of TMM (Trimmed Mean of M) values for genes observed on eccDNAs in VRC01 and CHOZN is presented in **Supplemental Tables S4.23** and **S4.24**, respectively.

Table 4.4: Genes with greatest expression level differences between PDL 10 and PDL 110 for VRC01. CPM – counts per million, logFC – log2 fold change (PDL 110/PDL 10), FDR – false discovery rate.

RefSeq ID	Gene	log2(CPM)		logFC	FDR	Description
		PDL 10	PDL 110			
XP_027280669.1	*Hcfc1r1*	0	7.34	7.34	1.21E-07	host cell factor C1 regulator 1 isoform X1
XP_027260267.1	*Tial1*	0	6.01	6.01	2.89E-11	nucleolysin TIAR isoform X6
XP_027270138.1	*Cdv3*	0	5.7	5.70	2.01E-08	protein CDV3 homolog isoform X7
XP_027266670.1	*Smco4*	0	5.57	5.57	1.22E-06	single-pass membrane and coiled-coil domain-containing protein 4 isoform X2
XP_027252491.1	*Aaas*	0	5.23	5.23	1.19E-05	aladin isoform X2
XP_027243991.1	*Pold2*	0	5.18	5.18	6.62E-07	DNA polymerase delta subunit 2 isoform X2
XP_027243992.1	*Pold2*	0	5.18	5.18	6.62E-07	DNA polymerase delta subunit 2 isoform X2
XP_027243993.1	*Fbxo9*	0	5.18	5.18	6.62E-07	F-box only protein 9 isoform X3
XP_027266907.1	*Fbxo9*	0	5.16	5.16	2.51E-06	F-box only protein 9 isoform X3
XP_027254036.1	*Tardbp*	0	4.91	4.91	0.001587	TAR DNA-binding protein 43 isoform X4
XP_027250346.1	*Oard1*	4.68	0	-4.68	1.96E-04	ADP-ribose glycohydrolase OARD1 isoform X3
XP_027283281.1	*Lyrm7*	4.79	0	-4.79	3.33E-05	complex III assembly factor LYRM7 isoform X2
XP_027248332.1	*Nfyb*	4.82	0	-4.82	3.01E-05	nuclear transcription factor Y subunit beta isoform X3
XP_027248333.1	*Nfyb*	4.82	0	-4.82	3.01E-05	nuclear transcription factor Y subunit beta isoform X3
XP_027252192.1	*Slc38a2*	4.87	0	-4.87	2.57E-05	sodium-coupled neutral amino acid transporter 2 isoform X2
XP_027282890.1	*Ndel1*	4.96	0	-4.96	2.85E-06	nuclear distribution protein nudE-like 1 isoform X2
XP_027266906.1	*Fbxo9*	5.21	0	-5.21	3.23E-07	F-box only protein 9 isoform X1
XP_027274670.1	*Zbtb18*	5.24	0	-5.24	3.23E-07	zinc finger and BTB domain-containing protein 18 isoform X1
XP_027265675.1	*Kiaa1191*	5.29	0	-5.29	3.23E-07	putative monooxygenase p33MONOX isoform X3
XP_027266790.1	*Syncrip*	5.91	0	-5.91	9.82E-11	heterogeneous nuclear ribonucleoprotein Q isoform X3

Table 5: Genes with greatest expression level differences between PDL 20 and PDL 90 for CHOZN. CPM – counts per million, logFC – log2 fold change (PDL 90/PDL 20), FDR – false discovery rate.

RefSeq ID	Gene	log2(CPM)		logFC	FDR	Description
		PDL 20	PDL 90			
XP_027266907.1	*Fbxo9*	0	5.4	5.4	1.02E-11	F-box only protein 9 isoform X3
XP_027251257.1	*Ahcyl1*	0.32	5.46	5.14	0.00299	S-adenosylhomocysteine hydrolase-like protein 1 isoform X3
XP_027250245.1	*Ppil1*	0	4.92	4.92	1.92E-06	peptidyl-prolyl cis-trans isomerase-like 1 isoform X2
XP_027256935.1	*Mpc1*	0.05	4.81	4.76	0.01551	mitochondrial pyruvate carrier 1 isoform X1
XP_027269723.1	*Kmt5a*	0	4.58	4.58	1.82E-07	N-lysine methyltransferase KMT5A isoform X4
XP_027251920.1	*Pacsin2*	0	4.24	4.24	1.06E-05	protein kinase C and casein kinase substrate in neurons protein 2 isoform X3
XP_035296909.1	*Rpe*	0	4.11	4.11	1.32E-05	ribulose-phosphate 3-epimerase isoform X4
XP_027251105.1	*Golga7*	0	4.09	4.09	3.81E-05	golgin subfamily A member 7 isoform X2
XP_027283240.1	*Fam114a2*	0	3.92	3.92	0.00011	protein FAM114A2 isoform X2
XP_035295752.1	*Ptbp3*	0	3.82	3.82	0.00093	polypyrimidine tract-binding protein 3 isoform X8
XP_035295258.1	*LOC100755823*	4.13	0	-4.13	1.06E-05	pleckstrin homology domain-containing family G member 5 isoform X6
XP_027270446.1	*Map4*	4.43	0.23	-4.2	0.00352	microtubule-associated protein 4 isoform X1
XP_035306543.1	*S100a5*	4.23	0	-4.23	4.05E-06	protein S100-A5 isoform X5
XP_027279375.1	*Strbp*	4.27	0	-4.27	3.53E-06	spermatid perinuclear RNA-binding protein isoform X2
XP_035304091.1	*Ddx5*	4.29	0	-4.29	7.60E-06	probable ATP-dependent RNA helicase DDX5 isoform X3
XP_027283239.1	*Fam114a2*	4.3	0	-4.3	2.74E-06	protein FAM114A2 isoform X1
XP_027252932.1	*Rpe*	4.41	0	-4.41	7.81E-07	ribulose-phosphate 3-epimerase isoform X3
XP_027253713.1	*C2H18orf32*	4.44	0	-4.44	2.74E-06	UPF0729 protein C18orf32 homolog
XP_027253096.1	*Dnpep*	4.86	0.07	-4.79	5.38E-08	aspartyl aminopeptidase isoform X3
XP_027266906.1	*Fbxo9*	5.36	0	-5.36	2.82E-11	F-box only protein 9 isoform X1

4.5 Discussion

Methods to improve long-term perfusion cultures of CHO cells are highly desired in the field of biomanufacturing[42,43,232,233]. Despite significant advancements in biomanufacturing to increase culture yields, intrinsic instability of CHO cell lines continues to be a hurdle toward the implementation of continuous biomanufacturing for CHO-based products. The CHO genome is highly prone to undesirable rearrangements, changes in ploidy, and transgene exclusion, which negatively affects product quality and process economics[2]. The genomic changes seen in CHO are also commonly observed in other cell types with instable genomes such as cancer cells[12,33,95,173,234,235] or cells of advanced age[104,135,182,228]. EccDNAs are an indicator of an unstable genome[3] and have been observed to be highly dynamic in the VRC01 cell line in fed-batch cultures[11].

This study utilized the cell lines VRC01 and CHOZN, which have both been recently characterized in detail to allow wide-spread study of industry-relevant cell lines[230]. Cordova et al, 2022 presented reference data for fed-batch shake flask and bioreactor cultures for both cell lines. VRC01 cultures show higher cell densities, while CHOZN cultures produced higher protein titers. Interestingly, when increasing VRC01 PDL cultures were examined, a q_p stability threshold of around PDL 24 was observed. For CHOZN, the q_p stability threshold was maintained until about PDL 60. Thus, based on cell-specific productivity, the CHOZN cell line was found to be more stable than VRC01. It should be noted that the cells used for this study were aged and banked from shake flask cultures, while samples from a perfusion culture would have been preferable;

however, cells aged in shake flasks are a suitable substitutes[210]. The less stringent control of process parameters, such as pH and DO, may have impacted the results of this study.

Cell division exposes cells to various environmental stressors, leading to DNA damage, telomere shortening, oncogene activation, and organelle stress. This is collectively referred to as cellular aging and can push cells into senescence[236,237]. The accumulation of eccDNAs has been observed during *in vitro* aging in rat lymphoblasts and human fibroblasts[238]. Accumulation of genomic rearrangements has been thought to be the key driver of cell aging[239]. Previous studies have reported genome instability in ageing cells in varying contexts such as, hypermutation in in tandem repeat-rich regions[240], loss of transgenes[241] and loss of repetitive sequences[242]. The genes and repeat regions that had disappeared were later discovered to be present on eccDNAs of varying size[243]. The contribution of eccDNAs to cellular aging has been documented in other species. One subclass of eccDNAs, referred to as telomeric circles (t-circles), play a role in maintaining telomere length. Gradual shortening of telomere regions hastens genome instability and drives cells toward senescence[244]. T-circles have been observed to reverse telomere shortening via recombination with telomeres[227]. This process could occur indefinitely if t-circles are capable of rolling circle amplification[245,246]. Recent research has revealed that t-circles arise from damage to telomeric regions, which creates internal loops (i-loops)[226].

The content of eccDNA has been extensively studied in yeast[4,121,229]. Yeast is one of the few cell types used in biomanufacturing that has been investigated for circular DNA

content. Young yeast cells exhibit high diversity in eccDNA sequence structure and composition; however, this heterogeneity decreases with cellular aging. Notably, aged yeast cells harbor rDNA genes within eccDNAs[228]. This rDNA accumulation leads to the formation of visible nucleolar fragments within cells. Furthermore, the insertion of plasmids containing these rDNA genes resulted in a reduction of cell division by up to 40%[229]. It was proposed that the extrachromosomal copies of rDNA genes in yeast may be generated by mutations to homologs of the *Sgs1* gene. *Sgs1* is associated with Werner's Syndrome in humans, a disease that accelerates the aging process and shortens patient lifespan[247]. Thus, eccDNA gene content may be directly correlated with the age of cells grown in culture.

The sequence characteristics and structural annotations conducted in this study (**Tables 4.1** and **4.2**) are directly comparable to those performed in a previous study on VRC01 eccDNA grown in fed-batch ambr®250 bioreactors[11]. The mean length of VRC01 eccDNAs in this study was greater than that observed in previous studies, but it is not clear that PDL significantly impacted mean sequence length. The disparity in average eccDNA length more likely can be attributed to factors such as culture methods, library preparation, and sequencing variations. In regard to repetitive elements, the VRC01 eccDNAs in this study exhibited a slightly lower proportion of short interspaced nuclear elements (SINEs) compared to the previous study, however, like other repeat distributions, the SINE content remained consistent with small variations. This study observed slight differences in the GC content and LINE repeat abundance between VRC01 and CHOZN eccDNAs. The increase in GC content for CHOZN eccDNA (>2%)

and higher abundance of LINE repeats in VRC01 (>3%) are hypothesized to be from differences in eccDNA biogenesis sites, as slight variations in the genome are to be expected when generating new cell lines.

The annotation of other sequence structures, such as genes, rDNA, and tRNA, were found to be consistent across the different PDLs and cell lines. Gene annotation showed that ~85% of genes identified for a particular cell line were only found in one of the three PDLs, which is consistent with other work that examined CHO cell eccDNA (**Figure 4.4**). Though rDNA annotation did not identify many rDNA genes, it must be reiterated that CIDER-seq is not a quantitative method; while a broad diversity of rDNA genes was not observed, comments cannot be made regarding the abundance of rRNA. While RNA-seq data is presented, it was aligned to the Chinese hamster reference transcriptome which may not have rDNA genes properly included. Note that when searching NCBI for rDNA genes, 804 human genes were available, but only 25 Chinese hamster rDNA genes were found (**Supplementary Table S4.2).** Thus, quantifying rRNA content would require better annotation of this gene family in the Chinese hamster. It was expected that both cell lines would have similar structural features on eccDNAs since both cell lines were derived from the CHO K-1 host. While all CHO cell lines are derived from the same tissue[13], these cell lines should be closely related as they diverged from the same lineage. Notably, there was a high degree of similarity in the distribution of eccDNA sequence structures, however the specific content of genes in eccDNA was highly variable between cell lines.

This study observed a high frequency of eccDNA biogenesis on chromosome 9 in both cell lines across all PDLs. The elevated levels of biogenesis could be attributed to instability in the chromosome due to the density of repeat structures[11,79]. Little variation in biogenesis frequency was observed as cells aged, however, there was an increase in frequency observed on chromosome 5 for CHOZN (**Table 4.3** and **Figure 4.2**). This could likely be attributed to structural variations between two cell lines. Repeat-rich regions of the genome, such as those found on chromosome 9, are highly susceptible to DNA circularization[8,123,136]. This raises questions about the impact of genome structure on eccDNA-mediated gene amplification and whether the CHO genome may be compartmentalized as a two-speed genome as reported in other organisms.

Currently, two-speed genomes have only been well characterized in plant pathogens[164,165,248,249]. No reports of a two-speed genome in mammalian species have been published as of this writing. Organisms with two-speed genomes exhibit differential evolution of certain genes and structures compared to others[164]. In filamentous plant pathogens, such as *Fusarium graminearum*, virulent effector genes were identified in regions of the genome with a higher density of repeats and transposable elements, whereas typical “housekeeping” genes reside in gene-dense regions[165,248]. Repeat-rich regions of the genome are more susceptible to variant accumulation[250], recombination[251], and the formation of eccDNA[122,142]. This arrangement allows for faster variant accumulation and subsequent evolution of virulent effector genes in repeat-rich regions. The hypothesis was that that eccDNAs could be an extension of the two-speed genome, given that eccDNA-mediated overexpression has been shown to enable adaptation in

other organisms, such as therapeutic resistance in cancer[154] and pesticide resistance in plants[5]. However, the data analyzed in this study is contradictory to this hypothesis (**Figure 4.3**). A possible explanation for the observation of eccDNA genes were in gene-rich regions of the genome in CHO is that circularization of sequences via transcription is a more frequent mechanism of eccDNA biogenesis in CHO cells[121]. This hypothesis would also account for overrepresentation of genes in eccDNA compared to the genome[211] (~3.5% versus ~1%). It should be noted that this would not exclude biogenesis of eccDNAs through other pathways, as the high levels of repeat content (~40%) in CHO cell eccDNAs provide evidence of potential recombination between repeat regions[7,142,214].

The contradictory behavior observed in the two-speed genome analysis may be because mammalian genomes evolve under different selective pressures compared to pathogen genomes. A review of six mammalian genomes revealed positive selection of genes related to immunity and defense, sensory perception, metabolism, and fertility[252], while a comparable study in Smut Fungus, *Sporisorium reilianum,* identified positive selection of virulent effectors, cell division, and DNA integration genes[253]. The evolution of two-speed genomes in filamentous plant pathogens, characterized by positive selection of DNA integration genes and virulent effectors, has contributed to the heterogeneity of their genome size through repeat expansion and gene accumulation[254]. While it is possible that mammalian genomes may exhibit similar compartmentalization, further studies examining genome structure and gene selection independent of the presence of eccDNA are necessary to draw conclusions. Our findings indicate that, if the CHO genome operates as a two-speed genome, it does not appear to influence the formation of eccDNAs.

Functional enrichment of eccDNA genes yielded few statistically significant GO terms due to the relatively low number of human orthologs identified (<100 per condition). Of significance, the VRC01 samples showed enrichment in terms associated with protein production. However, VRC01 PDL 10 did not have any significant enrichment in terms shared with VRC01 PDL 70 and PDL 110. Previous studies on the gene content of young VRC01 (~PDL 10) revealed statistically significant enrichment in biological process (BP) terms related to cytoplasmic translation, ribosomal small subunit assembly, and noncoding RNA (ncRNA) processing[11] (**Figure 4.5**). Additionally, functional enrichment analysis of VRC01 PDL 24 samples, as shown in **Supplemental Table S4.15**, revealed the same five CC GO terms enriched in VRC01 PDL 70 and 110 samples ($p \leq 0.1$) despite having only 15 human orthologs analyzed. It is important to note that VRC01 PDL 24 samples were not considered in this study due to insufficient library preparation compared to other samples.

These results indicate that the VRC01 cell line is likely to contain eccDNAs that enhance protein production. The difference in functional enrichment between VRC01 and CHOZN cell lines may be attributed to the selection of clones during the cell line development process, where cells with a high-productivity phenotype are preferred[213]. While some enrichment was identified in human orthologs of CHOZN eccDNA genes, no overlapping GO terms were identified between PDLs. The most notable observation was the presence of multiple significantly enriched BP GO terms in the CHOZN PDL 90 genes pertaining to mitochondrial transport. The comparison between the VRC01 and CHOZN

cell lines reveals a clear difference in the function of their respective eccDNA genes (**Figure 4.5**). Previous studies have shown that yeast eccDNAs become increasingly homogenous with increasing cellular age, as they near senescence and contain a high percentage of ribosomal DNA (rDNA) (>95%)[229]. Although a diverse range of rDNAs were not detected in the present study, the relatively homogenous nature of the eccDNA genes in VRC01 compared to the less consistent function in CHOZN suggests a potential difference in the age of the master cell banks.

The decline in DNA repair mechanisms in aging cells often leads to the accumulation of genetic damage over time[255]. Qian et al. previously described transcriptome shifts in a proprietary, recombinant CHO cell line generated from a glutamine synthetase double-negative host ($GS^{-/-}$)[221]; young cells were compared to those 130.5 PDLs later. In short, genes pertaining to cell cycle regulation and DNA repair were downregulated while genes related to lysosomes, redox metabolism, and lipid metabolism were upregulated[221]. However, none of the differentially expressed genes identified in the study were found to have substantial differential expression (≤ -2 or ≥ 2 log_2 fold change) in the VRC01 or CHOZN samples in this study. This unusual result could be influenced by several factors such as aging protocol, media composition, culture scale (shake flask versus 5-L bioreactor), and library preparation and sequencing methods. Further, the study conducted by Qian et al. used a proprietary cell line (Bristol Meyers Squib) that is not publicly accessible for open study.

Although the same genes identified in the previous study were not observed in our data, we observed differential expression in other genes related to genome stability that were not reported in Qian et al. (**Tables 4.4** and **4.5**). One significant change in gene expression observed in this study was the simultaneous up- and down-regulation of alternatively spliced proteins from the *Fbxo9* gene. *Fbxo9* functions in protein ubiquitin ligase; F-box proteins may play a role in DNA repair and synthesis[256]. Further, *Fbxo9* interacts with multiple intracellular signaling pathways and can act as both a tumor suppressor and an oncogene depending on the function of other signaling pathways[257]. Other notable shifts in gene expression include down regulation of *Nfyb*, a transcription factor responsible for controlling growth factors[258] in VRC01 and upregulation of *Pold2*, a DNA polymerase subunit critical for DNA replication and repair. Up-regulation of *Pold2* is considered a reliable biomarker in ovarian cancers[259]. Downregulation of *Ddx5*, an RNA helicase, was observed in CHOZN which has been linked to poor prognoses for pancreatic cancer patients[260].

Many studies of eccDNA report clear evidence of eccDNA-mediated gene overexpression in varying contexts[8,119,153,223]. Gene expression data of eccDNA-harbored genes is shown in **Supplemental Tables S4.23 and S4.24**, however, without a large change in gene expression, identifying eccDNA-mediated changes becomes difficult to discern. While none of the genes observed on an eccDNA could be correlated with the larger changes in gene expression, it is still possible that eccDNAs are impacting expression patterns. One study has shown that incomplete genes and promotor or enhancer elements amplified via circularization can act as transcription factor sponges

that divert necessary factors required for gene expression away from the genomic copy of the gene. These eccDNAs generate small, interfering RNAs that inhibit expression[261]. While characterizing the promoter and enhancer elements on eccDNA in CHO would provide tremendous insight into how eccDNA impacts gene expression, the current genome construction does not provide annotation of promoters or enhancers[79].

4.6 Conclusion

The distribution of structural features such as genes, repeats, tRNA, and ORIs is largely consistent across cell age and across CHO cell lines derived from the same lineage. EccDNA gene content has been shown to be highly variable with cell aging, with the most substantial differences being observed between cell lines, which may be due to genome structural variation during cell line derivation. While genome structure has been shown to influence gene circularization, genes susceptible to eccDNA biogenesis do not exhibit a bimodal architecture in the CHO genome. Our results provide no clear evidence that eccDNAs play a significant role in the shifts in gene expression observed with CHO cell aging. Despite inconsistent results with previous studies on aged CHO cells, our transcriptome data revealed the presence of additional genes related to genome instability, including prominent cancer biomarkers, that were differentially expressed.

CHAPTER FIVE

DYNAMICS OF AMINO ACID METABOLISM, GENE EXPRESSION, AND CIRCULOMICS IN A RECOMBINANT CHINESE HAMSTER OVARY CELL LINE ADAPTED TO MODERATE AND HIGH LEVELS OF EXTRACELLULAR LACTATE

5.1 Abstract

Accumulation of metabolic wastes in cell cultures can diminish product quality, reduce productivity, and trigger apoptosis. Limitation or removal of unintended waste products from Chinese hamster ovary (CHO) cell cultures has been attempted through multiple process and genetic engineering avenues with varied levels of success. One study demonstrated a simple method to reduce lactate and ammonia production in CHO cells by adaptation to extracellular lactate; however, the mechanism behind adaptation was not certain. To address this profound gap, this study characterizes the phenotype of a recombinant CHO K-1 cell line that was gradually adapted to moderate and high levels of extracellular lactate and examines the genomic content and role of extrachromosomal circular DNA (eccDNA) during and gene expression on the adaptation process. More than 500 genes were observed on eccDNAs. Notably, more than 1,000 genes were observed to be differentially expressed at different levels of lactate adaptation, while only 137 genes were found to be differentially expressed between unadapted and high levels of lactate adaptation; this suggests stochastic switching as a potential stress adaptation mechanism in CHO cells. Further, these data suggest alanine biosynthesis as a potential stress-mitigation mechanism for excess lactate in CHO cells.

5.2 Introduction

In 2021, global sales of biopharmaceuticals reached an all-time high of $343 billion. A majority of biopharmaceutical sales were generated from monoclonal antibody (mAb) products, which are often commercially produced in Chinese hamster ovary (CHO) cell lines. CHO cell lines are broadly utilized due to relative ease of culture, non-susceptibility to viruses, and thorough history of regulatory approval. The current industry-standard for CHO cell manufacturing is fed-batch cultures in which nutrients are gradually added to a culture over an extended period; however, fed-batch operation does not allow for easy removal of metabolic waste products, such as lactate and ammonia. Limiting lactate production in CHO cell cultures has been thoroughly studied as excessive lactate in cell cultures imparts a suboptimal environment by increasing acidity in uncontrolled cultures[70,208], negatively impacts viable cell density (VCD)[25], and reduces recombinant protein productivity[25,57]. Strategies to limit or eliminate lactate from cultures include pH-controlled delivery of glucose[73], cell line engineering[71,262], and controlled lactate feeding[75]. The detrimental impacts of lactate can be observed at concentrations as low as 20 mM, and cultures are often unable to recover when concentration exceeds 40 mM[209]. Often, lactate is produced in the early stages of a fed-batch culture and accumulates until a metabolic switch causes the cells to consume the excess lactate[60-62]. While the exact mechanisms behind this switch are unclear, multiple hypotheses have been proposed to explain this observation, such as depletion of glucose and/or glutamine[59,63-65,263], shifts in pH and/or temperature[66,67], and increased oxidative capacity[68].

In addition to metabolic waste accumulation, cells grown in suspension culture experience other stresses such as sheer stress[54] and nutrient depletion[264]. Multiple stress response mechanisms exist to mitigate these challenges at a variety of levels such as substrate[58], gene expression[76], and gene regulation[77]. Recent studies have identified a microevolutionary stress adaptation mechanism by gene overexpression via extrachromosomal circular DNAs (eccDNAs)[5,119,153,154]. EccDNAs are a common vehicle for gene amplification, which may correlate with gene overexpression in a subpopulation of cells[150,151,153]. EccDNAs are generated from the genome through multiple biogenesis pathways[121,122,128,214]. Genetic heterogeneity of cells is perpetuated by the randomness of eccDNA generation and recombination[265]. This study characterized the relationship between lactate adaptation and genetic changes within CHO cells that may mediate the adaptation via eccDNA microevolution transcriptomic shifts. A recombinant CHO K-1 cell line was gradually adapted to higher levels of extracellular lactate in 10 mM increments up to 60 mM in shake flask cultures. Batch cultures were sampled daily to characterize growth profiles, amino acids, and other key metabolites. Additionally, cells for eccDNA and transcriptome analysis were also harvested at two levels of lactate adaptation. EccDNAs were purified, sequenced, and annotated for gene content. Transcriptome data was then integrated to identify eccDNA-derived transcripts and expression profile patterns.

5.3 Materials and Methods

5.3.1 Cell adaptation

The cell line used in this study was a recombinant CHO K-1 cell line that expresses an anti-HIV monoclonal antibody (VRC01); the cells were generated and donated by the NIH. All cultures were grown in ActiPro media (Cytiva) supplemented with 6 mM of glutamine. During the adaptation process, cells were cultured in 125 mL baffled, vented shake flasks (VWR®, Radnor, PA) with a 30 mL working volume. Cultures were maintained in an incubator at 37°C and 5% CO_2 with a shake speed of 180 rpm (0.75 inch throw). Cells were progressively passaged into higher concentrations of lactate in 10 mM increments up to 60 mM. Adaptation was achieved when three criteria were met: 1) the growth rate in lactate-supplemented media was equivalent to that of unadapted cells in lactate-free media, 2) the VCD reached $4x10^6$ cells/mL by Day 3, and 3) the first two criteria were both sustained over three passages. After adaptation, cells were banked in a 10% dimethylsulfoxide (DMSO) media and stored in liquid nitrogen. Stocks from the liquid nitrogen were then thawed before beginning the next adaptation step (**Figure 5.1).**

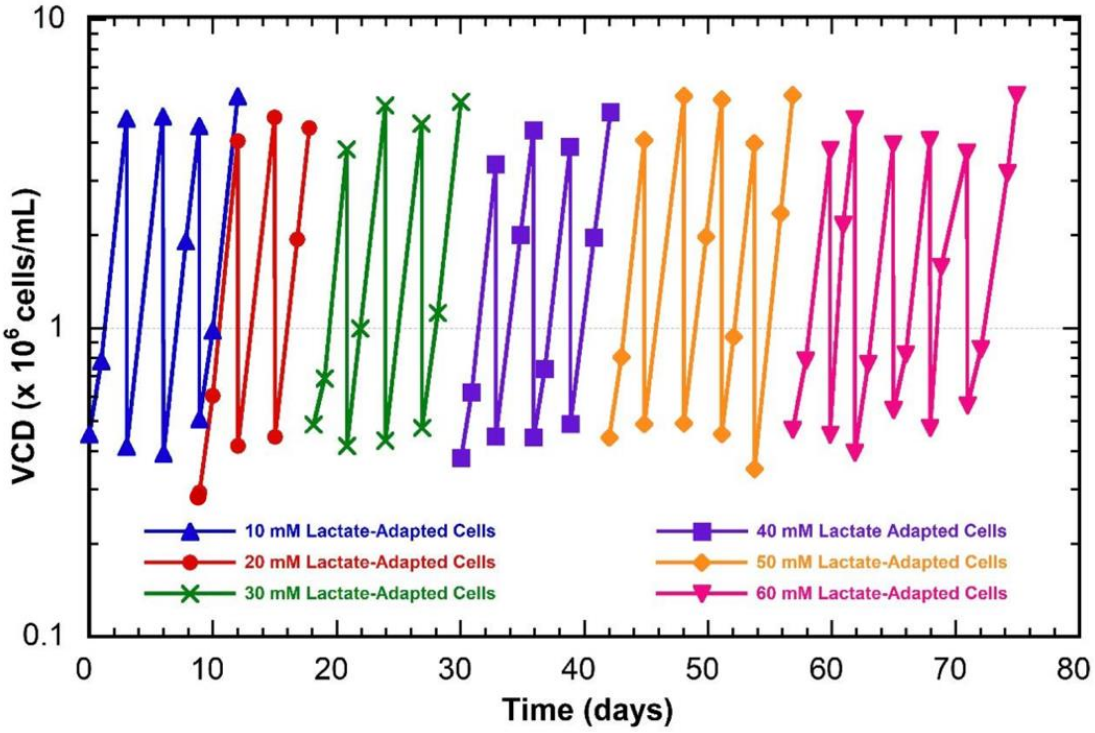


Figure 5.1: Growth profiles during the lactate-adaption process. Cells were gradually passaged into higher lactate concentrations at 10 mM increments. 10 mM adaptation (blue triangles), 20 mM adaptation (red circles), 30 mM adaptation (green crosses 40 mM adaptation (purple squares), 50 mM adaptation (yellow diamonds), and 60 mM adaptation (pink inverted triangles).

5.3.2 Batch cultures

Frozen stocks of unadapted, 30 mM-adapted, and 60 mM-adapted cells were rapidly thawed into media containing the respective amount of lactate. Cells were seeded into new flasks three days after thaw with a target seeding density of 0.5×10^6 cells/mL. Triplicate cultures (N=3) used for each condition; lactate-adapted cells were also grown in media without supplemental lactate. Incubators were maintained at 37°C and 5% CO_2 with an agitation rate of 180 rpm. Flasks were sampled daily for viable cell density (VCD),

viability, metabolites, and amino acids. Viability and VCD were measured on a Vi-CELL™ XR Cell Viability Analyzer (Beckman Coulter, Brea, CA). Glucose, lactate, glutamine, glutamate, ammonia, and IgG were measured on a Cedex Bio Analyzer (Roche Diagnostics, Basel, Switzerland). Concentrations of amino acids, with the exception of glutamine and glutamate, were determined via capillary electrophoresis with high pressure mass spectrometry (CE-HPMS) using the REBEL (908 Devices, Boston, MA); samples were diluted 1:100 in REBEL diluent before analysis.

5.3.3 Library preparation

Unadapted, 30 mM-adapted, and 60 mM adapted cells were thawed from liquid nitrogen stocks into respective medias. Cultures were passaged after three days into duplicate flasks (N=2) and harvested three days later in early exponential growth. Cell pellets of approximately $4x10^6$ cells were collected from each flask, placed in RNAlater (Thermo, AM7020), and kept at -20°C until library preparation. Approximately $1x10^6$ cells from each pellet were used for total RNA and genomic DNA (gDNA) extraction. Extractions were performed using an RNEasy midi kit (Qiagen, 74004) and a DNEasy Blood and Tissue kit (Qiagen, 69504) per the manufacturer's recommended protocols. Total gDNA was quantified with a Qubit fluorometer 2.0 (Thermo, Q32866) before it was used as starting material for a Phi-29-mediated rolling circle amplification (RCA) and magnetic bead purification (KAPA Pure Beads, Roche, KK8000) as described in the CIDER-seq protocol[10]. EccDNA libraries had SMRTbell barcodes adapted to the sequences by a third-party vendor before sequencing on a PacBio Sequel II with HiFi reads. RNA samples were treated with a DNAse solution prior to sample quantification.

NEBNext Ultra II RNA Library Prep Kit for Illumina was used for library preparation per the manufacturer's recommended procedures and pooled in equimolar ratios for sequencing. Paired-end reads for each sample (2x150bp) were collected on an Illumina NovaSeq 6000 S4 flow cell to an approximate depth of 20 million read pairs.

5.3.4 Bioinformatic pipeline

Raw eccDNA sequencing data was processed using the DeConcat algorithm from the CIDER-seq protocol[10] to confirm circularity and genomic origin. Sequences for biological replicates for each condition (N=2) were combined together before clustering to a 90% sequence identity threshold using CD-Hit[185]. Raw RNA-seq data had sequencing adapters removed with Trimmomatic[194] before checking data quality with FastQC. Reads were then aligned to the reference transcriptome with Bowtie2[195]. Transcript abundance was calculated with RSEM[196], which was used to calculate differential expression with edgeR[197].

5.3.4.1 EccDNA structural analysis

BLAST[193] was used to annotate origins of replication (ORIs) and ribosomal DNAs (rDNAs) on observed eccDNAs. Custom databases were made by collecting known mammalian origins of replication (N=118, retrieved on 07/14/2022) and rDNA genes in human (N=804), mouse (N=81), rat (N=153), and Chinese hamster (N=25) (retrieved on 01/09/2023) from NCBI. BLAST searches were conducted with an e-value of 1e-50. Databases for custom ORI and rDNA BLAST searches are shown in **Supplemental Tables S5.1 and S5.2,** respectively. Transfer RNAs (tRNAs) were identified using

tRNAscan-SE 2.0[189]. Repetitive motifs on eccDNAs were annotated and masked using RepeatMasker (Smit, AFA, Hubley, R & Green, P. *RepeatMasker Open-4.0*.2013-2015 <http://www.repeatmasker.org>). Repeat-masked sequences had genes annotated using *Maker*[186].

5.3.4.2 EccDNA genomic origins

Observed eccDNAs were BLASTed against the PICRH Chinese hamster reference genome[79] to find the most likely genomic origin of sequences. Parameters used for BLAST searches are the same as those for ORI and rDNA annotation. The highest-scoring BLAST results for each eccDNA were then assigned to 500 kbp windows of the genome created using BEDTools[187]. Counts of eccDNAs per genome window were visualized with Rideogram[266]. Z-scores were assigned to windows to identify statistically significant regions of eccDNA biogenesis.

5.4 Results

5.4.1 Phenotypic cell culture data

Unadapted, 30 mM-adapted, and 60 mM-adapted cells were grown in 5 day batch cultures to characterize the lactate-adapted phenotype. Cells were grown in media supplemented with lactate0 that corresponds to adaptation levels. Further, 30 mM-adapted and 60 mM-adapted cells were also grown in media with no supplemental lactate. Lactate adaptation had a small effect on growth rate for the 60 mM-adapted cultures with supplemental lactate as the maximum viability was 9.6×10^6 cells/mL by Day 5; however, all other cultures achieved a maximum VCD of 1×10^7 cells/mL (**Figure 5.2A).**

Cell viability was also similar across conditions by the end of the cultures (**Figure 5.2B)**. Glucose was a primary carbon source in all cultures, however lactate-adapted cells without supplemented with lactate utilized more glucose than undapted cells and lactate-adapted cells in media with supplemented lactate. There is no indication lactate is being utilized for gluconeogenesis as there is no observed accumulation of glucose (**Figure 5.2C).** Lactate levels in all cultures increased before cells switched to lactate consumption on Day 3 regardless of lactate adaptation or supplementation (**Figure 5.2D).** The metabolic switch to lactate consumption correlated with glutamine depletion (**Figure 5.2E)**. Alanine increased in all cultures until Day 3. After Day 3, alanine declines in cultures without supplemental lactate, but plateaus in 30 mM-adapted cultures supplemented with lactate and continues increasing in 60 mM-adapted cultures supplemented with lactate (**Figure 5.2F).** The ammonia profiles pair closely in all experimental conditions until Day 3. Lactate-adapted cultures in media with supplemented lactate did not experience the late culture accumulation of ammonia as seen in all other conditions (**Figure 5.2G).** Glutamate concentration paired closely among all experimental conditions throughout the culture. Consumption of glutamate was first observed on Day 3 in line with glutamine depletion and lactate consumption (**Figure 5.2H).** Titer was observed to be significantly higher in unadapted cells. Lactate-adaptation resulted in lower titer with 60 mM-adapted cultures showing lower titer than 30 mM-adapted cultures (**Figure 5.2I).**

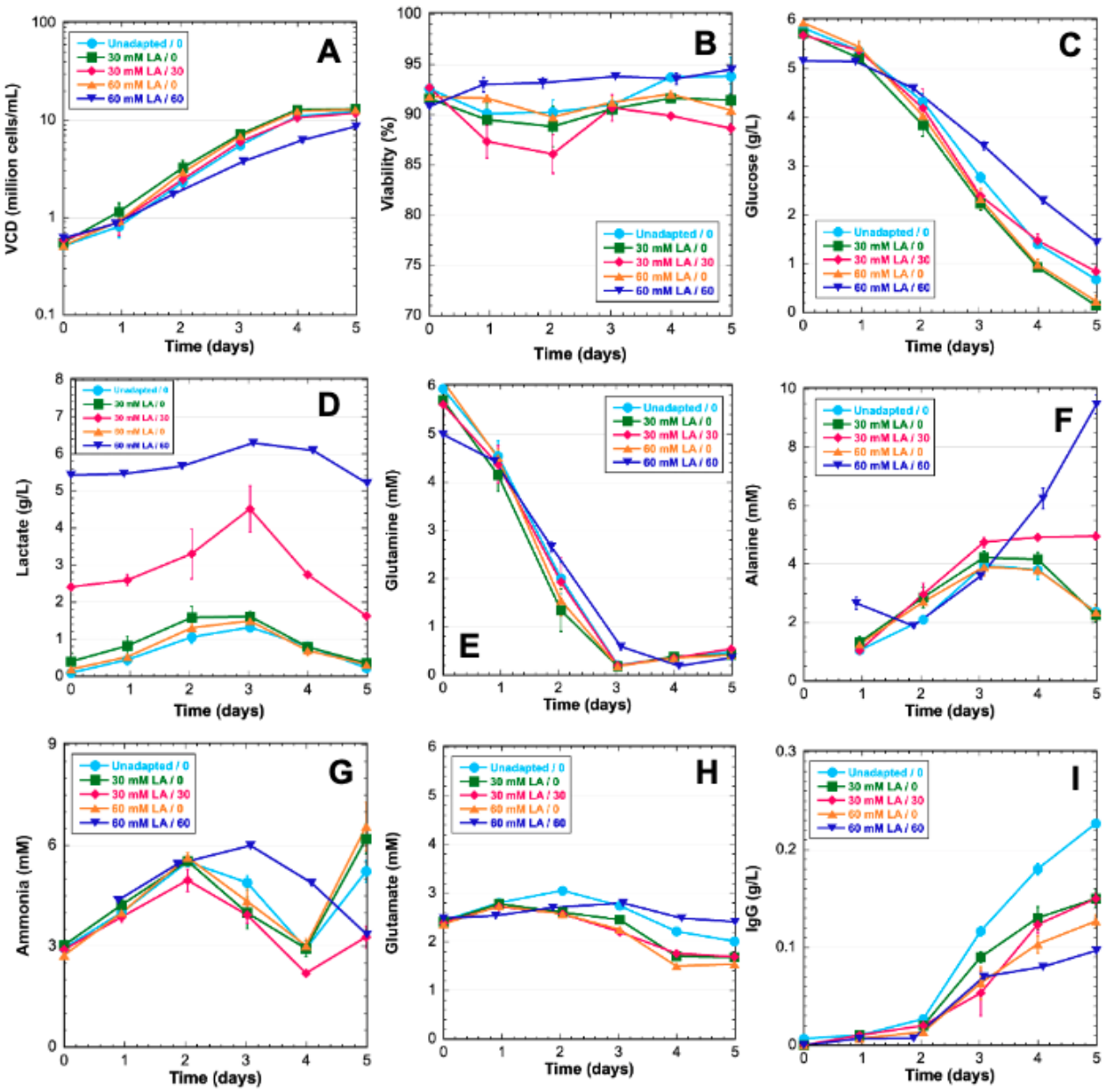


Figure 5.2: Growth characteristics for unadapted and lactate-adapted VRC01 CHO cells cultured in various levels of lactate. A) Viable cell density (VCD) **B)** Viability **C)** Glucose **D)** Lactate **E)** Glutamine **F)** Alanine **G)** Ammonia **H)** Glutamate **I)** IgG. LA – lactate-adapted. Unadapted cells in normal media – light blue, 30 mM-adapted cells in normal media – green, 30 mM-adapted cells in lactate supplemented media – pink, 60 mM-adapted cells in normal media – yellow, 60 mM-adapted cells in lactate supplemented media – dark blue .

5.4.2 EccDNA sequence composition

Cell pellets for eccDNA analysis were collected on Day 3 during the exponential growth phase. The number of observed eccDNAs decreased as lactate adaptation level increased, though this may be due to natural variation as individual library preparations of the same sample can show a high degree of heterogeneity[11]. The average length of observed eccDNAs was similar between the unadapted and 30 mM-adapted, however, the observed average eccDNA sequence length was approximately 500 bp higher for the 60 mM-adapted samples. Many sequence structures were observed to be similarly distributed for the unadapted and lactate-adapted cultures including: percentage of bases pertaining to repeats (~38.0%), GC content (~40.5%), and tRNA genes (~9.3%). Full summaries of tRNA annotation for unadapted, 30 mM-adapted, and 60 mM-adapted eccDNAs are available in **Supplemental Tables S5.3-S5.5**, respectively. Notably, the proportion of eccDNAs encoding genes increased to 5.2% in the 60 mM adapted samples from 4.0% in the 30 mM-adapted samples, a clear overrepresentation of gene content relative to the genome and other eccDNA samples. Origins of replication (ORIs) and ribosomal DNAs (rDNAs) were identified in all three conditions, but sequences with ≥ 95% identity to ORI or rDNA sequences were not substantial (≤6 per condition). Full BLAST output tables for ORI analysis are available in **Supplemental Tables S5.6-S5.8.** Full BLAST output tables for rDNA annotation are available in **Supplemental Tables S5.9-S5.11** for unadapted, 30 mM-adapted, and 60 mM-adapted eccDNAs, respectively. An analysis of repeat and retrotransposable elements revealed that short interspaced nuclear elements (SINEs) showed a gradual decrease in abundance as adaptation progressed, while long interspaced nuclear elements (LINEs) showed a gradual increase

in abundance. However, these changes were within less than a 1% change between unadapted and 60 mM-adapted samples and could be within the range of normal variation. Gene annotation identified 567 genes across the three conditions. EccDNAs were analyzed for genuine coding sequence. Our analysis revealed the presence of 240 genes in unadapted samples, 217 genes in 30 mM-adapted samples, and 174 genes in 60 mM-adapted samples. Notably, minimal overlap of eccDNA genes was observed across lactate adaptation levels, with only 53 genes found to be present in multiple conditions. Homology-based functional annotations for unadapted, 30 mM-adapted, and 60 mM-adapted eccDNAs are available in **Supplemental Tables S5.12-S5.14**, respectively. Full summaries of eccDNA structural annotations are shown in **Table 5.1,** and the distribution of eccDNA-encoded genes is summarized in a Venn diagram in **Figure 5.3**.

Table 5.1: Sequence characteristics of eccDNA sequences observed in CHO cells at varying levels of lactate adaptation. Repeat motif content, gene content, tRNA content, rDNA content, and potential origins of replications are included. Sequences were pooled by lactate adaptation level and clustered for sequence similarity (>90%).

Condition	Sequences	Sequences Clustered	Max Length (bp)	Average Length (bp)	Repeat bases masked	GC (%)	eccDNA with genes	eccDNA with rDNA	eccDNA with tRNA	ORI (>95%)
Unadapted	8,202	8,109	24,436	4,905	15,159,445 bp (37.9%)	40.6%	295 (3.6%)	6	800 (9.9%)	3
30 mM Adapted	6,476	6,419	27,326	4,976	12,132,410 bp (37.9%)	40.5%	255 (4.0%)	4	618 (9.6%)	3
60 mM Adapted	4,563	4,511	24,598	5,515	9,671,540 bp (38.9%)	40.3%	233 (5.2%)	4	484 (10.7%)	4

		Unadapted			30 mM Adapted			60 mM Adapted		
Repeat Structure	**Subcategory**	**Number of Sequences**	**Base pairs (bp)**	**Percent of total bases**	**Number of Sequences**	**Base pairs (bp)**	**Percent of total bases**	**Number of Sequences**	**Base pairs (bp)**	**Percent of total bases**
SINEs:		23,088	2,994,270	7.49	17,928	2,322,980	7.25	13,529	1,750,013	7.03
	Alu/B1	9,915	1,172,210	2.93	7,715	910,743	2.84	6,008	712,601	2.86
	MIRs	1,263	142,664	0.36	972	108,753	0.34	688	77,202	0.31
LINEs:		12,326	5,818,281	14.6	9,816	4,729,537	14.8	7,735	3,856,083	15.5
	LINE1	11,416	5,678,568	14.2	9,140	4,623,778	14.4	7,284	3,781,556	15.2
	LINE2	725	111,589	0.28	564	88,862	0.28	359	62,787	0.25
	L3/CR1	136	20,991	0.05	81	11,098	0.03	68	8,645	0.03
	RTE	45	6,525	0.02	26	5,024	0.02	22	2,683	0.01
LTR elements:		13,196	3,972,926	9.94	10,381	3,175,847	9.91	8,222	2,552,906	10.3
	ERVL	1,223	349,001	0.87	888	267,321	0.83	795	217,908	0.88
	ERVL-MaLRs	5,550	1,498,301	3.75	4,444	1,207,001	3.77	3,434	927,879	3.73
	ERR_class I	1,259	262,964	0.66	972	189,106	0.59	799	172,688	0.69
	ERV_class II	5,043	1,804,631	4.52	3,991	1,464,909	4.57	3,115	1,182,229	4.75
DNA elements:		2,334	431,383	1.08	1,899	358,406	1.12	1,363	250,375	1.01
	hAT-Charlie	1,415	247,910	0.62	1,173	205,506	0.64	832	147,135	0.59
	TcMar-Tigger	598	123,340	0.31	456	98,914	0.31	326	69,009	0.28
Unclassified:		626	244,548	0.61	464	197,627	0.62	389	141,496	0.57
Total Interspaced Repeats:		-	13,461,408	33.7	-	10,784,397	33.7	-	8,550,873	34.4
Small RNA:		581	41,990	0.11	467	34,506	0.11	412	31,344	0.13
Satellites		1,694	727,205	1.82	1,338	579,860	1.81	1,112	491,622	1.98
Simple Repeats		16,679	836,271	2.09	13,098	665,925	2.08	10,325	541,927	2.18
Low Complexity		2,008	105,297	0.26	1,514	77,218	0.24	1,231	62,244	0.25

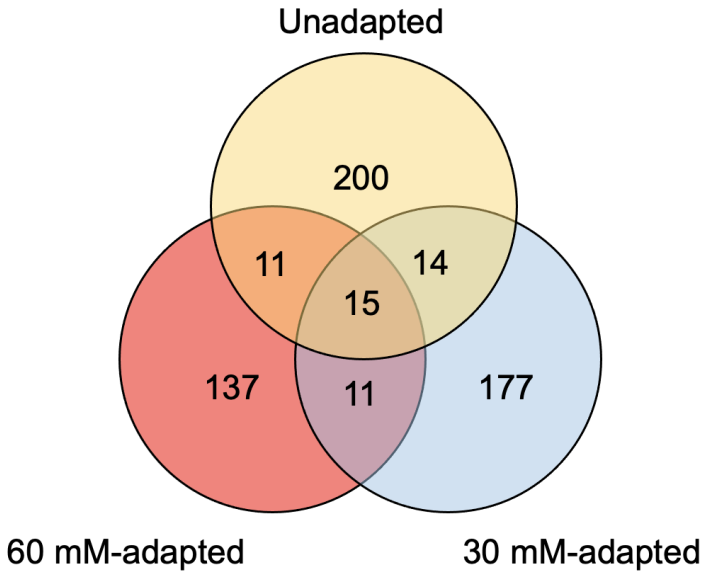


Figure 5.3: EccDNA-encoded gene distributions for CHO cells gradually adapted to higher levels of extracellular lactate. Unadapted (yellow), 30 mM-adapted (blue), and 60 mM-adapted (red).

5.4.3 EccDNA sequence origins

To determine the genomic origins of eccDNA, a non-overlapping window approach was employed. Specifically, the genome was partitioned into consecutive 500 kbp windows, resulting in 4,602 windows for assigning eccDNAs. The mean number of eccDNAs binned per window per condition were 2.50 ± 5.03, 1.97 ± 4.42, and 1.42 ± 3.89 for unadapted, 30 mM-adapted, and 60 mM-adapted, respectively. Windows in each condition with a Z-score ≥ 2 were considered statistically significant regions of high eccDNA biogenesis that may be considered hotspots. Approximately 1% of windows were considered significant across the three conditions with 48, 38, and 43 windows for unadapted, 30 mM-adapted, and 60 mM-adapted, respectively. Analysis of the genomic

regions with the highest frequency of eccDNA biogenesis for each experimental condition revealed notable representation of chromosome 9, spanning from 13 Mbp to 18 Mbp, with four windows ranging from 14 Mbp to 16 Mbp observed in all three conditions ($Z \geq 10$). In the 60 mM-adapted sample, two windows exhibited higher biogenesis frequencies compared to the other samples. One of these windows was located in the telomeric region of chromosome 7, and the other was on chromosome X, containing nine and four genes, respectively, none of which were observed on eccDNAs. Although some variability was observed in other regions, the most prominent region of eccDNA formation remained unaffected by lactate adaptation. The top 15 windows with the highest frequency of eccDNA biogenesis for each condition are listed in Table 5.2. A comprehensive statistical analysis of eccDNA biogenesis mapping for all experimental conditions can be found in Supplemental Table S5.15, while genome-scale maps for each condition are presented in Supplemental Figures S5.1-S5.3.

Table 5.2: The 15 highest frequency eccDNA biogenesis windows observed for each condition. The start and stop positions for each window with the respective chromosome, eccDNA count, and Z-score.

Unadapted					30 mM-adapted					60 mM-adapted				
Chr	Start	End	Value	Z-score	Chr	Start	End	Value	Z-Score	Chr	Start	End	Value	Z-score
9	15000000	15500000	217	42.66	9	15000000	15500000	185	41.41	9	15000000	15500000	141	35.89
9	14500000	15000000	113	21.98	9	14500000	15000000	117	26.03	9	14500000	15000000	98	24.84
9	14000000	14500000	89	17.2	9	14000000	14500000	84	18.56	9	14000000	14500000	85	21.49
9	15500000	16000000	88	17	9	13500000	14000000	76	16.75	9	13500000	14000000	77	19.44
9	13500000	14000000	84	16.21	9	15500000	16000000	61	13.36	9	15500000	16000000	53	13.26
9	13000000	13500000	51	9.64	9	16000000	16500000	55	12	7	134000000	134359064	41	10.18
10	25500000	26000000	47	8.85	9	12500000	13000000	47	10.19	X	106500000	107000000	38	9.41
9	16500000	17000000	36	6.66	9	17500000	18000000	41	8.83	9	12500000	13000000	35	8.64
9	16000000	16500000	35	6.46	9	13000000	13500000	36	7.7	9	16000000	16500000	33	8.12
9	17500000	18000000	35	6.46	1_2	196000000	196500000	34	7.25	2	41000000	41500000	32	7.86
9	25500000	26000000	32	5.87	9	16500000	17000000	29	6.12	7	88000000	88500000	32	7.86
4	131500000	132000000	29	5.27	2	34500000	35000000	21	4.31	9	17500000	18000000	32	7.86
7	134000000	134359064	27	4.87	4	72500000	73000000	20	4.08	2	34500000	35000000	29	7.09
9	17000000	17500000	27	4.87	7	134000000	134359064	19	3.86	4	5000000	5500000	29	7.09
1_2	195000000	195500000	23	4.08	9	25500000	26000000	19	3.86	9	17000000	17500000	28	6.84

5.4.4 Transcriptome analysis

5.4.4.1 EccDNA-encoded genes

It is plausible that genes encoded on eccDNAs may be transcriptionally active if the structures required for transcription are accurately replicated on the eccDNA from the template. To identify shifts that may be attributed to gene gain or loss, transcriptome data for genes encoded on eccDNAs were correlated. Out of 567 identified genes encoded on eccDNAs, 35 exhibited elevated levels of expression that were positively associated with the presence of an eccDNA-encoded copy of the gene. For the purpose of this analysis, an increase in transcript abundance was defined as a 1.2-fold change compared to the other conditions. Notably, these eccDNA-encoded genes were exclusively observed in one of the three experimental conditions.

Out of the 567 identified genes encoded on eccDNAs, 17 were solely observed in unadapted samples and displayed reduced transcript abundance between unadapted and 30 mM-adapted samples. Among these genes were *Pde12*, a phosphodiesterase that regulates mRNA stability in the mitochondrion[267] and functions as an exoribonuclease[268], *Gpatch11*, a nucleic acid binding protein assumed to be present in kinetochores[269], and *Ptbp3*, an RNA binding protein used as a biomarker in lung adenocarcinoma[270] and colorectal cancer[271]. In contrast, six eccDNA-encoded genes exhibited increased transcript abundance in the 30 mM-adapted samples, including *Ndufs8*, a subunit of the NADH:ubiquinone oxidoreductase core that facilitates NADH oxidation and ubiquinone reduction in the electron transport chain[272], and *Exosc4*, a

component of exosomes that has been shown to stimulate cell proliferation and act as an oncogene[273]. In the 60 mM-adapted samples, 12 genes were exclusively detected on an eccDNA and displayed an elevated transcript abundance relative to the 30 mM-adapted samples. *Lin52*, which encodes a protein involved in DNA transcription and is part of the DREAM complex that inhibits cell cycle genes unless the oncogene *Mybl2* is overexpressed[274], was among the identified genes. Notably, the selectable marker for the VRC01 cell line, *Dhfr*^{35}, was observed on eccDNAs in unadapted and 30 mM-adapted samples, although the criteria for a 1.2-fold difference in transcript abundance was not met. The expression of *Dhfr* was comparable between these two conditions (TMM=9.65 and 9.11, respectively) but decreased in the 60 mM-adapted cells, where it was not identified on an eccDNA (TMM=7.76, 1.17-fold difference). Heatmaps of gene expression data that correlate with the presence of eccDNA-encoded genes can be found in **Figure 5.4**, and numerical data used to generate the figure can be located in **Supplemental Table S5.16**. Global gene expression values are provided in **Supplemental Table S5.17**.

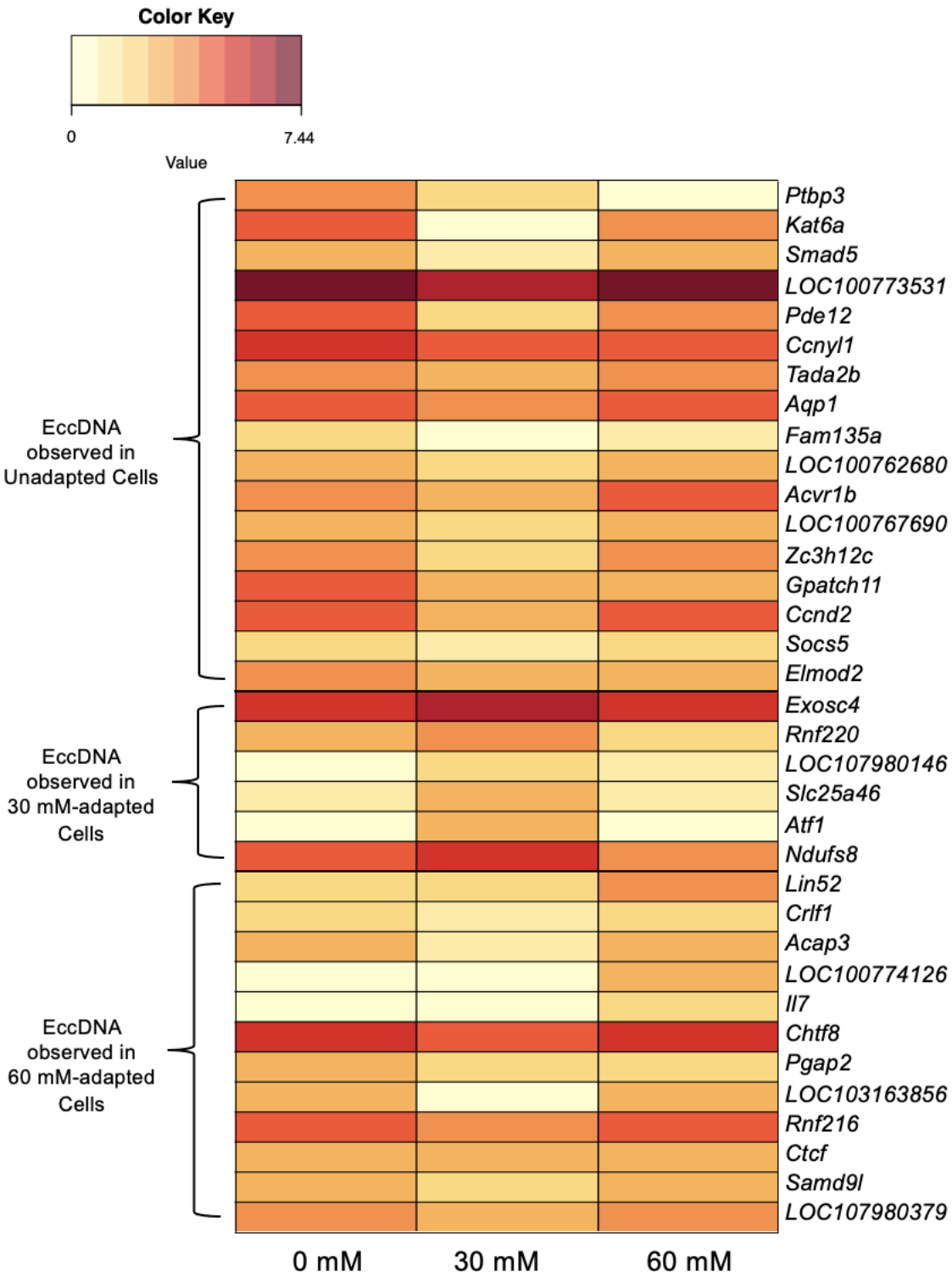


Figure 5.4: Heatmap of transcriptome abundance for eccDNA-encoded genes that show correlated expression with gene presence.

5.4.4.2 Metabolism-linked genes

Although genes related to the lactate-adapted phenotype were not found on eccDNAs, we investigated the expression of genes involved in lactate and alanine metabolic pathways. These genes were categorized into six groups: monocarboxylate transporters (Slc16 family[275]), amino transferases[276], glucose transporters (Slc5 family[277]), alanine and glutamine transporters (Slc6 family[278]), pyruvate metabolism[276], and glutamate transporters (Slc1 family[279]). This simplified pathway is shown in **Figure 5.5.** Transcriptome data showed no substantial up or downregulation in these genes in response to lactate adaptation aside from minor variation. *Slc16a6* and *Ldhd* showed slightly elevated levels of transcript abundance in 60 mM-adapted samples compared to unadapted samples, however, neither gene was determined to be differentially expressed by edgeR. A heatmap of RNA-seq data for lactate and alanine metabolism-associated genes is shown in **Figure 5.6.** Numerical data used to generate **Figure 5.6** is available in **Supplemental Table S5.15.**

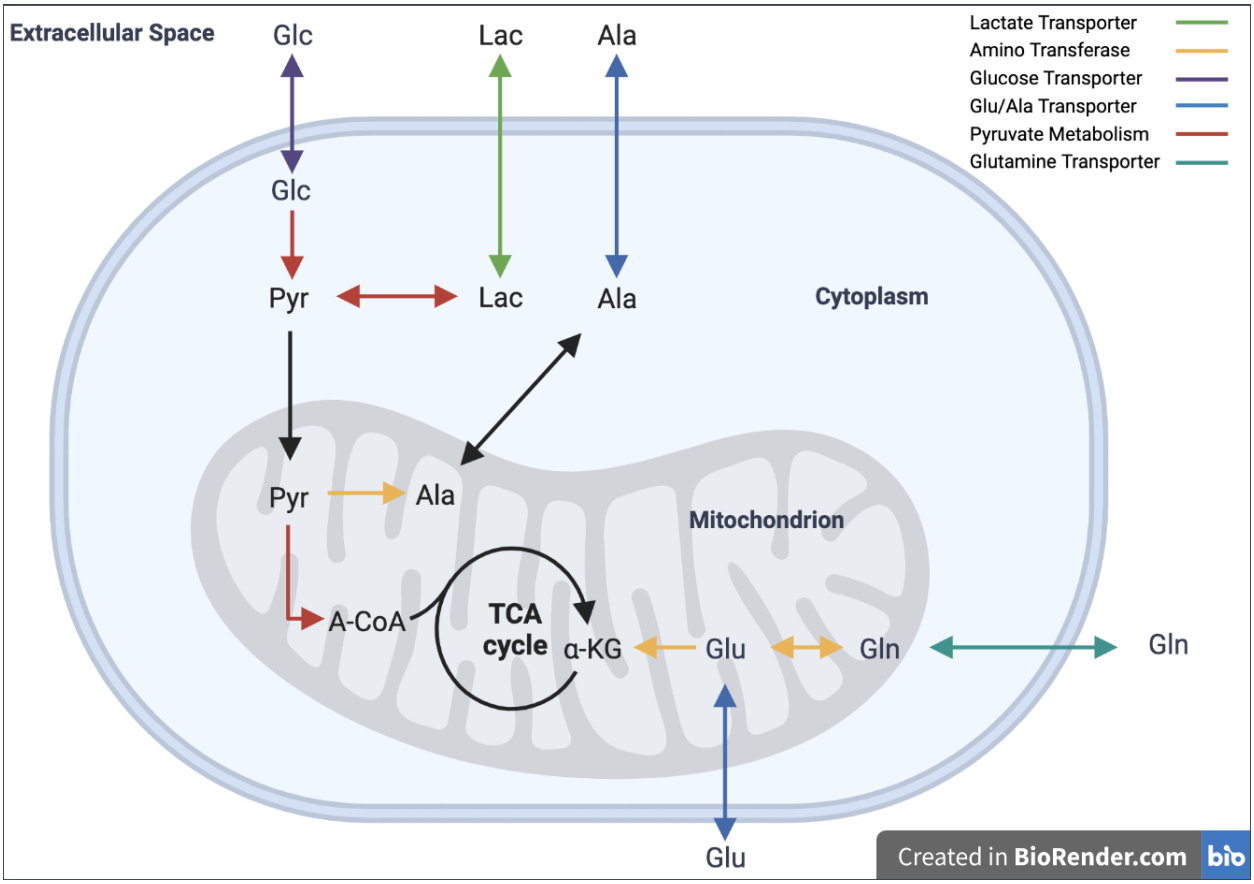


Figure 5.5: Simplified diagram of lactate and alanine metabolism. Arrows are representative of genes and gene families that facilitate reactions and transport. Green – monocarboxylate transporters, yellow – aminotransferases, purple – glucose transporters, blue – glutamate and alanine transporters, red – pyruvate metabolism-associated genes, and teal – glutamine transporters. Black arrows reflect TCA cycle pathways and mitochondrial transport. Genes and gene families depicted in color were examined for differential expression (**Figure 5.6**). Figure generated using BioRender.

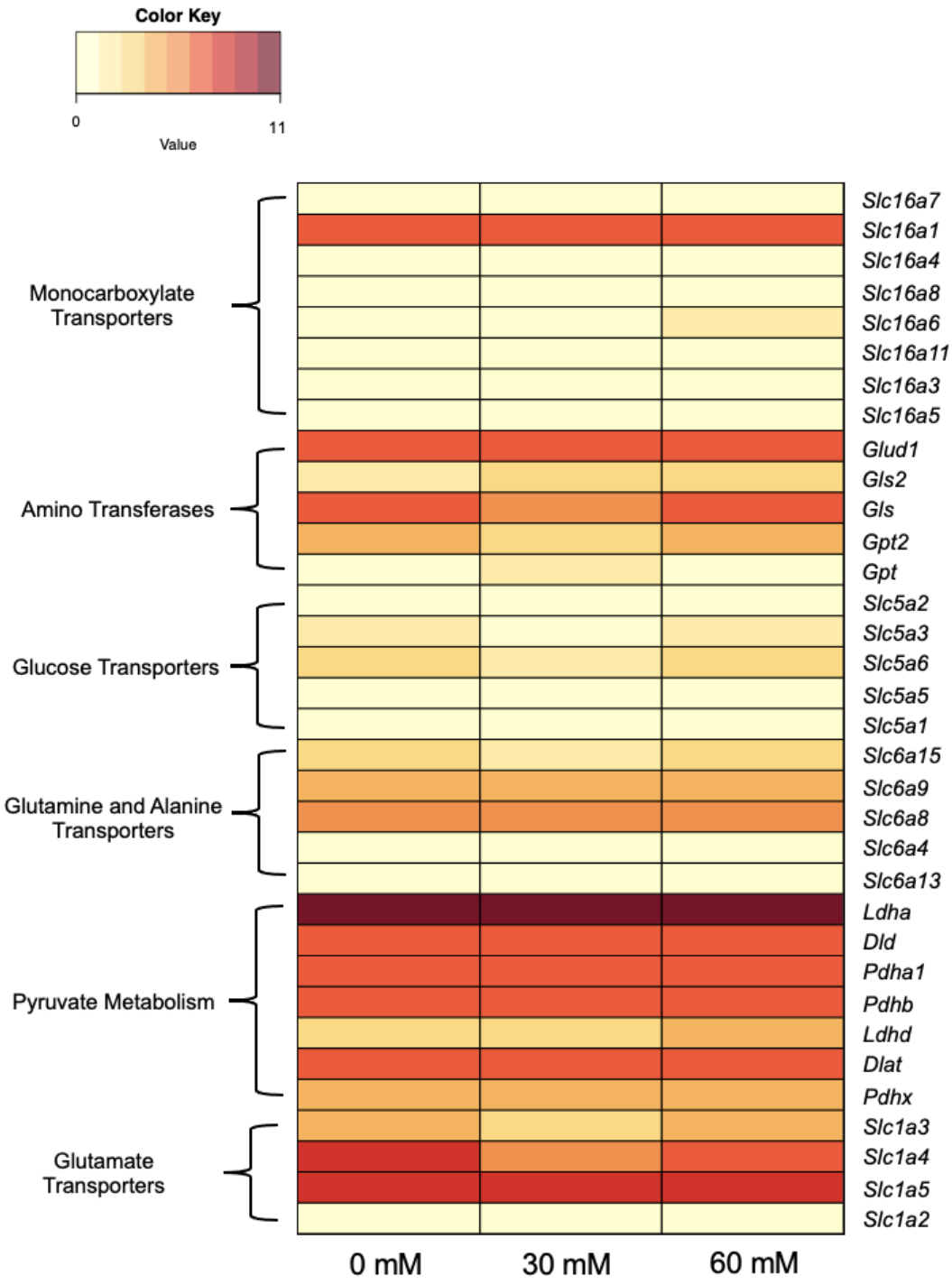


Figure 5.6: Heat map of RNA-seq data for genes facilitating lactate and alanine metabolism. Genes are grouped by function on the left y-axis and gene names are shown on the right y-axis. Lactate adaptation level is shown on the x-axis

5.4.4.3 Differentially expressed genes

Although the expression of genes related to lactate metabolism appears to remain consistent, significant alterations in gene expression (≤ -2 or ≥ 2 $\log_2$ fold change) were detected throughout the adaptation process. Specifically, 30 mM-adapted cells exhibited 762 downregulated and 374 upregulated genes relative to unadapted cells. Comparison between 30 mM-adapted and 60 mM-adapted cells revealed 333 downregulated and 794 upregulated genes for 60 mM-adapted cells. Gene clustering using edgeR identified 1,069 differentially expressed genes for 30 mM-adapted samples, with no net expression changes between unadapted and 60 mM-adapted samples. Among these, 868 genes were expressed at lower levels and 201 were expressed at higher levels in 30 mM-adapted samples. Analysis of the 1,069 genes via KEGG pathway mapping revealed small numbers of genes involved in a broad range of functions across hundreds of pathways. Lists of genes upregulated and downregulated in 30 mM-adapted samples can be found in **Supplemental Tables S5.19** and **S5.20,** respectively. Summaries of identified KEGG pathways for genes differentially expressed for 30 mM-adapted samples can be found in **Supplemental Tables S5.21** and **S5.22** and edgeR clusters are shown in **Supplemental Figure S5.4.**

Upon comparing unadapted cells to 60 mM-adapted cells, only 108 genes and 134 proteins were differentially expressed with 67 proteins downregulated and 67 upregulated. Genes observed to have a net change in expression impacted many biological processes and may have a role in genome stability. *Rras,* a GTPase that regulates angiogenesis and is often overexpressed in cancer[280] was found to have a 3.06-

fold increase in expression levels. Another gene with notable net overexpression was NADPH oxidase 4. *Nox4* catalyzes the production of reactive oxygen species (ROS) and plays an important role in hypoxia signaling. Further, *Nox4* allows for oncogenic adaptation in cancer by facilitating a metabolic shift that is less dependent on aerobic conditions[281]. A critical gene for DNA double-strand break repair, *Brca1*[282], had a 2.71-fold decrease in expression by the end of lactate adaptation. The gene with the greatest net downregulation, *Hnrnpc,* is a regulator of pre-mRNA processing and has been observed to be overexpressed in some cancers; however, reduced expression of *Hnrnpc* has been shown to limit tumor proliferation[283,284]. Genes with a net differential expression change is shown in **Table 5.3.**

Table 5.3: Proteins with significant shifts in net gene expression between unadapted and 60 mM-adapted CHO cells. Log2FC – log_2 fold-change is normalized to the unadapted cells.

RefSeq ID	Gene	log2FC	RefSeq ID	Gene	log2FC	RefSeq ID	Gene	log2FC
XP_027259257.1	*Pnisr*	5.08	XP_027263110.1	*Mms19*	2.3	XP_027262108.1	*C3H16orf87*	-2.43
XP_027271342.1	*LOC107977511*	4.72	XP_027268884.1	*Yeats2*	2.29	XP_035302467.1	*Asip*	-2.43
XP_035300595.1	*Camk2n2*	4.07	XP_027266487.1	*Banp*	2.28	XP_027250346.1	*Oard1*	-2.44
XP_027260962.1	*Usp6nl*	4.01	XP_027272584.1	*Trip10*	2.26	XP_035296751.1	*Prkag1*	-2.45
XP_027243872.1	*Spred2*	3.97	XP_027260324.1	*Kat8*	2.26	XP_027246344.1	*Tinf2*	-2.45
XP_035299495.1	*Vps11*	3.81	XP_027281260.1	*Baiap2*	2.25	XP_035309594.1	*LOC118237753*	-2.54
XP_027270425.1	*Ccdc51*	3.77	XP_027281261.1	*Baiap2*	2.25	XP_035296374.1	*Tmem161b*	-2.62
XP_027265866.1	*Tmem41b*	3.64	XP_027269839.1	*LOC100765617*	2.2	XP_027243712.1	*Mpv17l*	-2.62
XP_035309292.1	*Tdrd3*	3.52	XP_027279239.1	*Zswim3*	2.19	XP_027243713.1	*Mpv17l*	-2.62
XP_027284541.1	*Spr*	3.51	XP_027256563.1	*Spata24*	2.18	XP_035308661.1	*Mpv17l*	-2.62
XP_027274795.1	*Nit1*	3.33	XP_027253398.1	*Asb1*	2.18	XP_027285208.1	*Ino80b*	-2.7
XP_035301857.1	*Nit1*	3.33	XP_035301484.1	*Rbbp5*	2.18	XP_027285383.1	*Tra2a*	-2.71
XP_027258717.1	*Ccn2*	3.32	XP_027264466.1	*LOC100757535*	2.15	XP_027283850.1	*Brca1*	-2.71
XP_027278256.1	*Pex16*	3.16	XP_027277209.2	*Asip*	2.14	XP_027283851.1	*Brca1*	-2.71
XP_027282741.1	*CUNH17orf49*	3.09	XP_035296265.1	*Nhsl1*	2.08	XP_027283852.1	*Brca1*	-2.71
XP_027276511.1	*Rras*	3.06	XP_027258948.1	*Anp32b*	2.06	XP_027283853.1	*Brca1*	-2.71
XP_027268403.1	*Slc35a5*	3.05	XP_027254583.1	*Il7*	2.06	XP_027256348.1	*Rnf138*	-2.71
XP_027263390.1	*Nox4*	3.05	XP_027264703.1	*Mob2*	2.05	XP_027274946.1	*Atxn3*	-2.75
XP_027279804.1	*Coro7*	2.98	XP_027273144.1	*Acbd6*	2.04	XP_035298824.1	*Hivep1*	-2.77
XP_027254088.1	*H6pd*	2.95	XP_027281267.1	*Chmp6*	2.03	XP_027243022.1	*Ctbp1*	-2.82
XP_027268965.1	*Klhl22*	2.95	XP_027266778.1	*Mthfs*	2	XP_035298483.1	*Cpne2*	-2.83
XP_027247336.1	*Agk*	2.94	XP_027266779.1	*Mthfs*	2	XP_027245634.1	*Fam133b*	-2.87
XP_027246345.1	*Tinf2*	2.92	XP_035307155.1	*LOC100774954*	-2.04	XP_027271986.1	*Hip1*	-2.89
XP_027286836.2	*Zc3h4*	2.9	XP_027274041.1	*Trmt5*	-2.05	XP_027252162.1	*Pphln1*	-3.06
XP_027252267.1	*Cacnb3*	2.88	XP_027250449.1	*Nr2c2ap*	-2.07	XP_027287883.1	*Mospd1*	-3.11
XP_035306485.1	*Ptpn12*	2.87	XP_027258438.1	*Poli*	-2.08	XP_027244120.1	*Dnajb12*	-3.13
XP_027278991.1	*Ptpa*	2.86	XP_027274796.1	*Nit1*	-2.13	XP_027244121.1	*Dnajb12*	-3.13
XP_027259074.1	*Stoml2*	2.84	XP_027274797.1	*Nit1*	-2.13	XP_035304186.1	*Top3a*	-3.2
XP_027255528.1	*Id3*	2.83	XP_027274798.1	*Nit1*	-2.13	XP_027271702.1	*Pus1*	-3.2
XP_027275696.1	*Slco4a1*	2.81	XP_027274799.1	*Nit1*	-2.13	XP_027247306.1	*Hipk2*	-3.28
XP_027243830.1	*Commd1*	2.73	XP_027274800.1	*Nit1*	-2.13	XP_027258058.1	*Rtn4ip1*	-3.41
XP_035295692.1	*Ift74*	2.68	XP_027274801.1	*Nit1*	-2.13	XP_027262953.1	*Ldb1*	-3.47
XP_027252455.1	*Eif4b*	2.6	XP_027258090.1	*Pde4b*	-2.15	XP_027250779.1	*LOC103162709*	-3.66
XP_027288617.1	*Mospd2*	2.57	XP_027266034.1	*Eef2k*	-2.16	XP_027255726.1	*Zbtb17*	-3.67
XP_027265269.1	*Zkscan8*	2.56	XP_027288241.1	*Fam3a*	-2.18	XP_027270426.1	*Ccdc51*	-3.73
XP_035294787.1	*LOC113833870*	2.55	XP_035295624.1	*Ndc1*	-2.19	XP_027285126.1	*Atp6v1e1*	-3.87
XP_027282857.1	*Vamp2*	2.55	XP_035307073.1	*LOC100755006*	-2.21	XP_035303760.1	*Vamp2*	-3.91
XP_027283079.1	*Top3a*	2.46	XP_027267297.1	*Pde4a*	-2.21	XP_027266816.1	*Ube3d*	-3.95
XP_027283080.1	*Top3a*	2.46	XP_027288197.1	*Taz*	-2.23	XP_035298723.1	*Usp6nl*	-4.16
XP_027258844.1	*Palm2akap2*	2.42	XP_027265636.1	*Uimc1*	-2.27	XP_027266818.1	*Ube3d*	-4.33
XP_027271581.1	*Pnpla6*	2.4	XP_027269840.1	*LOC100765617*	-2.3	XP_027262827.1	*Smndc1*	-4.34
XP_027255969.1	*Exoc3*	2.4	XP_027248483.1	*Baz2a*	-2.32	XP_027250669.1	*Smim7*	-4.38
XP_027278905.1	*Rapgef1*	2.39	XP_027248484.1	*Baz2a*	-2.32	XP_027257570.1	*Tmem167a*	-5.29
XP_027248513.1	*Ankrd52*	2.34	XP_035306520.1	*Baz2a*	-2.32	XP_027242456.1	*Hnrnpc*	-5.62
XP_027257344.1	*Ppwd1*	2.34	XP_035300748.1	*Khsrp*	-2.38			

5.5 Discussion

Controlling lactate and ammonia concentrations in industrial cell cultures is of great interest to the bioprocessing community as elevated concentrations can limit cell growth and lead to reduced productivity and product quality[11,57,70,285,286]. Numerous process engineering efforts have been made to mitigate waste production such as feeding alternative carbon sources[63,287], using sensors in controlled bioreactors to tune feeding of glucose and/or glutamine[288,289], and varying culture pH[45,290]. Other studies have focused on choosing clones that have higher lactate consumption rates[68,291,292]. Genetic engineering attempts include overexpression of pyruvate carboxylase 2 (*Pyc2*)[293,294], and galactokinase (*Galk1*)[262], as well as knockdown or downregulation of lactate dehydrogenase A (*Ldha*) and/or pyruvate dehydrogenase kinases (*Pdhk*)[295-298]. While there is a large body of literature on controlling lactate levels and adapting cells to grow in lactate, results have been varied with little consensus regarding the most efficient method. Further, the underlying mechanisms of these effects have not been deciphered.

Few studies have addressed the issue by feeding cultures with lactate to shift pH[73] or adapting cells to lactate by media supplementation[72]. Freund et al. is the only published study characterizing a CHO cell line that is adapted to grow in elevated extracellular lactate. Notably, Freund et al. proposed that adaptation was likely a result of mass action rather than a specific metabolic shift that enabled adaptation. Notably, previous work in CHO cells has demonstrated that amination of pyruvate to alanine is a metabolic stress response to excessive ammonia accumulation[58]. Phenotypic data of lactate-adapted cells in this work suggests that excessive lactate is likely mitigated in a similar fashion; by

dehydrogenating lactate to produce pyruvate and transamination with α-ketoglutarate to form alanine. This hypothesis is supported by the reduced levels of ammonia observed in lactate-supplemented cultures of lactate-adapted cells (**Figure 5.2G**) which is consistent with the findings presented in Freund et al.'s lactate-adapted cells[72]. Alanine biosynthesis allows for simultaneous removal of both lactate and ammonia (**Figure 5.5D** and **5.5G**) and is likely a stress response mechanism for metabolic waste accumulation in CHO cells.

Overall, eccDNA structure does not show significant signs of change with lactate adaptation (**Table 5.1**). Small shifts were observed, such as LINE and SINE elements increasing and decreasing, however the shifts are so small they could easily be attributed to natural variation or variation from sample preparation and/or sequencing. The most substantial difference observed with respect to lactate adaptation was the increase in average eccDNA sequence length combined with an increase in the proportion of eccDNAs harboring genes. An increase in average sequence length could be indicative of recombination between smaller eccDNAs[119], however an in-depth, sequence-by-sequence structural analysis would be required to confirm that hypothesis. The proportions of repeat content and distribution, GC content, tRNA, and gene content exhibit consistency with minimal variation noted between lactate-adapted eccDNAs and the three conditions previously observed in VRC01 eccDNAs[11]. Collectively, these data suggest that the composition of eccDNA sequence structures in VRC01 are minimally impacted by stress adaptation and fed-batch culturing.

Identification of eccDNA biogenesis sites indicated a genomic hotspot on chromosome 9, specifically between 12.5 Mbp and 18 Mbp in all levels of lactate adaptation (**Table 5.2** and **Supplemental Figures S5.1-S5.3**). One genomic region on chromosome X exhibited a high biogenesis frequency only in the 60 mM-adapted samples, ranking among the top 15 regions with highest frequencies. This region, which has been previously identified as a high-frequency window in other studies, was otherwise consistent in VRC01[11]. Previous studies have demonstrated that eccDNA biogenesis frequencies can significantly shift in response to stress adaptation[119]; however, in this study, no significant changes in biogenesis frequencies were observed in response to lactate adaptation. The observed increase in frequencies from windows on chromosomes 7 and X in 60 mM-adapted eccDNAs cannot be attributed to transcriptome-driven biogenesis, as none of the genes located in those regions were found on an eccDNA. Notably, the chromosome 7 window with high frequencies is located at the telomeric region of the chromosome and may indicate decreasing chromosome stability via the production of t-circles[227,245].

EccDNAs have been shown to mediate stress response and adaptation via gene overexpression across species in response to many stimuli[118,119,138,151,152,223]. Further, some evidence has been shown that eccDNAs may play a minor role in modifying gene expression patterns in VRC01 cells grown in lactate-stressed conditions by amplifying *Akr1b1*[11]. It is important to note that in the previous study, genes that correlated with eccDNA gene presence had a more substantial difference in transcript abundance (≤ -2 or ≥ 2 $\log_2$ fold change). Initially, employing the same threshold to this dataset showed no

correlation, which resulted in the relaxed criteria utilized to identify correlation. Further, genes with a low level of expression (TMM<2 in all three conditions) were excluded to limit false positives. Correlating transcript abundance with eccDNA gene content and establishing a 1.2x difference in TMM values identified 35 eccDNA-encoded genes that may be transcriptionally active.

RNA-seq data shown in this study suggests that certain changes in gene expression may be attributed to eccDNA, however none of the eccDNA-encoded genes with correlated gene expression changes do not appear to respond to stress. In the 30 mM-adapted samples, *Ndufs8*, which is encoded on an eccDNA, is the only gene related to redox balancing that may have an impact on lactate-adaptation, but this gene was not found on an eccDNA in the 60 mM-adapted samples. Two eccDNA-encoded genes that correlated with gene expression changes exhibit potential to enhance genome stability. *Lin52* is a subunit of the DREAM complex, which is a crucial regulator of gene expression shifts that are dependent on cell cycle progression[299]. Overabundance of DREAM complex components could result in tighter control of cell cycle genes that regulate growth and DNA synthesis. Downregulation of *Ptbp3* is also a positive sign for improving genome stability as overexpression is a highly reliable biomarker for cancers[270,271,300,301]. While overexpression of *Ptbp3* is correlated with many cancer phenotypes, knockdown or silencing of *Ptbp3* has been shown to induce cell cycle arrest and apoptosis[302].

Key metabolism and transporter genes were examined to identify shifts in gene expression that may accommodate the lactate-adapted phenotype. Despite adaptation of

cells to levels of extracellular lactate far beyond what is normally tolerable in CHO cell cultures, no evidence was found that indicates a transcriptome-level response. Freund et al. proposed that lactate adaption occurred by mass action; excess lactate in culture would make conversion to pyruvate more thermodynamically favorable. While the lactate consumption phenotype is likely due to substrate-level control, lactate accumulated in all cultures until Day 3 regardless of lactate adaptation and/or supplementation which would not agree with Freund's proposed mechanism.

Significant changes in gene expression were observed between 60 mM-adapted and 30 mM-adapted samples, although lactate metabolism genes were not significantly modulated. In contrast, no such changes were observed between unadapted samples and 60 mM-adapted samples. Eukaryotic gene expression is regulated by complex signaling cascades[183]. Changes in expression are effected greatly by external signaling pathways, transcription factors and can be short or long-term[303]. The large shifts in gene expression with regard to 30 mM-adapted cells was unexpected; however, when comparing unadapted to 60 mM-adapted cells, some of the expression changes seen in 30 mM-adapted cells are likely attributed to gene expression tuning in response to the lactate stress[304]. Responding to a temporary stress can cause shifts in complex gene regulation pathways; adaptation to chronic stresses can cause drastic rebalancing of regulatory cascades[303]. In yeasts, stochastic switching has been observed as a survival tactic in response to environmental stresses[305]. The observed random gene expression shifts in 30 mM-adapted cells could be attributed to the rapid pace of adaptation to lactate. Stochastic expression in response to stress has also been observed in Arabidopsis. For

instance, when exposed to heat stress, more than 16,000 genes exhibited differential expression, with only 43 conferring a selective advantage for adaptation[306].

Examining net differential gene expression between unadapted and 60 mM-adapted samples identified 134 genes with a net change in transcript abundance by the end of the adaptation process. The increased expression of *Nox4* is likely the closest indicator of a gene expression shift mediating the lactate adaptation process as *Nox4* has been shown to encourage tumor growth in anaerobic environments[281]. Also notable is the downregulation of four splice variants of *Brca1*, a critical DNA repair gene. Maintenance of *Brca1* is critical as excess stress can result in unintentional genomic rearrangements[27]. Further, differential expression of *Rras* and *Hnrnpc* in opposite patterns is also unusual, as overexpression of these genes typically encourages proliferation. The data presented in this study provides a single point in time snapshot of a highly dynamic lactate adaptation process, and although gene expression may become more finely tuned at higher levels of adaptation, it is probable that stochastic switching may still play a role. To identify precise changes in gene expression in response to lactate adaptation, time-series data from extended cultures of lactate-adapted cells would be necessary.

5.6 Conclusions

This study characterizes phenotypic behavior, circulome dynamics, and gene expression shifts in CHO cells gradually adapted to grow in extreme levels of extracellular lactate. The generation of alanine in lactate-adapted cultures that are supplemented with lactate suggests that alanine biosynthesis may be an efficient mitigation mechanism of

both lactate and ammonia in CHO cells. EccDNA content in the VRC01 cell line remains highly heterogeneous and dynamic with no clear evidence of contributing to overall differential gene expression in a substantial way. Notably, eccDNAs were observed to be longer and less abundant as lactate adaptation progressed, which may be indicative of eccDNA recombination. The underlying mechanism of lactate adaptation in CHO cells remains unclear, though expression of lactate metabolism genes do not appear to change in response to lactate stress. Finally, 134 genes were found to have a net change in expression through the adaptation process, though more than 1,000 genes were observed to be differentially expressed between unadapted, 30 mM-adapted, and 60 mM-adapted cultures. This data could suggest stochastic switching of gene expression as a stress adaptation mechanism in CHO cells. Further study of these gene expression profiles in these cells grown in extended culture is needed to fully assess the contribution to the lactate-adaptated phenotype.

CHAPTER SIX

CHARACTERIZATION OF METABOLIC RESPONSES, GENETIC VARIATIONS, AND MICROSATELLITE INSTABILITY IN AMMONIA-STRESSED CHO CELLS GROWN IN FED-BATCH CULTURES

6.1 Abstract

As bioprocess intensification has increased over the last 30 years, yields from mammalian cell processes have increased from 10's of milligrams to over 10's of grams per liter. Most of these gains in productivity can be attributed to increasing cell densities within bioreactors. As such, strategies have been developed to minimize accumulation of metabolic wastes, such as lactate and ammonia. Unfortunately, neither cell growth nor biopharmaceutical production can occur without some waste metabolite accumulation. Inevitably, metabolic waste accumulation leads to decline and termination of the culture. While it is understood that the accumulation of these unwanted compounds imparts a suboptimal culture environment, little is known about the genotoxic properties of these compounds that may lead to global genome instability. In this study, we examined the effects of high and moderate extracellular ammonia on the physiology and genomic integrity of Chinese hamster ovary (CHO) cells. Through whole genome sequencing, we discovered 2,394 variant sites within functional genes comprised of both single nucleotide polymorphisms and insertion/deletion mutations as a result of ammonia stress with high or moderate impact on functional genes. Furthermore, several of these *de novo* mutations were found in genes whose functions are to maintain genome stability, such as *Tp53,*

Tnfsf11, Brca1, as well as *Nfkb1.* Furthermore, we characterized microsatellite content of the cultures using the CriGri-PICR Chinese hamster genome assembly and discovered an abundance of microsatellite loci that are not replicated faithfully in the ammonia-stressed cultures. Unfaithful replication of these loci is a signature of microsatellite instability. With rigorous filtering, we found 124 candidate microsatellite loci that may be suitable for further investigation to determine whether these loci may be reliable biomarkers to predict genome instability in CHO cultures. This study advances our knowledge with regards to the effects of ammonia accumulation on CHO cell culture performance by identifying ammonia-sensitive genes linked to genome stability and lays the foundation for the development of a new diagnostic tool for assessing genome stability.

6.2 Introduction

Biopharmaceutical manufacturing represents nearly 2% of the total US GDP[307] which makes it an important driver of the US economy. Biopharmaceuticals include monoclonal antibodies, recombinant proteins, and assemblies of proteins produced by biological means. Commercial products are used as blood factors, thrombolytic agents, therapeutics, growth factors, interferons, and vaccines [308,309]. The most common mammalian cell line used is the Chinese hamster ovary (CHO) cell line, due to its ability to produce biopharmaceutical molecules with post-translational modifications required in humans [310]. However, it is well understood that recombinant CHO cell lines are susceptible to genome instability that is often observed after approximately 70 generations [2,37,311,312]. Previous studies have characterized genomic variants across

various CHO cell lines that may be a contributing factor to genome instability [213,313,314]. An unstable genome can result in reduced productivity in continuous cultures and fed-batch systems [30,315]. A common occurrence in both continuous cultures and fed-batch systems is the accumulation of metabolic waste products, such as ammonia and lactate. The role these waste products play in cellular processes, such as glycosylation, metabolism, and productivity have been characterized [61,316-318]; however, the effects of these waste products on genome stability have not been directly assessed.

Microsatellite instability (MSI) is described as genetic hypermutability at microsatellite loci where a high frequency of insertion or deletion (indel) mutations accumulate in daughter cells during cell division [97,235]. MSI results from improperly functioning mismatch repair (MMR) pathways which are key to maintaining genome stability [112]. Rather than correcting DNA mismatch errors that occur spontaneously during DNA replication, cells with impaired MMR systems accumulate these errors over the course of subsequent propagation. The prevalence of these errors allow for MSI loci to be utilized as stable genetic biomarkers that are capable of diagnosing many human cancers [111,319]. Studies have shown that approximately 15% of human patients with colorectal cancer [110,112], 20% of patients with stomach cancer [113], and 30% of patients with endometrial cancer [114] could attribute their disease to genome instability that can be diagnosed with MSI biomarkers. The clinical uptake of MSI-based diagnostics, such as the Bethesda Panel, demonstrates the reliability and clinical utility of MSI loci as biomarkers [106].

In this study, we investigated the effects of exogenous ammonia exposure on genome stability during fed-batch cultures of CHO cells. Specifically, the accumulation of DNA mutations in cells exposed to elevated ammonia were compared to cultures grown under standard fed-batch conditions. Ammonia was added to duplicate parallel cultures at 10 mM and 30 mM final concentrations at 12 hours of culture time to establish mild and high ammonia stresses respectively. After 72 hours of elevated ammonia exposure, samples were taken for whole genome sequencing (WGS). These sequences were then analyzed for MSI, single nucleotide polymorphisms (SNPs), and insertion/deletion (indel) variations. The SNPs and indels were mapped to the Chinese hamster genome and assessed for functional impact in both coding and regulatory genetic regions. Microsatellite regions were analyzed to identify loci with dose-dependent indel mutations that could be used as potential biomarkers.

6.3 Materials and Methods

6.3.1 Culture conditions

A recombinant CHO-K1 Clone A11 from the Vaccine Research Center at the National Institutes of Health (NIH), which expresses the anti-HIV antibody VRC01 (IgG_1) was used. The inoculum train was expanded in 250 mL shake flasks with 70 mL ActiPro media (GE Healthcare) that were maintained at 5% CO_2 and 37°C. The bioreactors were ambr® 250 bioreactors (Sartorius Stedim, Göttingen, Germany) with two pitched blade impellers and an open pipe sparger (vessel part number: 001-5G25). The bioreactors were inoculated at a target cell density of 0.4×10^6 cells/mL in ActiPro batch media and fed daily beginning on Day 3 (3% (v/v) Boost 7a and 0.3% (v/v) Boost 7b (GE

Healthcare)). Duplicate cultures were ammonia-stressed at 12-h post inoculation with 0 mM, 10 mM, or 30 mM NH_4Cl. The 0 mM and 10 mM cultures used saline to normalize the volume of the 0 mM and 10 mM cultures to the 30 mM cultures. Dissolved oxygen was controlled at 50% of air saturation using PID control that increased the O_2 mixture in the gas sparge to 100%, then the stir speed from 300 to 600 rpm. Antifoam (10% solution in media; SH30897.41 – GE Healthcare) was added as needed to control foaming. All gases were supplied through the open pipe sparger; an overlay was not used. The pH was controlled via sparging CO_2 and air, and base pump (1 M NaOH). The pH setpoint was 7.0 with a 0.2 deadband. Temperature was controlled at 37°C. Samples for WGS and MSI analysis were harvested at 84 hours culture time (72 hours post-stress) and centrifuged at approximately 2,000 x g for 15 minutes at 4°C. The supernatant was removed, and the pellet was stored at -80°C.

6.3.2 DNA extraction, whole genome sequencing, and microsatellite variant discovery

Pellets of approximately 0.5 x 10^6 cells were pre-washed with 1X phosphate buffered saline (PBS) prior to extraction. Total genomic DNA (gDNA) was purified from 2 replicate samples per condition with the DNAeasy Blood and Tissue Kit (Qiagen), following the manufacturer's recommended procedures and combined prior to sequencing. Whole genome shotgun sequencing was performed on an Illumina NovaSeq (2x150 paired end) through a third- party vendor to approximately 30x genome coverage. Raw sequence data was assessed for quality with the FASTQC software (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Raw sequence data was

preprocessed to remove low quality bases and adapter sequences with the Trimmomatic software v.0.38 [194]. Preprocessed reads were aligned to the CriGri-PICR version assembly of the Chinese hamster genome (*Cricetulus griseus*) (RefSeq assembly accession: GCF_003668045.1) with the Bowtie2 v.2.3.4.1 short read aligner [195]. Alignments were coordinate sorted and indexed with SamTools v1.3.1 [320]. SNPs and indels were determined with the HaploTypeCaller Walker from the Genome Analysis Toolkit (GATK v.4.0) [321]. Genetic variants in the resulting VCF file (SNP and INDEL) were hard filtered according to the following criteria: read depth (DP 6) and mapping quality (MQ 30). Variant sites were further filtered to remove variant loci that were common between the sample groups but differed from the CHO PICR reference assembly. Finally, variant sites were kept in the final VCF file only if one or both of the treatment samples differed from the control. Functional SNPs were characterized with the SnpEff software, v4.3 [322]. Genome-wide microsatellite loci were determined against the PICR CH assembly with MISA, a microsatellite finder software [323]. Microsatellite loci were intersected with indel coordinates using BedTools Intersect command 2.27.1 [187] to identify indel variants associated with microsatellites.

6.3.3 Text/data mining and functional enrichment analysis

The query “genomic instability [MeSH Terms]“ was used to search PubMed to retrieve the abstracts with PMIDs (14,968 PMIDs). The PubTator [324] tool was used to collect genes annotated in these abstracts with Entrez Gene IDs (ftp://ftp.ncbi.nlm.nih.gov/pub/lu/PubTator/gene2pubtator.gz released 2/14/2020). Among the 5,073 retrieved genes, 3,131 genes were human (*Homo sapiens*), 882 mouse

(*Mus musculus*), 435 yeast (*Saccharomyces cerevisiae*) representing the three top species with the largest number of genes mapped in those PubMed abstracts. The other top mapped species include seven vertebrates: rat, chicken, zebrafish, frog, Chinese hamster, dog and pig as well as three non-vertebrates: fly, *Arabidopsis*, and worm. The ortholog pairs between human and the eight other above vertebrates were mapped with NCBI ortholog assignment (ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/gene_orthologs.gz released 07/20/2020), whereas those between human and the four non-vertebrate species were mapped with OMAbrowser (https://omabrowser.org/oma/genomePW/). Altogether, 2,897 human genes linked to "genomic instability" were matched with corresponding Chinese hamster orthologs. For the SNP list with high and moderate impact mutations by SnpEff, 273 Chinese hamster genes were mapped with their corresponding human orthologs that had been linked to "genomic instability". ClusterProfiler [192,325] was used to obtain the enriched KEGG pathways and GO annotations for the given gene lists.

6.3.4 Identification of candidate MSI loci

Candidate MSI loci were determined with a filtering strategy that leverages several criteria as follows: First, each novel indel-variable genomic locus was assigned a mutation score which is a proportion of the number of variant reads (allelic depth) by the total depth of reads for each site extracted from the .vcf file. Second, the mutation scores of the control cultures were subtracted from the mutation scores of the treated loci in order to generate a mutation score relative to the control. This allowed for the removal of loci that did not exhibit dose-dependent responses to the exogenous ammonia. Concurrently, loci

with nonpositive relative mutation scores were also removed. The remaining loci were then intersected with the genome-wide microsatellite coordinates determined by MISA with the Intersect command of BedTools v2.27.1 [187] to identify loci within known microsatellites. The final ranked set of candidate MSI loci contain sites where control samples have fewer to no variant reads in comparison to the treated samples.

6.4 Results

6.4.1 Growth and metabolite profiles

Recombinant CHO cells expressing the monoclonal antibody VRC01 were cultured in tightly controlled ambr® 250 bioreactors for 12 hours prior to the addition of ammonia to stress the cultures. Up to 1.5 days post-inoculation, there were no observable differences in the viable cell densities (VCD); however, at 2.5 days, the 30 mM ammonia-stressed cultures had substantially lower VCDs compared to the control and 10 mM stressed cultures (**Figure 6.1A**). The 10 mM ammonia-stressed cultures had similar VCDs to the control cultures until Day 7; yet cell viabilities were similar to the control cultures for the entire culture durations. In contrast, the 30 mM ammonia-stressed cultures reached peak VCDs on Day 4 and gradually declined until the cultures were harvested on Day 8.5 due to low viability (< 70%); a cell viability below 70% is a standard harvesting threshold. Samples for genome sequencing were taken 84 hours post inoculation (Day 3.5), i.e., 72 hours post-stress. At the time of harvest of genome sequencing samples, the viability for all samples was greater than 90% (**Figure 6.1A**). The mildly stressed (10 mM) cultures had no significant change in the ammonia levels between 12 and 84-h, while the high stress (30 mM) cultures had a gradual decline in the

ammonia concentration until harvested (**Figure 6.1B**). The glucose and lactate profiles (**Figure 6.1D,E**) confirm that the control and 10 mM ammonia-stressed cultures were paired closely throughout the entire cultures, although the 10 mM stressed culture had slightly lower VCD beginning on Day 6. In contrast, the 30 mM ammonia-stressed cultures began to accumulate glucose and lactate after Day 5, most likely due to the set feeding protocol based on culture volume, and the significantly lower cell growth (**Figure 6.1A**). It is well-known that excessive glucose inevitably leads to lactate accumulation [61], which was observed for the 30 mM stressed cultures. Amino acid profiles were also obtained for these cultures [326]. The amino acid profile that showed the greatest differences between the control and 10 mM cultures was alanine; both the 10 mM and 30 mM ammonia-stressed culture alanine profiles were very similar through Day 6, while the control cultures had profiles that represented a higher consumption rate, as alanine was fed starting on Day 3 (**Figure 6.1C**). The glutamine profiles for the control and 10 mM cultures were similar up to Day 7, until the 10 mM cultures began to accumulate glutamine (**Figure 6.1G**). The glutamine accumulation can be attributed to the feeding of glutamate (**Figure 6.1H**), which when in excess can be aminated to form glutamine [327]. The 30 mM cultures were terminated prior to any significant differences in the glutamine accumulation being observed between the control and 30 mM cultures. Therefore, the glutamate feeding, based on volume, caused glutamate to accumulate due to lower VCD relative to the control cultures for the 10 mM and 30 mM cultures. This in turn impacted the glutamine profile. A global measure of cell health is the overall protein production and cell-specific productivity. The monoclonal antibody titer at the end of the cultures was about 50% lower for the 10 mM cultures compared to the control cultures, whereas the 30 mM cultures had

negligible protein productivity in (**Figure 6.1F**). Furthermore, cell-specific productivity (picograms of IgG per cell per day) was found to be substantially higher in the control cultures when compared to the 10 mM sample. It should also be noted that the control and 10 mM cultures had a relatively stable production rate, whereas the 30 mM cultures declined (**Figure 6.1I**). Overall, the samples for the genome sequencing analysis were taken at culture times when there were no substantial VCD, viability, or metabolic differences between the control and 10 mM ammonia-stressed cultures; however, the VCD was significantly lower for the 30 mM ammonia-stressed cultures (**Figure 6.1**).

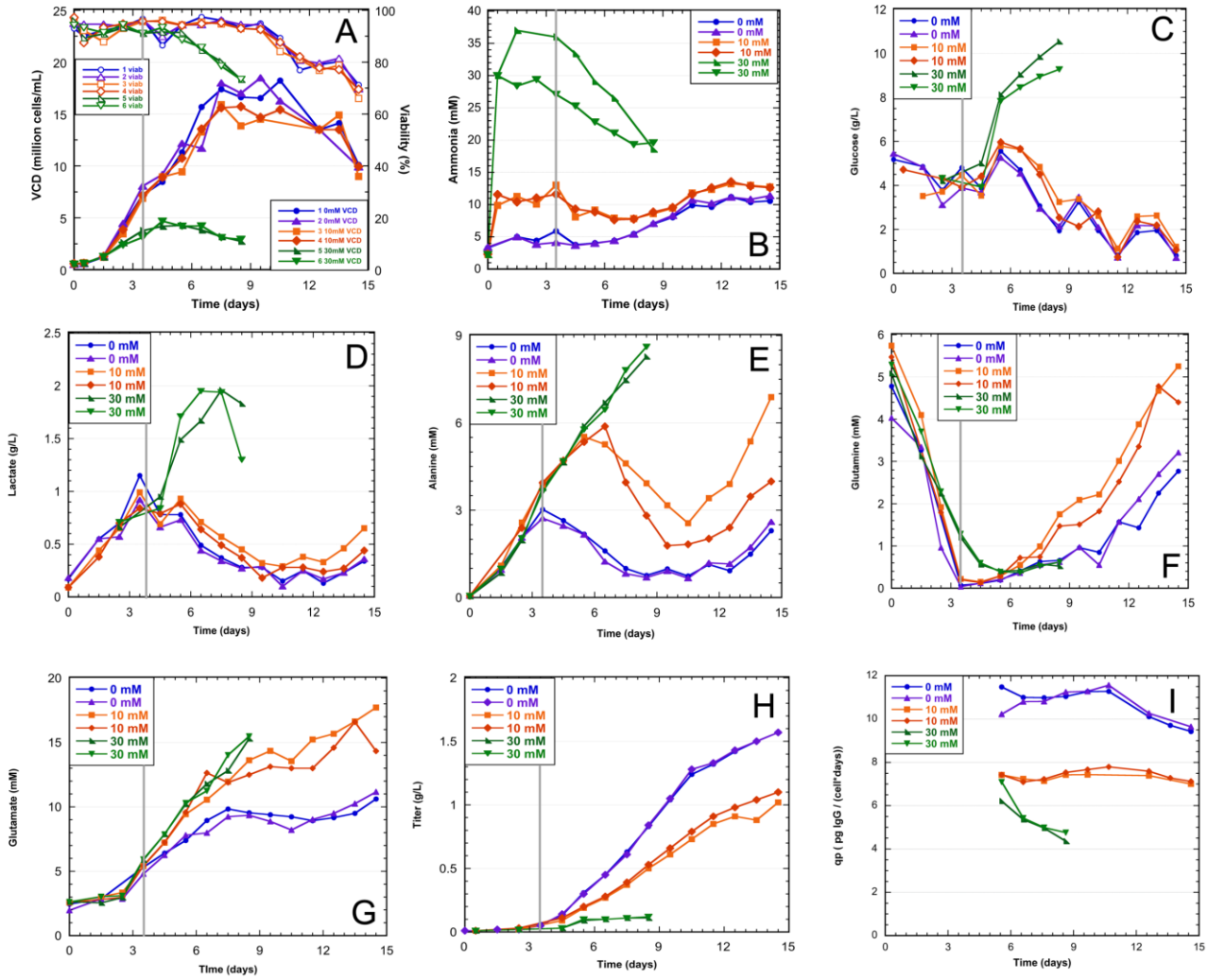


Figure 6.1: **Cell growth, ammonia, titer and metabolic profiles for CHO K-1 VRC01 cells cultured in duplicate in the ambr® 250 bioreactor.** The ammonia stresses (10 mM and 30 mM) were added at 12 hours. Samples for genomic analysis were harvested at 84 hours (3.5 days) as shown by the solid grey line. **A)** Viable cell density (VCD) and viability (filled and hollow symbols, respectively), **B)** ammonia, **C)** alanine, **D)** glucose, **E)** lactate, **F)** titer of recombinant monoclonal antibody, **G)** glutamine, **H)** glutamate, and **I)** Cell-specific productivity (qp). Due to low levels of the recombinant protein in culture prior to day 5, the qp value is not shown until a significant titer has been reached. In industry, it is common to only measure titers starting at day 7. Control - 0 mM (blue and purple lines); 10 mM (orange and red lines); 30 mM (green and dark green lines).

6.4.2 Whole genome shotgun sequencing and variant discovery in stressed conditions

Whole genome shotgun sequences were collected for the control and treated samples to an approximate depth of 30X coverage to assess the genomic impact of exogenous ammonia exposure. A total of 389,694 variant sites were identified across both stress levels that were composed of 310,597 SNPs and 79,097 indels, (**Supplemental Tables S6.1** and **S6.2** respectively). Of the 389,694 variant sites, a total of 135,913 variant sites reside in protein coding genes (**Supplemental Table S6.3**). The variant sites were seemingly randomly distributed in both intergenic and genic positions across the genome. A distribution and density map of variant positions relative to annotated coding genes is depicted in **Figure 6.2**. These variants were further filtered to remove sites annotated as a modifier or low impact variant (e.g. synonymous mutations) predicted by SnpEff. This led to the discovery of 2,394 variants within protein coding genes with significant impact variations (high/moderate impact predicted by SnpEff) due to ammonia stress (**Supplemental Table S6.4**)

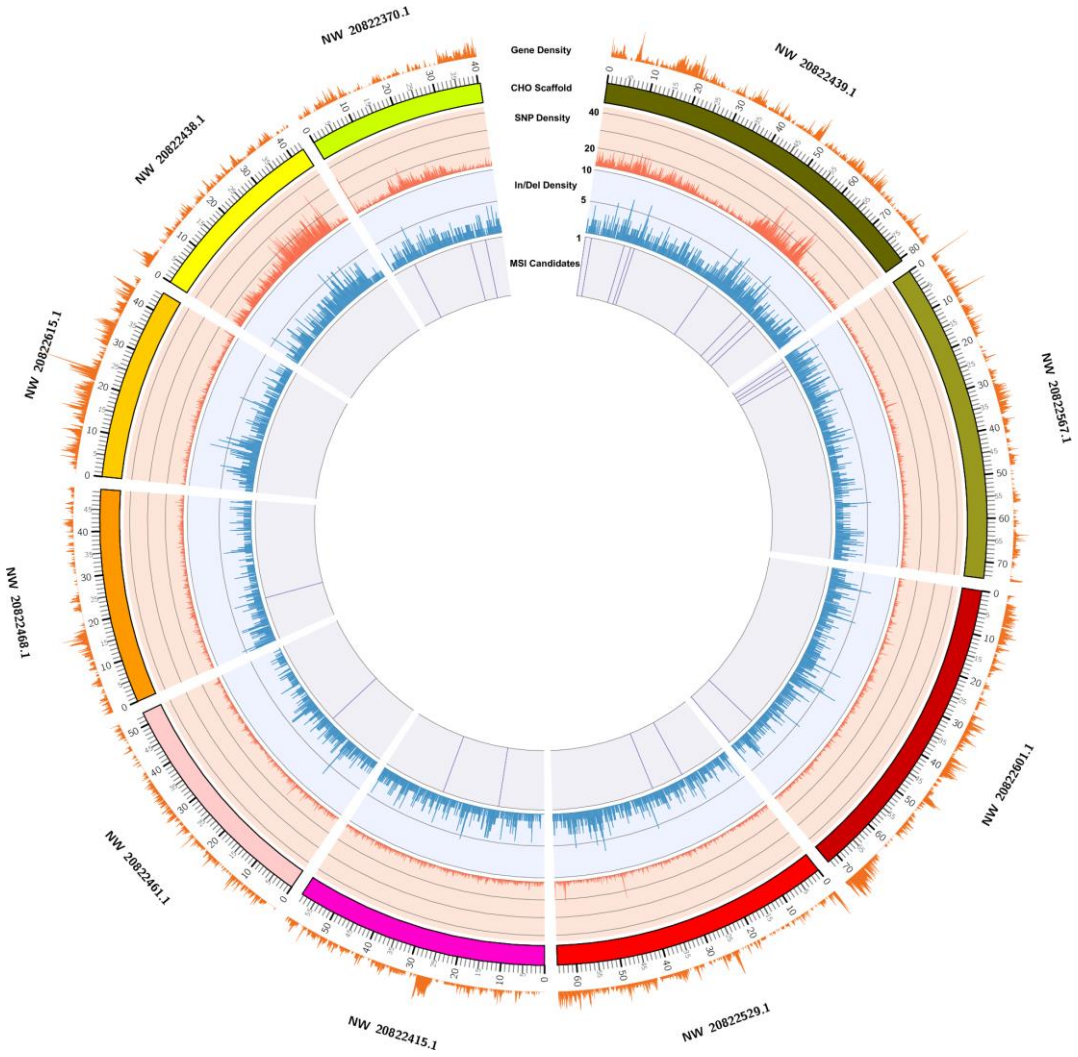


Figure 6.2: Genome coverage map of genetic variants and MSI loci in the 10 longest CHO scaffolds. A circos plot of the 10 longest CHO genome scaffolds (parsed into 100kb windows that represents approximately ~20% of the CHO genome) that depicts distribution of genetic variants and MSI loci. The innermost track (light purple) depicts candidate MSI loci (23 out of the 124); Indel density and distribution is depicted in light blue; SNP density and distribution is in light red; and gene density is plotted outside of the CHO ideograms in light orange. Ideogram ticks are scaled in megabases.

6.4.3 Functional impact of ammonia-induced variants in genome stability genes

The above described 2,394 variants were assigned to 1,843 Chinese hamster protein-coding genes with certain functional impact. Through mapping of human orthologs for those Chinese hamster genes, we found 273 genes that are linked to genome instability terms via text mining (**Supplemental Table S6.5**). **Figure 6.3A** shows the map for KEGG enrichment result of over-representation test of the 273 genes. The five most significant KEGG pathways include breast cancer, cellular senescence, longevity regulating pathway, MAPK signaling pathway, and cell cycle. It is critical to note that the KEGG enrichment analysis (**Figure 6.3A**) combines all variants found in the 10 mM and 30 mM stress cultures, whereas gene lists for variants exclusively detected in 10mM or 30mM stress samples generated no enrichment of KEGG pathways. **Figure 6.3B** shows the KEGG comparison between three gene lists: one for all variants in 10mM sample, another for 30mM sample, and one for variants from the combined list (i.e., the above mentioned 273 genes). Breast cancer, cellular senescence, longevity regulating pathway are the three KEGG pathways common to the three gene lists (**Figure 6.3B**). These corresponding genes are listed in **Table 6.1**. **Figure 6.4** summarizes the significant GO terms enriched among the genes for variants existing at both stress levels. Notable GO biological process terms in **Figure 6.4** include DNA recombination, cell cycle checkpoint, regulation of response to DNA damage stimulus, telomere organization, and DNA damage checkpoint. Additionally, notable functions of genes include double-strand break repair (*Brca1*), mismatch repair (*Mlh3*), and centromere generation (*Cenpc*) (**Supplemental Table S6.5**). More detailed enrichment analysis on KEGG and GO of variant genes can be found in **Supplemental Tables S6.6-S6.9**.

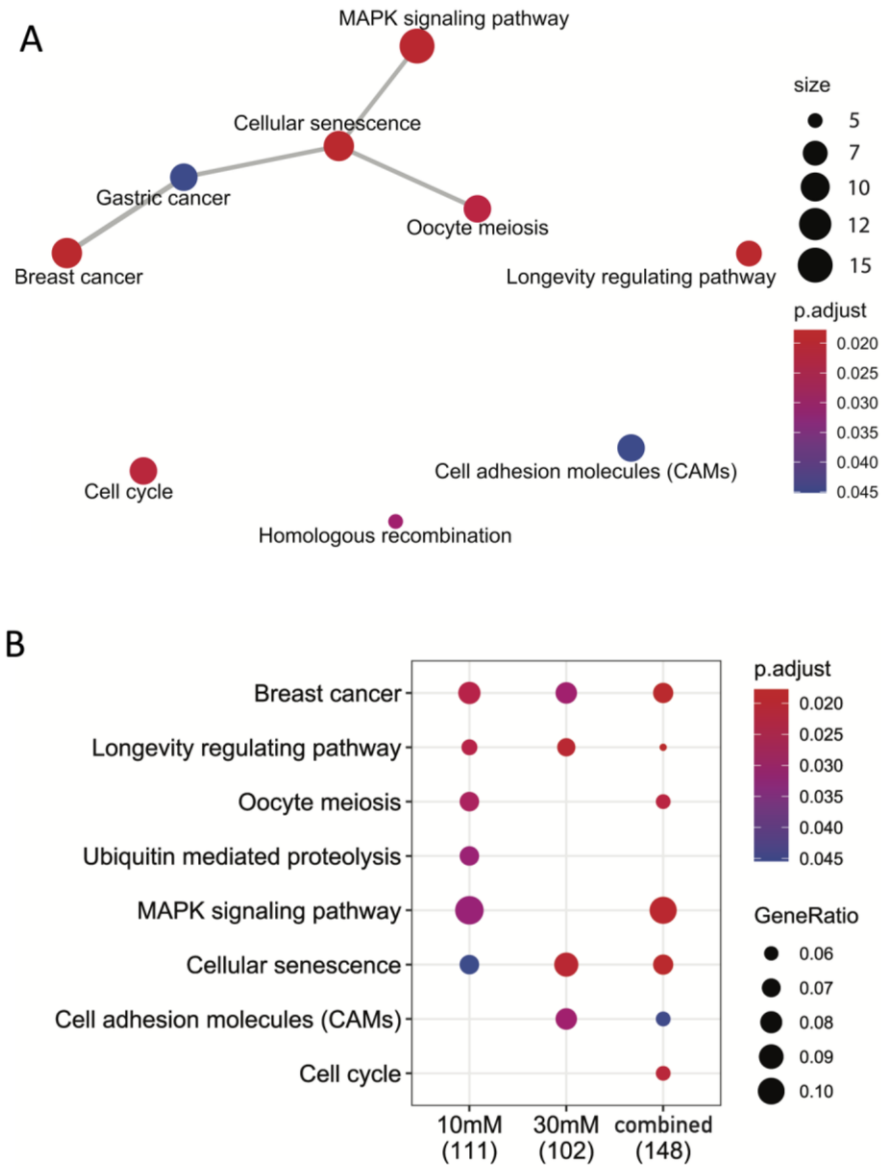


Figure 6.3: KEGG enrichment results from over representation analysis of ammonia-sensitive genes linked to genomic instability. A) Network plot of most significant enriched KEGG pathways. Enrichment scores (i.e., adjusted p values) and gene counts (the number of genes enriched in a KEGG term) are depicted by dot color and size. **B)** Comparison of enrichment results of KEGG pathways for genes identified with significant variants in the 10mM Ammonia-stressed culture (111 genes), in the 30mM culture (102 genes), and in the union of 10mM and 30mM cultures (148 genes). Enrichment scores (i.e., adjusted *p*-values) and gene ratios (the percentage of total genes in the given KEGG term) are depicted by dot color and size. The plots are made with clusterProfiler.

Table 6.1: **A summary of select KEGG enrichment genes discovered in ammonia-stressed cultures that can be linked to genome instability in humans via text mining.**

Gene Name (Human)	Entrez Gene ID (Human)	Entrez Gene ID (Chinese hamster)
tumor necrosis factor superfamily member 11 (TNFSF11)	8600	100768715
peroxisome proliferator activated receptor gamma (PPARG)	5468	100689245
interleukin 1 alpha (IL1A)	3552	100769260
Wnt family member 1 (WNT1)	7471	100766046
protein phosphatase 1 catalytic subunit alpha (PPP1CA)	5499	100760810
transforming growth factor beta receptor 1 (TGFBR1)	7046	100772727
E2F transcription factor 4 (E2F4)	1874	100765561
frizzled class receptor 2 (FZD2)	2535	100763109
LDL receptor related protein 6 (LRP6)	4040	100772150
tuberous sclerosis 2 (TSC2)	7249	100755849
lin-9 DREAM MuvB core complex component (LIN9)	286826	100774401
BRCA1, DNA repair associated(BRCA1)	672	100770724
mitogen-activated protein kinase kinase 1 (MAP2K1)	5604	100689403
protein kinase AMP-activated non-catalytic subunit gamma 3 (PRKAG3)	53632	100770459
activating transcription factor 2 (ATF2)	1386	100754663
RB1 inducible coiled-coil 1 (RB1CC1)	9821	100763340
progesterone receptor (PGR)	5241	100757656
fms related tyrosine kinase 4 (FLT4)	2324	100766609
Klotho (KL)	9365	100758189
notch 1 (NOTCH1)	4851	100761880
beta-transducin repeat containing E3 ubiquitin protein ligase (BTRC)	8945	100750426
nuclear factor kappa B subunit 1 (NFKB1)	4790	100770607
tumor protein p53 (TP53)	7157	100682525
cyclin E1 (CCNE1)	898	100753358
notch 2 (NOTCH2)	4853	100771788

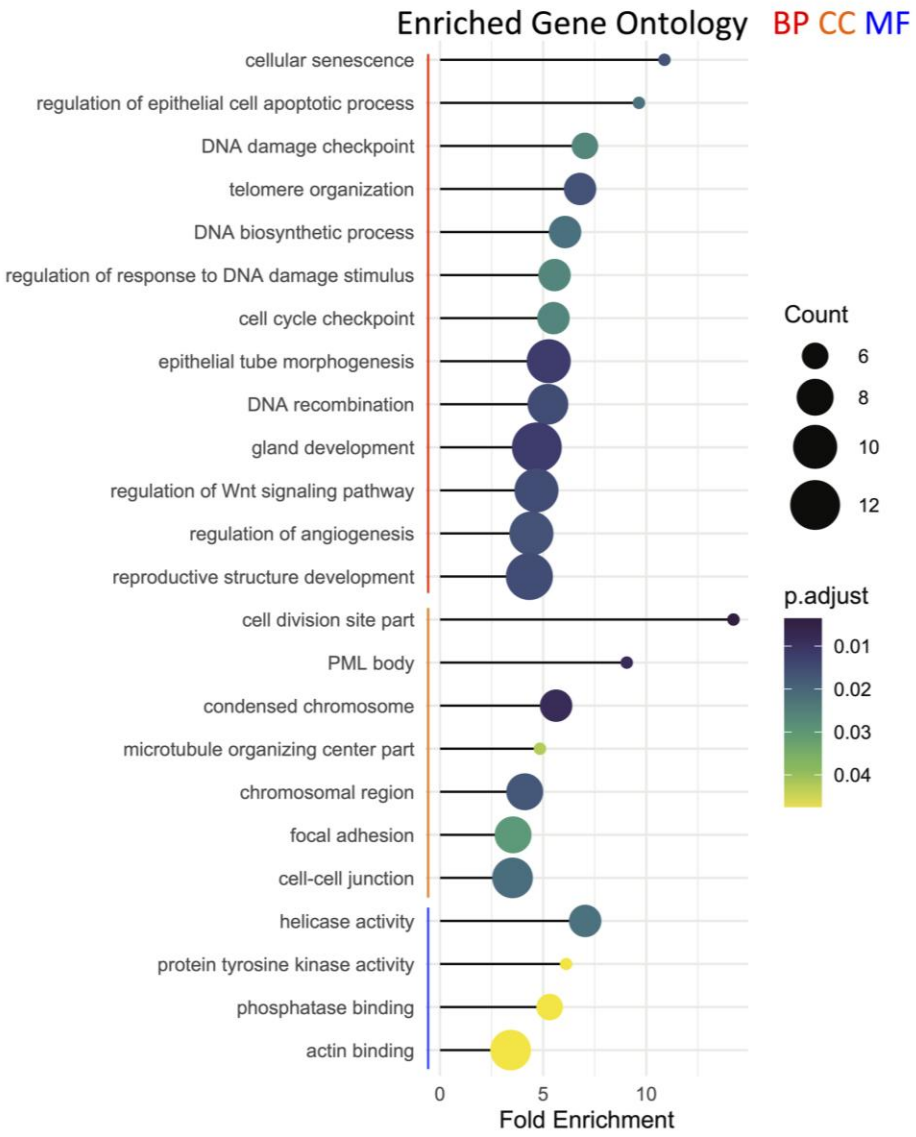


Figure 6.4: Gene Ontology (GO) enrichment results from over representation analysis of ammonia-sensitive genes linked to genomic instability. Terms in the three GO categories are grouped with color bars: biological process (red), cellular component (orange), and molecular function (blue). Fold enrichment is shown in x-axis. Enrichment scores (i.e., adjusted *p*-values) and gene counts (the number of genes enriched in a GO term) are depicted by dot color and size. The plot is made with clusterProfiler.

6.4.4 Microsatellite and candidate MSI loci

A whole-genome scan for microsatellites discovered a total of 409,628 loci, with motifs that included di-, tri-, and tetranucleotide repeats (**Supplemental Table S6.10**). As expected, the microsatellites composed of dinucleotide repeats were the most prevalent with a total of 287,124. Trinucleotide and tetranucleotide motifs were less abundant with 46,602 and 75,902 occurrences, respectively. An analysis of genome-wide indels in ammonia treated and control samples revealed 1,022 microsatellites that were lengthened or shortened due to the ammonia stress (**Supplemental Table S6.11**). An example microsatellite locus with desirable length variation resulting from elevated ammonia is shown in **Figure 6.4.** In this example, there is a higher abundance of mapped reads with deletions for the 30mM ammonia stressed cultures, suggesting a dose-dependent response. Furthermore, we developed a custom mutation score and stringent filtering criteria (see methods) to identify a candidate set of 124 MSI loci where stable mutations were present in both ammonia-stressed cultures, but were not present in the control cultures. These 124 MSI loci can be used for future research as diagnostics for genome instability (**Figure 6.2**). It is important to note that because the mutation score was calculated using the allelic depth; loci with more reads are statistically more significant than those with fewer reads. With this in mind, the 124 candidate MSI loci may not be all inclusive of the optimal loci due to the variation in mapped read depth across the genome. The remaining loci after each filter step is summarized in **Table 6.2**. A full list of loci in each step can be found in **Supplemental Tables S6.12-S6.15**, while the location of all candidate loci are summarized in **Table 6.3**.

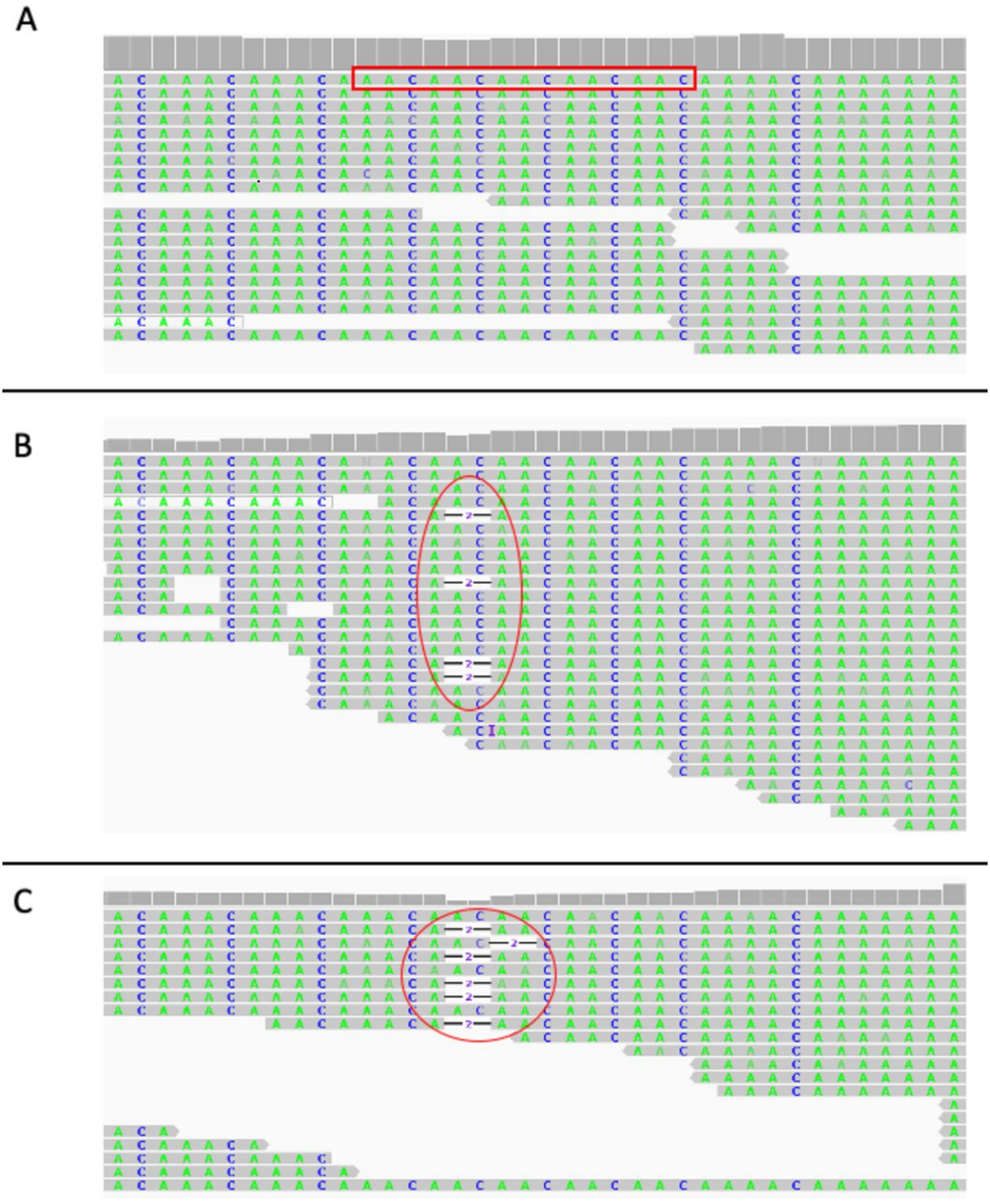


Figure 6.5: An Integrated Genome Viewer (IGV) image of a microsatellite located on scaffold NW_020822544.1 at position 4,160,116. A) control, **B)** 10 mM, and **C)** 30 mM cultures. This microsatellite contains five repeats of an AAC motif. They grey lines above each nucleotide are indicative of the total read depth at that location; note that it is significantly lower in regions where deletions were detected.

Table 6.2: Numerical representation of filter progression.

Filter Criteria	Remaining Loci
Genome-wide INDELs	79,097
30 mM Relative Mutation Score > 10 mM Relative Mutation Score	35,437
Loci with positive Relative Mutation Score	16,678
Loci confirmed to be microsatellites	187
Loci with genotypes that vary from the control	124

Table 6.3: Location and composition of all candidate microsatellites. More detailed information on candidate loci can be found in **Supplemental Table S6.15.**

			Mutation Scores (%)						Mutation Scores (%)		
Scaffold	**Position**	**Motif**	**10 mM**	**30 mM**	**Control**	**Scaffold**	**Position**	**Motif**	**10 mM**	**30 mM**	**Control**
NW_020822370.1	34498325	(GA)31	85.71	94.44	83.33	NW_020822439.1	16584719	(GT)8	20	36.36	7.14
NW_020822370.1	29116483	(AG)22	14.29	62.5	0	NW_020822439.1	2390672	(TG)62	12.5	33.33	0
NW_020822370.1	8110802	(GT)33	14.29	20	7.14	NW_020822439.1	243530	(CA)25	25	50	0
NW_020822373.1	4520712	(TC)27	10.34	12.4	4.59	NW_020822439.1	17884454	(AGAT)11	13.33	20	7.14
NW_020822375.1	17984519	(GT)23	90	100	77.78	NW_020822440.1	3560698	(GA)28	88.89	90	83.33
NW_020822375.1	24391224	(CA)8	41.67	70	0	NW_020822443.1	2115154	(AC)27	27.27	43.75	9.09
NW_020822376.1	2957094	(CA)7	13.04	21.05	8.33	NW_020822446.1	382267	(CA)32	66.67	75	33
NW_020822382.1	3087333	(CCTC)5	14.29	21.43	0	NW_020822450.1	9390312	(AC)22	16.67	44.44	11.11
NW_020822403.1	19387296	(AC)25	28.57	50	4.76	NW_020822452.1	9210374	(CT)31	90	93.33	81.82
NW_020822403.1	15933029	(AC)26	25	46.15	0	NW_020822458.1	15628677	(AG)27	91.67	100	73.33
NW_020822406.1	3701096	(GT)23	14.29	42.86	9.09	NW_020822461.1	21799265	(TCTT)16	16.67	25	0
NW_020822407.1	7711795	(AAAC)6	16.67	26.67	0	NW_020822464.1	1598541	(AC)26	84.62	100	83.33
NW_020822409.1	5651224	(AAT)12	14.29	44.44	0	NW_020822464.1	6242934	(TG)26	27.78	36	0
NW_020822410.1	13888699	(GT)28	83.33	90.32	63.64	NW_020822465.1	1645147	(TG)9	39.29	54.55	5.88
NW_020822410.1	10858329	(AC)19	94.12	96.15	75	NW_020822468.1	15157395	(GT)35	87.5	100	66.67
NW_020822412.1	3775426	(TG)26	21.43	28.57	0	NW_020822469.1	3552281	(GATG)8	31.25	54.55	0
NW_020822415.1	36220614	(TC)31	91.67	92.31	57.14	NW_020822484.1	1703831	(TG)7	95.65	100	77.27
NW_020822415.1	16052939	(ACA)5	19.23	23.33	5.26	NW_020822486.1	16687223	(GT)23	21.43	22.22	0
NW_020822421.1	5779917	(TC)29	91.3	100	80	NW_020822487.1	21884245	(CT)34	89.66	96.43	80
NW_020822426.1	5559446	(GA)31	93.75	100	90.91	NW_020822487.1	17786455	(TC)30	91.67	100	87.5
NW_020822426.1	3827444	(TCT)30	71.43	72.73	55.56	NW_020822487.1	25648949	(AC)15	25	37.5	10
NW_020822426.1	1539468	(GT)22	27.27	28.57	0	NW_020822487.1	8954697	(CA)29	10	18.18	0
NW_020822428.1	1367872	(AC)26	93.75	100	85.71	NW_020822488.1	1550207	(TG)26	85.71	100	75
NW_020822428.1	8085398	(TCCT)11	16.67	27.27	0	NW_020822499.1	2670448	(TG)31	95.65	100	84
NW_020822434.1	512274	(GT)28	18.18	44.44	0	NW_020822499.1	21810339	(CT)20	35.71	56.25	0
NW_020822436.1	3837117	(AC)16	90.32	93.33	38	NW_020822499.1	16894388	(TTA)5	11.9	13.73	6.67
NW_020822439.1	68421902	(AC)29	91.67	100	83.33	NW_020822499.1	10821740	(TTGT)8	14.29	38.46	7.14
NW_020822439.1	47065858	(TC)35	84.62	94.74	80	NW_020822501.1	11357080	(GT)29	91.67	100	85.71
NW_020822439.1	62174741	(GT)23	22.22	35.48	11.11	NW_020822501.1	14044688	(TTG)7	14.29	24.32	3.33
NW_020822439.1	14184660	(AAAC)7	12	50	0	NW_020822501.1	14071142	(TTA)12	20	28.57	5.56
NW_020822439.1	66094811	(AC)24	25	28.57	0	NW_020822501.1	17408028	(TG)30	55.56	57.14	7.14

Table 6.3 continued

			Mutation Scores (%)						Mutation Scores (%)		
Scaffold	Position	Motif	10 mM	30 mM	Control	Scaffold	Position	Motif	10 mM	30 mM	Control
NW_020822503.1	10248411	(CTTT)14	22.22	25	0	NW_020822567.1	3062112	(CA)7	17.39	29.41	3.85
NW_020822503.1	4513778	(GT)14	10.53	20	0	NW_020822567.1	1362729	(GA)24	15.38	25	0
NW_020822503.1	17353421	(GT)6	22.22	25	9.09	NW_020822570.1	21823495	(TG)28	14.29	44.44	0
NW_020822504.1	9433369	(TC)30	95	100	88.89	NW_020822591.1	7123790	(TATT)6	94.44	100	80
NW_020822504.1	13319850	(TGTA)5	12.5	15.56	5.88	NW_020822591.1	7780615	(TAAA)8	33.33	37.5	0
NW_020822505.1	16324767	(CA)24	93.75	100	77.78	NW_020822591.1	14941	(GA)34	9.09	25	0
NW_020822505.1	16510213	(AC)30	85.71	87.5	77.78	NW_020822591.1	8820175	(TTAT)11	9.52	27.27	0
NW_020822505.1	10642584	(TG)27	27.27	37.5	0	NW_020822592.1	4061125	(AC)31	42.86	44.44	0
NW_020822508.1	1001369	(AC)22	90.91	100	71.43	NW_020822595.1	5415636	(CAAA)6	37.93	38.46	7.69
NW_020822508.1	2728629	(TATC)10	16.67	27.78	0	NW_020822597.1	5771854	(TG)31	88.24	94.44	75
NW_020822508.1	15865728	(GAAG)13	7.69	31.25	0	NW_020822601.1	72176322	(CA)8	91.18	100	85.71
NW_020822508.1	15865731	(GAAG)13	7.14	26.32	0	NW_020822601.1	61385941	(AC)19	21.43	22.73	0
NW_020822511.1	9568885	(TG)29	92.31	100	83.33	NW_020822602.1	6751043	(TC)30	91.67	96.55	83.33
NW_020822512.1	6724926	(TG)30	87.5	93.33	85.71	NW_020822602.1	9765422	(AATA)7	15.38	15.79	0
NW_020822519.1	6274345	(AC)25	91.67	100	84.62	NW_020822603.1	9112470	(AC)25	92.31	100	66.67
NW_020822519.1	11516391	(AG)36	15.79	37.5	0	NW_020822604.1	8845685	(GT)6	10	37.5	0
NW_020822520.1	2511201	(AC)24	20	25	0	NW_020822610.1	821820	(AC)27	31.25	41.18	10
NW_020822526.1	18794354	(AC)26	91.67	100	83.33	NW_020822614.1	8982333	(ATAG)16	13.64	13.79	0
NW_020822526.1	25646840	(CA)7	90	100	77.78	NW_020822629.1	5127703	(AG)28	12.5	20	0
NW_020822526.1	24206438	(GT)30	87.5	90	70	NW_020822629.1	6170860	(CA)22	12.5	33.33	0
NW_020822526.1	16873752	(AG)32	83.33	100	69.23	NW_020822634.1	2791562	(TC)33	95.24	100	62.5
NW_020822526.1	17494091	(AC)16	17.65	57.14	0	NW_020822638.1	3030009	(AC)21	92.31	100	58.33
NW_020822526.1	4205760	(ATGT)11	16.67	37.5	0	NW_020822698.1	6755673	(GT)27	90	100	80
NW_020822529.1	28253360	(TG)23	91.67	100	86.67	NW_020822698.1	121373	(TG)30	28.57	37.5	0
NW_020822529.1	16865510	(TC)27	42.86	54.55	0	NW_020822785.1	9557	(CA)14	14.29	75	0
NW_020822530.1	10618047	(TG)22	18.18	28.57	0	NW_020822967.1	23702	(CA)17	15.38	33.33	9.09
NW_020822531.1	6287546	(AC)27	90.91	93.75	81.82	NW_020823044.1	10660	(TTG)6	15.29	15.45	2.53
NW_020822531.1	1037419	(GT)44	45.45	50	12.5	NW_020823531.1	97528	(AG)22	11.76	15.22	5.41
NW_020822544.1	4160116	(AAC)5	22.73	60	0	NW_020823768.1	45602	(TC)26	93.75	95.83	78.95
NW_020822567.1	4418183	(TG)29	91.67	100	80	NW_020824031.1	35819	(AC)23	20	25	0
NW_020822567.1	6942408	(AC)24	25	35.29	0	NW_020824065.1	30045	(AG)6	12	14.29	4.65

6.5 Discussion

Ammonia is a common metabolic waste product in cell cultures. The accumulation of ammonia most often leads to decreased cell and recombinant protein productivity. Typical fed-batch cultures last for 14 to 20 days, where in the exponential phase, cell division can occur daily. *De novo* mutations that occur early in culture will be amplified and have the potential to dominate the cell population as the culture approaches harvest. In this study, two ammonia stresses (10 mM) and (30 mM) were used to investigate the genotoxic effects of this byproduct on CHO cell fed-batch cultures. Further, the role of ammonia stress on genome instability was investigated. Despite a relatively short exposure duration of 72 hours, MSI loci were identified, which have the potential to be biomarkers for genome instability.

The VCD, cell viability, and metabolic profiles indicated that the 30 mM ammonia stress significantly impacted the culture health, as the characteristic cell growth and metabolic profiles were significantly different from the control cultures. The effects of the 10 mM ammonia stress were less profound, yet the metabolic profiles and protein productivity were more sensitive to these changes than the cell viability and VCD profiles. At the time of sampling for genetic analysis (3.5 days), VCDs for the control and 10 mM cultures appeared to be matched, whereas the 30 mM ammonia-stressed cultures had lower VCDs. The decreased consumption of alanine observed for the ammonia-stressed cultures was the only metabolic difference observed at sampling for the whole genome sequencing. Alanine metabolic changes are known to occur under ammonia stress [328]. Therefore, the whole genome sequencing would identify changes due to the ammonia

stresses, and not due to potential other culture condition differences that might accumulate.

Until now, efforts to characterize ammonia stress effects on CHO cells have mainly focused on transcriptome, proteome, and product characteristic changes [56,316-318,329-332]. In this study, the effects of ammonia stress were further characterized by examining variants within functional genes and microsatellites. Whole genome sequencing allowed for variant SNPs and indels to be identified. Moreover, greater than 2,300 high or moderate-impact novel gene variants were identified from the ammonia-stressed cultures that may impact cellular functions of critical pathways. KEGG and GO enrichment analyses confirmed that many of the variant genes affected pathways could lead to suboptimal clone performance. Though thousands of variant genes were identified, this list was narrowed to focus on genes pertaining to pathways involved in genome stability (**Figures 6.3** and **6.4**, **Table 6.1**).

Alterations in critical genes responsible for a wide variety of processes such as transcription regulation, cell cycle regulation, tumor suppression, and signaling pathways may lead to global genome instability. *De novo* genomic SNPs and indels accumulating is typically the result of replication errors which can result from a variety of mechanisms such as replication stalling [85], replication fork collapse [86,87], double-strand breaks [102,333], environmental stressors, transcription regulation errors, or other replication errors [334]. All of these replication mechanisms can be linked to error correction fidelity of DNA repair mechanisms. These DNA repair errors, in turn, can lead to an accelerated variant

accumulation rates and loss of genome stability [335]. Mutations, such as synonymous base changes in coding and regulatory regions, normally have little to no effect on gene transcription and translation, however, non-synonymous changes can have functional effects on the subsequent amino acid sequence and folding or function that ultimately can be linked to loss in cell viability.

Through text mining approaches, 273 of the variant genes found in the CHO genome were linked to human orthologs; whose function are related to genome stability maintenance. One gene identified is exceptionally well-known for its role in double-strand break repair and tumor suppression, *Brca1* (**Table 6.1**); loss of *Brca1* function has been associated with increased breast cancer incidence and metastasis, which demonstrates its critical function in maintaining stability [99]. Genome instability can be further exacerbated by the loss of tumor suppressor function. For example, *Lin9* (**Table 6.1**) is a tumor suppressor that inhibits DNA synthesis and acts synergistically with the well-known *Rb1* gene to prevent rapid, uncontrolled cell division [215]. Therefore, loss of function in *Lin9* can lead to cancer-like growth of mutant cells that would eventually dominate the culture population.

Some variant genes belonged to three significantly enriched KEGG pathways related to genome instability in humans - cellular senescence, cell cycle, and homologous recombination (**Figure 6.3**). Cellular senescence occurs as a result of multiple stimuli such as DNA damage and oxidative stress [336]. By forcing the cells into a non-replicative state, senescence can severely limit the productivity of cell culture, especially when it

occurs before or in the early exponential growth phase. The 30 mM stressed cultures had more genes enriched in the senescence pathway (**Supplemental Table S6.6**), which makes variant genes in this pathway a likely contributor to the poor growth observed. The second pathway, cell cycle, was observed to have significant enrichment in union genes of the 10 mM and 30 mM ammonia-stressed cultures. The cell cycle contains multiple checkpoints to ensure daughter cells are healthy and contain undamaged DNA[98,103]. Significant enrichment in this pathway indicates that damaged or otherwise improperly replicated DNA could be passed on to daughter cells. Finally, the homologous recombination pathway repairs damage caused by double strand breaks by using an identical sequence as a template [337]. This repair method is much more accurate than non-homologous end joining and is less prone to variant generation [91].

It should be noted that while the mismatch repair (MMR) pathway was not found to be significantly enriched, three notable MMR genes accumulated variants: *Mlh3* (a *MutL* homolog), *Rpa1*, and *Abl1* (**Supplemental Table S6.5**). An impaired or inefficient MMR system can lead to the accumulation of mutations in functional genes over cell divisions that are critical to the cell's survival and can lead to loss of genetic stability [33] or disease states, such as cancer [96]. The need for a highly conserved MMR system can be observed by the presence of multiple orthologs of *MutS* and *MutL* in eukaryotic genomes [95]. *MutS* binds to base mismatches or small indels [36,338] while *MutL* is responsible for communicating the identification of mismatch events to downstream elements of MMR such as exonucleases [339].

Variants in genes that regulate the MMR pathway may be an origin to the cascade of events that leads to genome instability. When the MMR pathway in a cell is compromised, mistakes can occur and propagate indiscriminately across the genome as cell division occurs[36,95,338]. Unfaithful replication of genomic repeats, such as microsatellite repeats, have been used as effective biomarkers in predicting certain diseases, such as cancer In this study, we found 1,022 microsatellites with variable repeat lengths in the WGS reads from ammonia stressed samples. We developed criteria and an approach to identify microsatellite loci that have variant length that can be attributed to ammonia stress. For each site, we considered read depth, mutation type, and frequency in affected and control samples to subset 124 candidate MSI loci that contain indels with a dose dependent response to the ammonia concentrations that were not observed in the control cultures. This set of 124 MSI loci represent a potential biomarker set that could have utility to predict genome instability in CHO cell cultures under stressful culture conditions.

6.6 Conclusion

The accumulation of metabolic wastes, such as ammonia, can have a profound effect on CHO cell culture viability, transcriptome, and recombinant protein productivity. Additionally, past work, as well as this study, have observed shifts in growth patterns and metabolic profiles due to the ammonia stress. Further, in this study, it was observed that high levels of exogenous ammonia caused *de novo* mutations, such as SNPs and indels, within functional genes. More importantly, these mutations persisted throughout the culture population. Variants were identified in the genes that regulate critical cellular

processes, such as DNA repair; which is a hallmark of genome instability. In addition to characterizing the microsatellite content of the Chinese hamster genome, potential MSI loci that exhibited unfaithful replication in the presence of exogenous ammonia were identified; these microsatellites could be utilized as a tool to diagnose genome instability in future work.

CHAPTER 7

CONCLUSIONS AND FUTURE DIRECTIONS

7.1 Conclusions

The objective of this study was to establish the foundational knowledge of eccDNA structure, function, and microevolutionary dynamics for CHO cells in multiple contexts. In Chapter 3, eccDNA evolution was found to be highly dynamic in a tightly controlled, two-week fed-batch culture in control and lactate-stressed conditions. Further, CHO cell eccDNAs were shown to be highly heterogeneous, potentially have a small impact on global gene expression, and may be modulated by environmental shifts. In Chapter 4, it was observed that eccDNA gene content was not largely impacted by PDL, but was greatly different between cell lines derived from the same lineage. Interestingly, eccDNA biogenesis was found to be more frequent in gene dense regions of the CHO genome; this contradicts existing evidence of eccDNAs in filamentous plant pathogens that are more likely to arise from repeat-dense regions of the genome potentially due to different drivers of genome evolution. In Chapter 5, eccDNAs were not observed to contribute to significant gene expression changes associated with lactate adaptation. Surprisingly, no gene expression changes of lactate metabolism genes were observed; however, differential expression of genes responsible for maintaining genome stability was observed. Data shown indicate that mass action, as suggested by Freund et al., is unlikely to be the mechanism of lactate consumption in CHO cells. It is more likely the mechanism behind the lactate metabolic switch is modulated by glutamine depletion and/or redox

balancing rather than mass action or gene expression shifts. In Chapter 6, *de novo* mutations, such as SNPs and indels, were observed in key genome instability-linked genes in response to an ammonia stress. Notably, a panel of microsatellite loci were observed to exhibit instability when cultured in ammonia-stressed conditions in a dose-dependent manner. These microsatellites may be usable for quick biomarkers for genome stability in CHO cell cultures.

7.2 Future Directions

This work characterized eccDNA libraries that were generated using an economic, yet non-quantitative, method. Future experiments characterizing eccDNAs in CHO cell lines should be performed at a larger scale to have sufficient material for quantitative sequencing methods to better understand the abundance of specific eccDNAs. Establishing an efficient, quantitative eccDNA sequencing method would enable accurate identification of biomarkers that may be indicative of genome stability. Additionally, beneficial information could be gained by further study of CHO cell eccDNAs in perfusion cultures, in multiple host CHO cell lines, and in fed-batch culture of lactate-adapted cells. Studying eccDNAs of CHO cells grown in perfusion culture would allow for an understanding of how PDL and long-term culturing impacts the circulome and gene expression patterns in a tightly-controlled controlled environment. Also, by studying eccDNAs in multiple host CHO cell lines would provide insight to the evolution of the CHO genome. Understanding how cell line derivation impacts eccDNA sequence structure, biogenesis, and function could explain why some cell lines are more stable and identify engineering targets that may be useful in all CHO cell lines if eccDNA engineering is

pursued. While eccDNAs were not observed to have a substantial impact on gene expression shifts in lactate-adapted CHO cells, lactate-adapted cells should be studied in fed-batch cultures to characterize longer term stability. Lactate-adapted cells show a more rapid lactate consumption and reduced ammonia production, however downregulation of genome stability genes may impact stability in extended cultures.

While no native eccDNAs in any species have been successfully engineered as of this writing, the data presented in this work provides the baseline understanding that would be required to engineer eccDNAs in CHO cell lines. Regions of the genome with a high abundance of eccDNA biogenesis may be of particular interest for generating recombinant cell lines. While insertion of transgenes into high biogenesis regions may not be ideal for stable expression, it could be desirable to attempt to circularize transgenes or other regulatory sequences. For instance, it may be ideal to clone a transgene into a stable part of the genome, while a short enhancer sequence and origin of replication are inserted into a region of high eccDNA biogenesis. This could increase recombinant protein production by generating autonomously-replicating eccDNAs harboring structures to enhance transgene expression.

APPENDICES

Appendix A

LIST OF SUPPLEMENTAL TABLES

Appendix A: List of Supplemental Tables (continued)

Appendix A: List of Supplemental Tables (continued)

Appendix A: List of Supplemental Tables (continued)

Appendix A: List of Supplemental Tables (continued)

S6.5 Human-Chinese hamster gene orthologs that can be linked to genome instability via text mining. ..... C6S1_DGC.xls

S6.6 KEGG enrichment of variant genes that can be linked to genome instability. ..... C6S2_DGC.xls

S6.7 Statistically significant Biological Process GO terms of variant genes. ..... C6S2_DGC.xls

S6.8 Statistically significant Molecular Function GO terms of variant genes. ..... C6S2_DGC.xls

S6.9 Statistically significant Cellular Component GO terms of variant genes. ..... C6S2_DGC.xls

S6.10 Complete list of microsatellites found in the Chinese hamster genome. ..... C6S2_DGC.xls

S6.11 Microsatellites containing insertion and or deletion mutations. ..... C6S2_DGC.xls

S6.12 Indel loci with higher mutation frequencies In 30 mM ammonia-stressed cultures. ..... C6S2_DGC.xls

S6.13 Indel loci from table S6.12 where mutation frequency is higher in the 10 mM samples compared to the control. ..... C6S2_DGC.xls

S6.14 Intersection of dose-dependent indels (S6.13) and genome-wide microsatellites (S6.10). ..... C6S2_DGC.xls

S6.15 The 124 candidate microsatellite loci that exhibited dose-dependent variation in response to amminia stress. ..... C6S2_DGC.xls

Appendix B

Chapter 3 Supplemental Figures

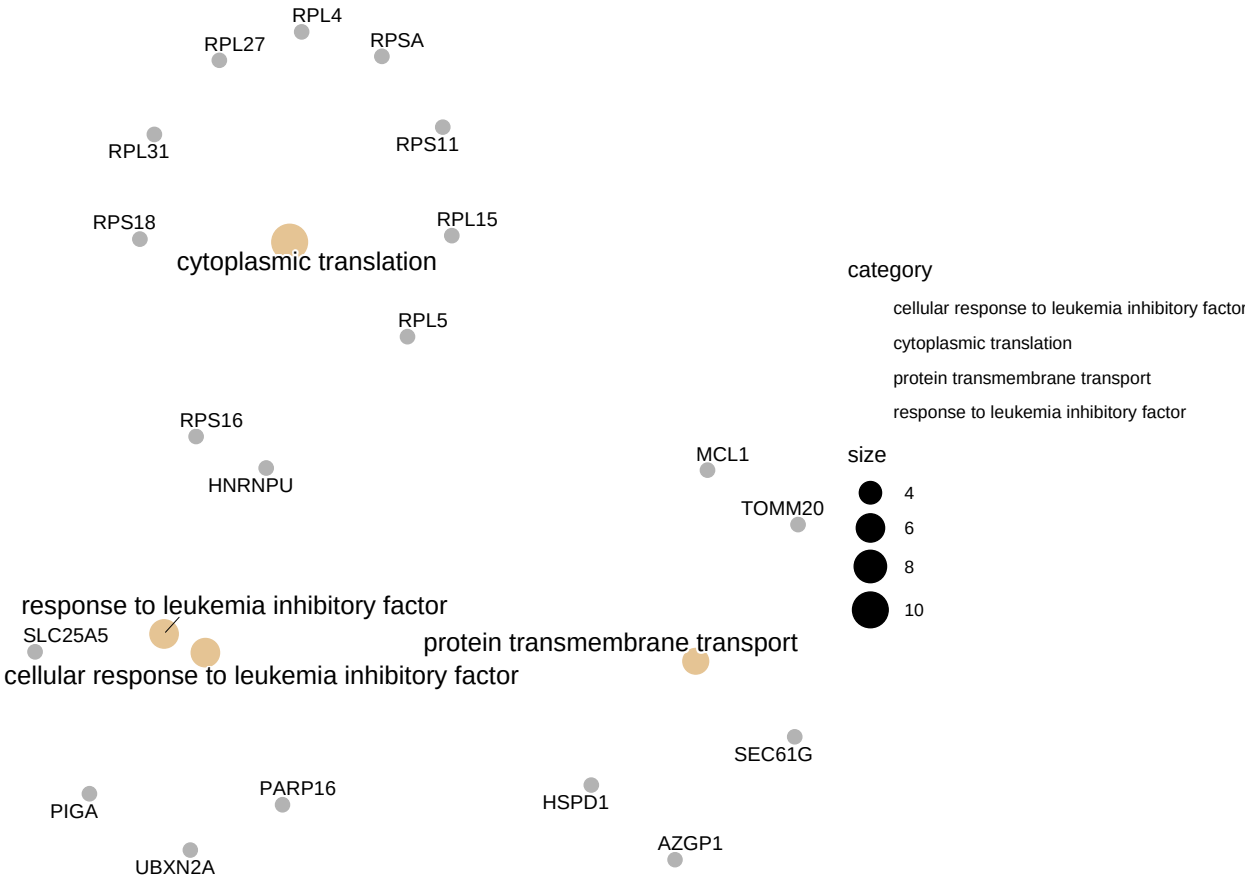


Figure S3.1: Control Day 12 BP. Network diagram of significantly enriched GO biological process terms for the human orthologs of Chinese hamster genes detected in Control Day 12 samples. The small gray nodes show individual genes and larger beige nodes indicate GO terms. The size of the beige nodes is proportional to the number of genes with that GO term and the colored lines indicate the GO category for which a gene belongs to.

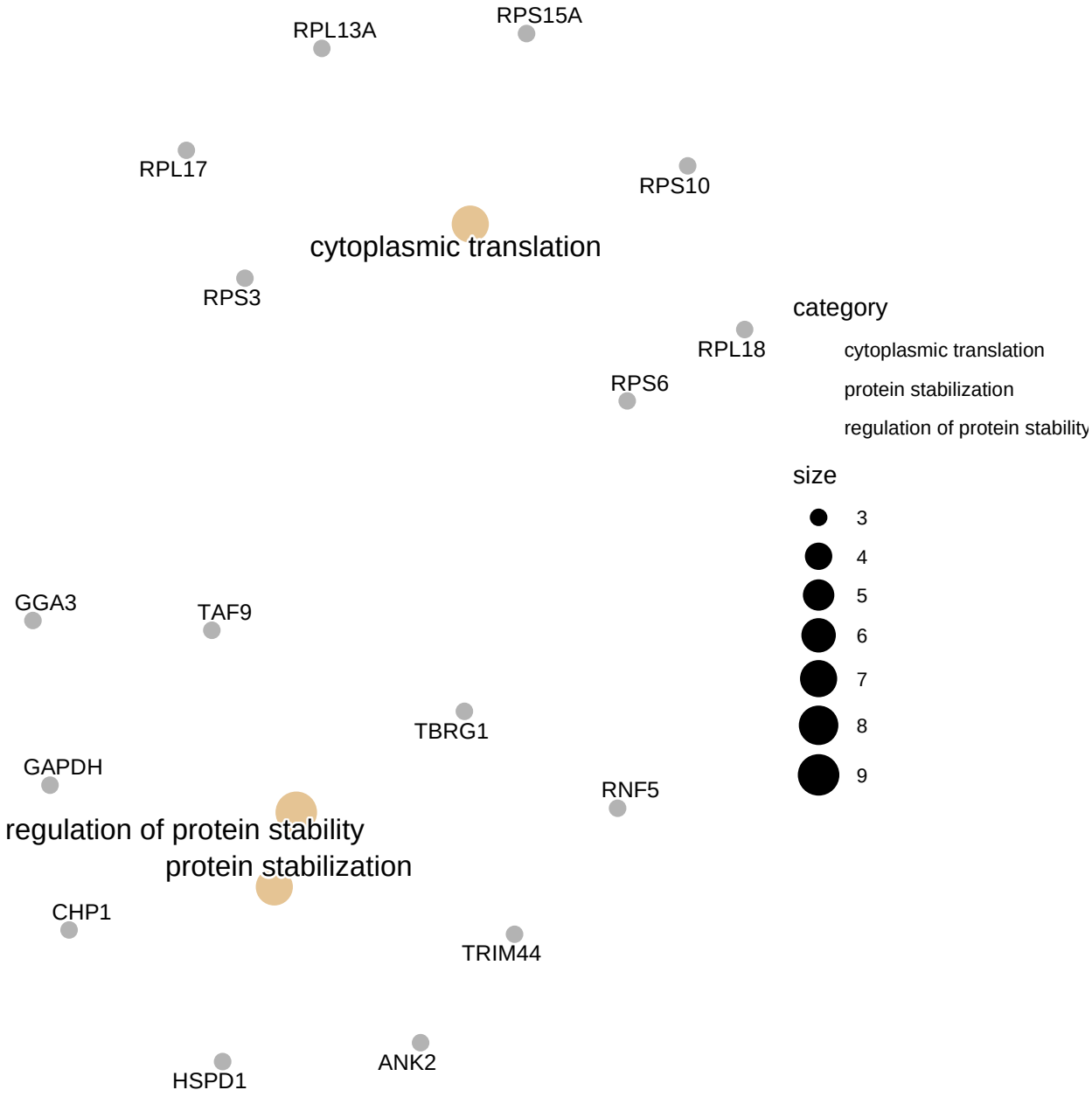


Figure S3.2: Lactate-stressed Day 12 BP. Network diagram of significantly enriched GO biological process terms for the human orthologs of Chinese hamster genes detected in Lactate-stressed Day 12 samples. The small gray nodes show individual genes and larger beige nodes indicate GO terms. The size of the beige nodes is proportional to the number of genes with that GO term and the colored lines indicate the GO category for which a gene belongs to.

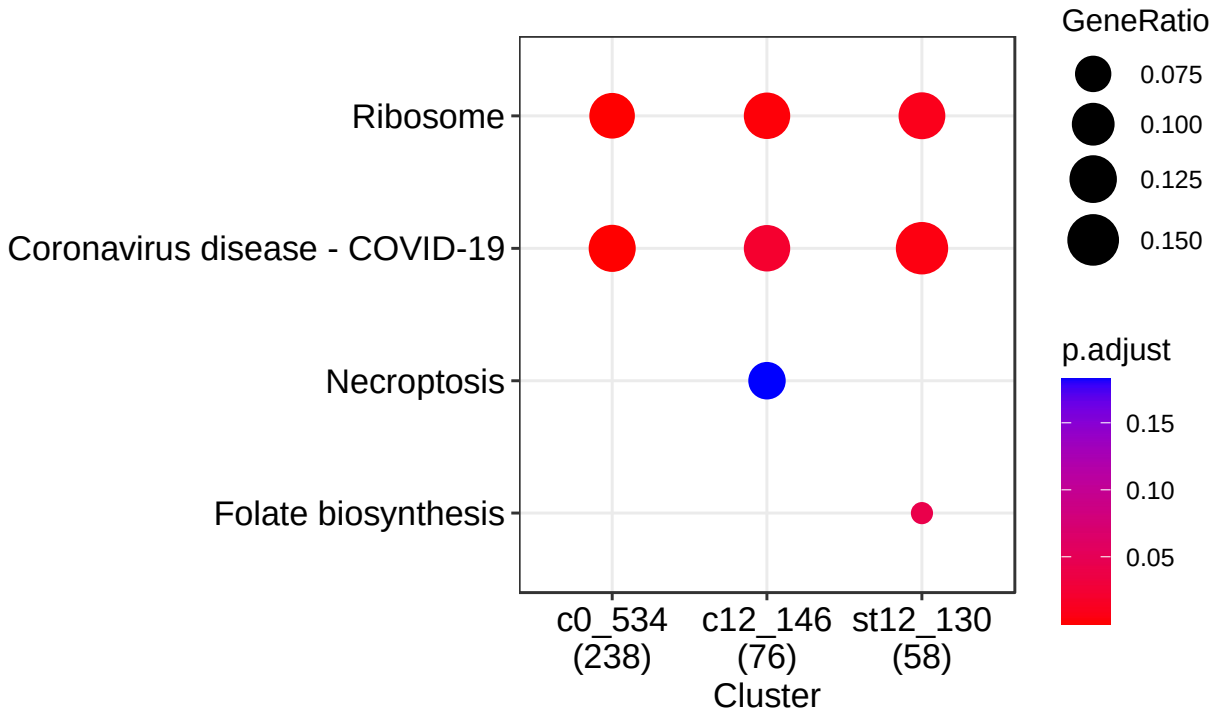


Figure S3.3: KEGG Pathway analysis of observed eccDNA genes. Node size is proportional to the number of genes found in the pathway, while the node color indicates the pathway's statistical significance.

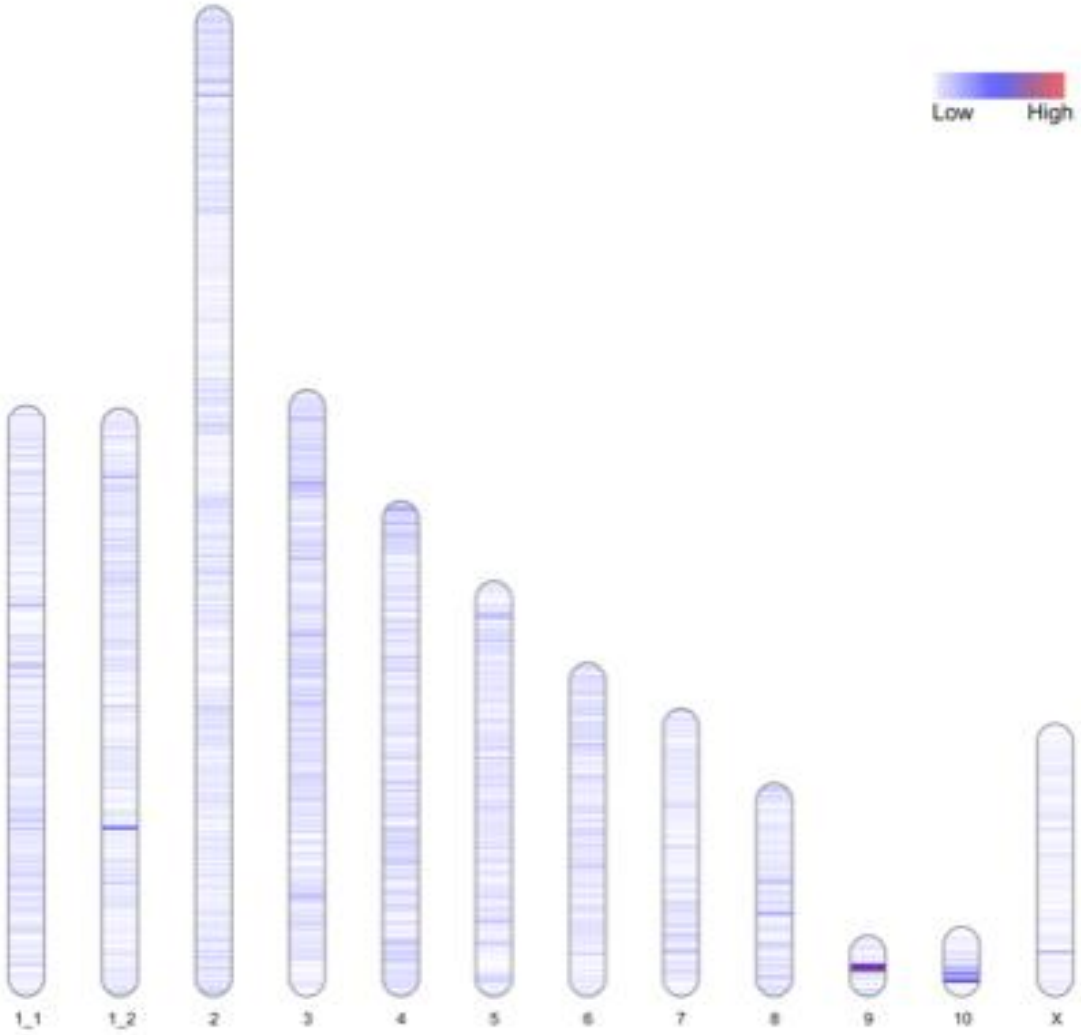


Figure S3.4: Control Day 12 eccDNA biogenesis map. Chromosome-scale heatmap of eccDNA sequences observed for the Control Day 12 samples. Frequency of observed eccDNA is shown in color; low (white) to high (red).

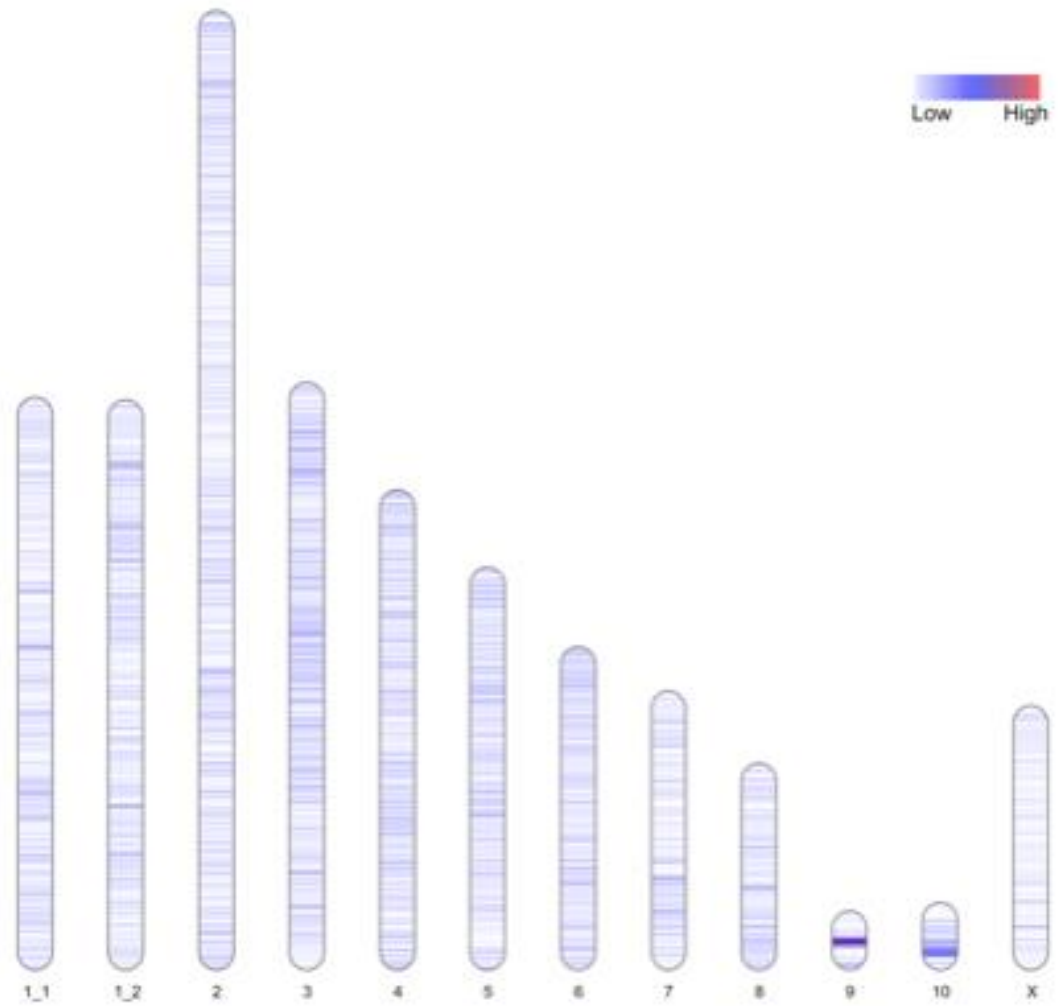


Figure S3.5: Lactate-stressed Day 12 eccDNA biogenesis map. Chromosome-scale heatmap of eccDNA sequences observed for the Lactate-stressed Day 12 samples. Frequency of observed eccDNA is shown in color; low (white) to high (red).

Appendix C

Chapter 4 Supplemental Figures

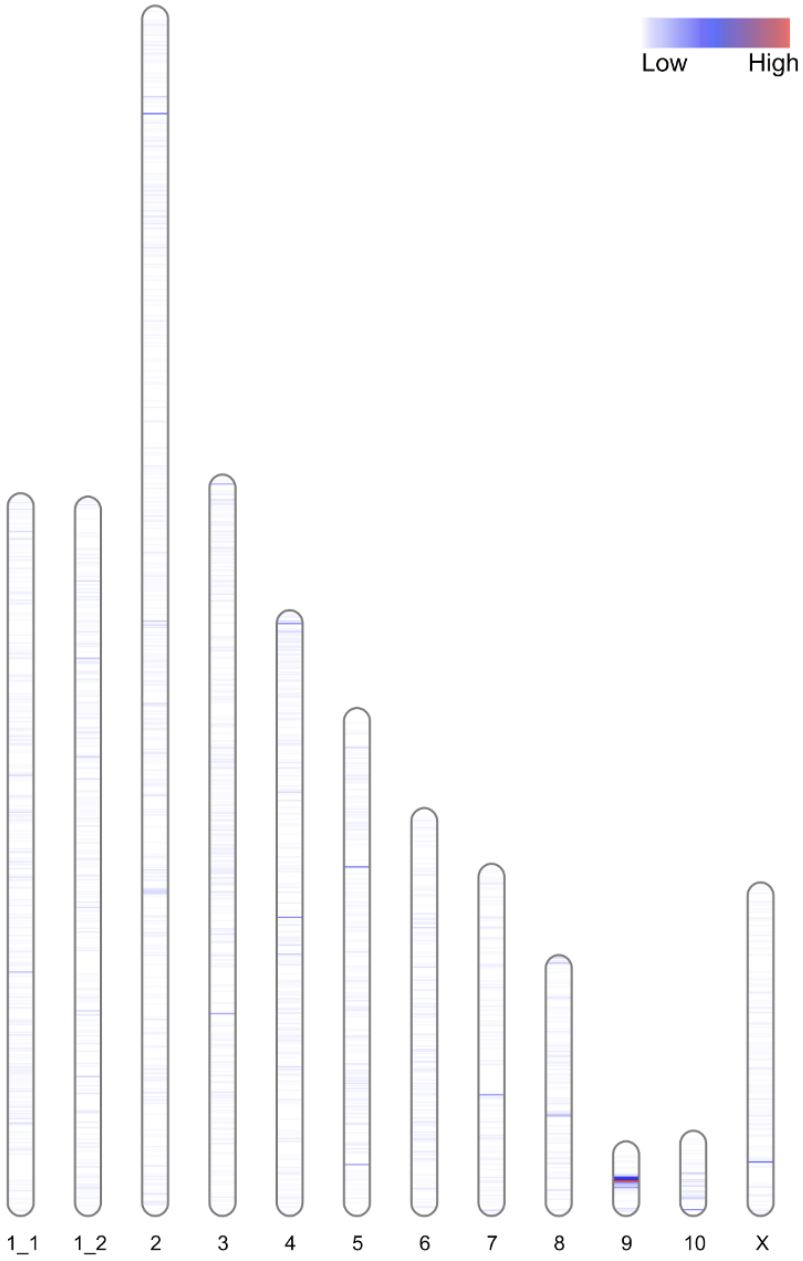


Figure S4.1: VRC01 PDL 10 eccDNA biogenesis map. Chromosome-scale heatmap of eccDNA sequences observed for the VRC01 PDL 10 samples. Frequency of observed eccDNA is shown in color; low (white) to high (red).

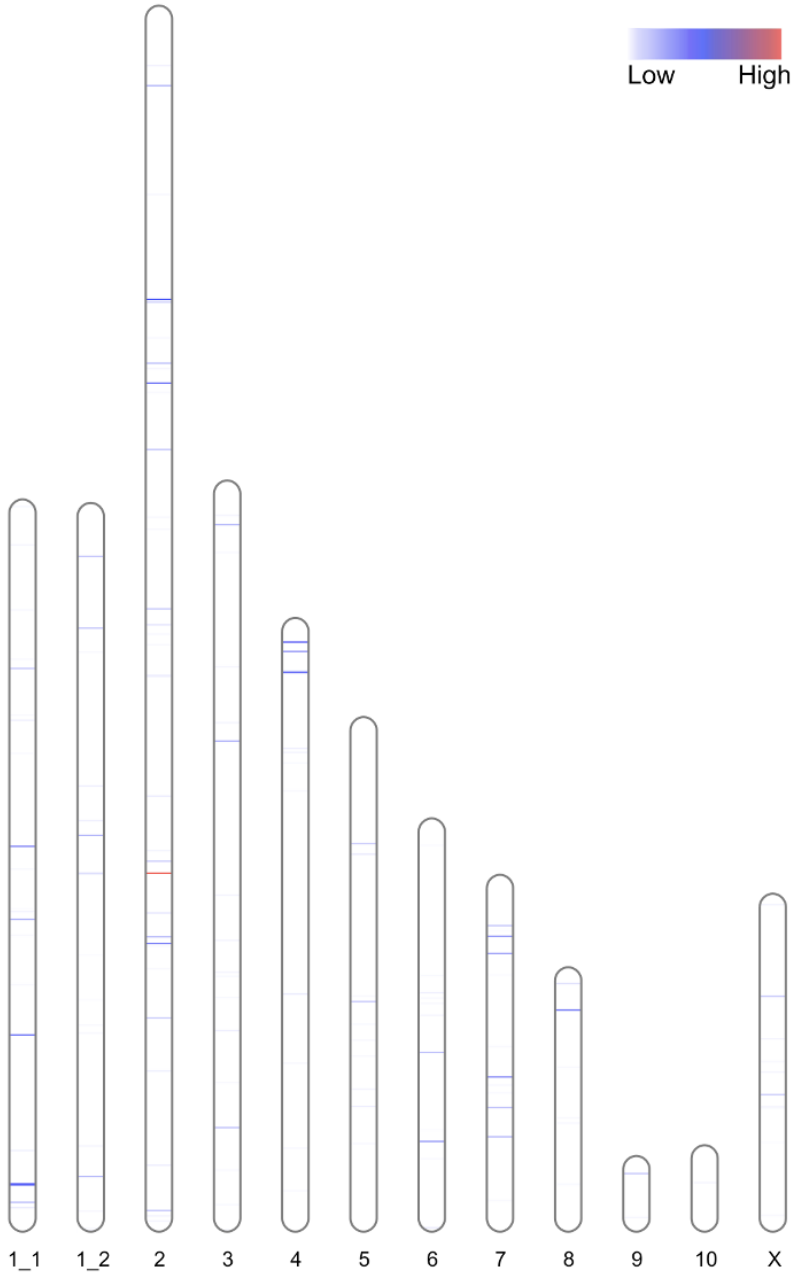


Figure S4.2: VRC01 PDL 24 eccDNA biogenesis map. Chromosome-scale heatmap of eccDNA sequences observed for the VRC01 PDL 24 samples. Frequency of observed eccDNA is shown in color; low (white) to high (red).

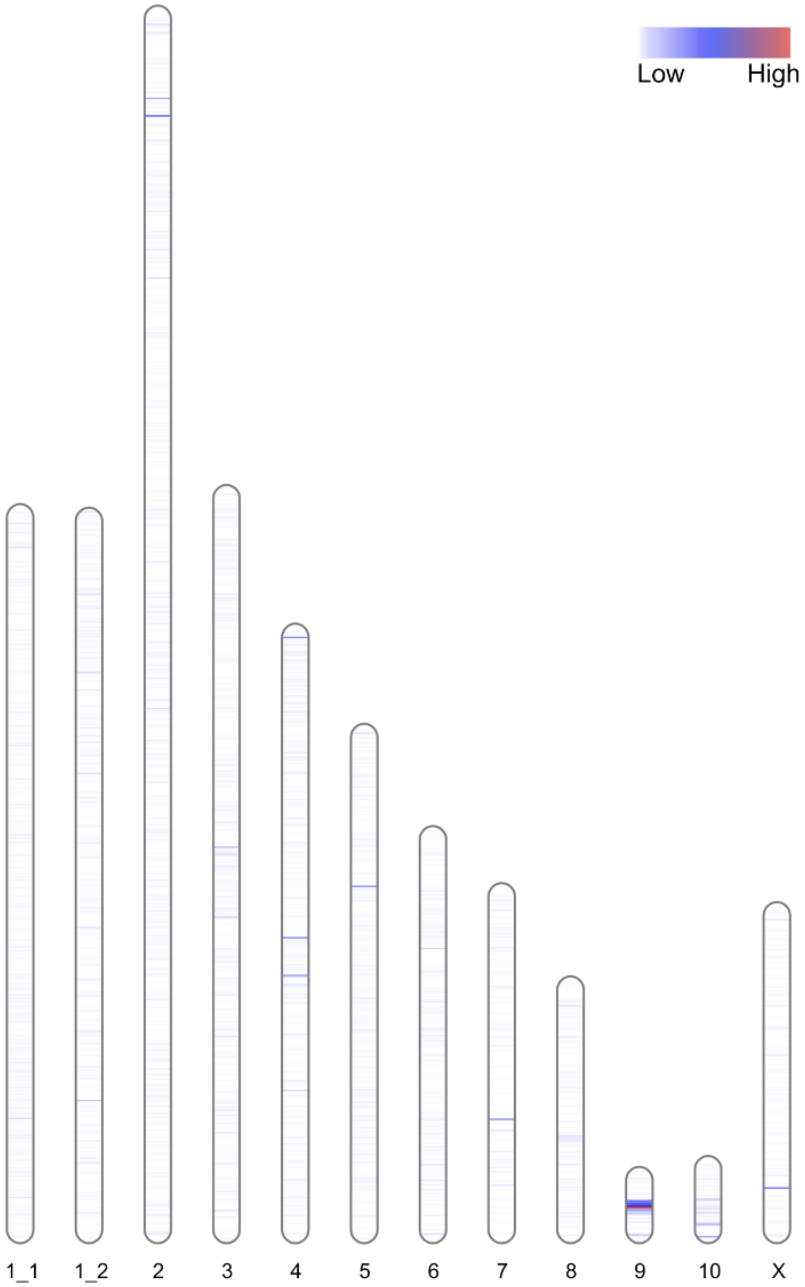


Figure S4.3: VRC01 PDL 70 eccDNA biogenesis map. Chromosome-scale heatmap of eccDNA sequences observed for the VRC01 PDL 70 samples. Frequency of observed eccDNA is shown in color; low (white) to high (red).

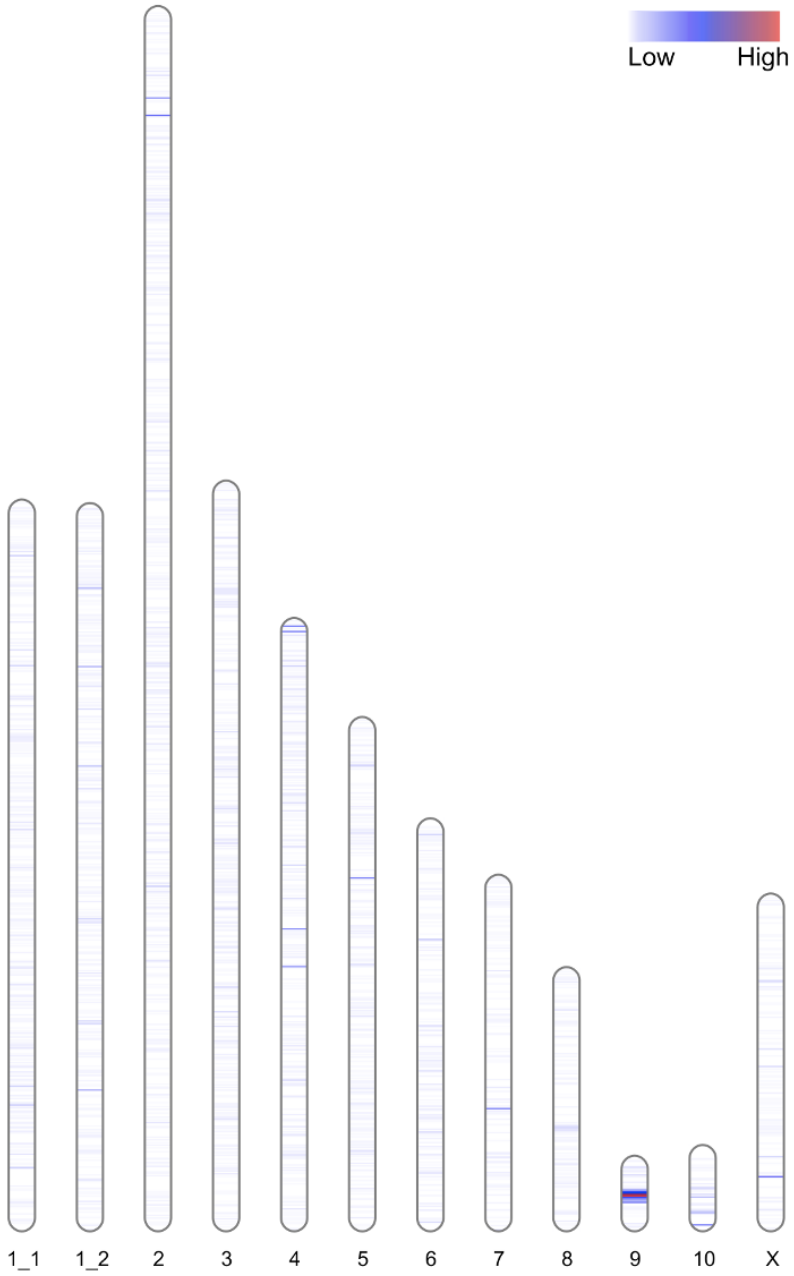


Figure S4.4: VRC01 PDL 110 eccDNA biogenesis map. Chromosome-scale heatmap of eccDNA sequences observed for the VRC01 PDL 110 samples. Frequency of observed eccDNA is shown in color; low (white) to high (red).

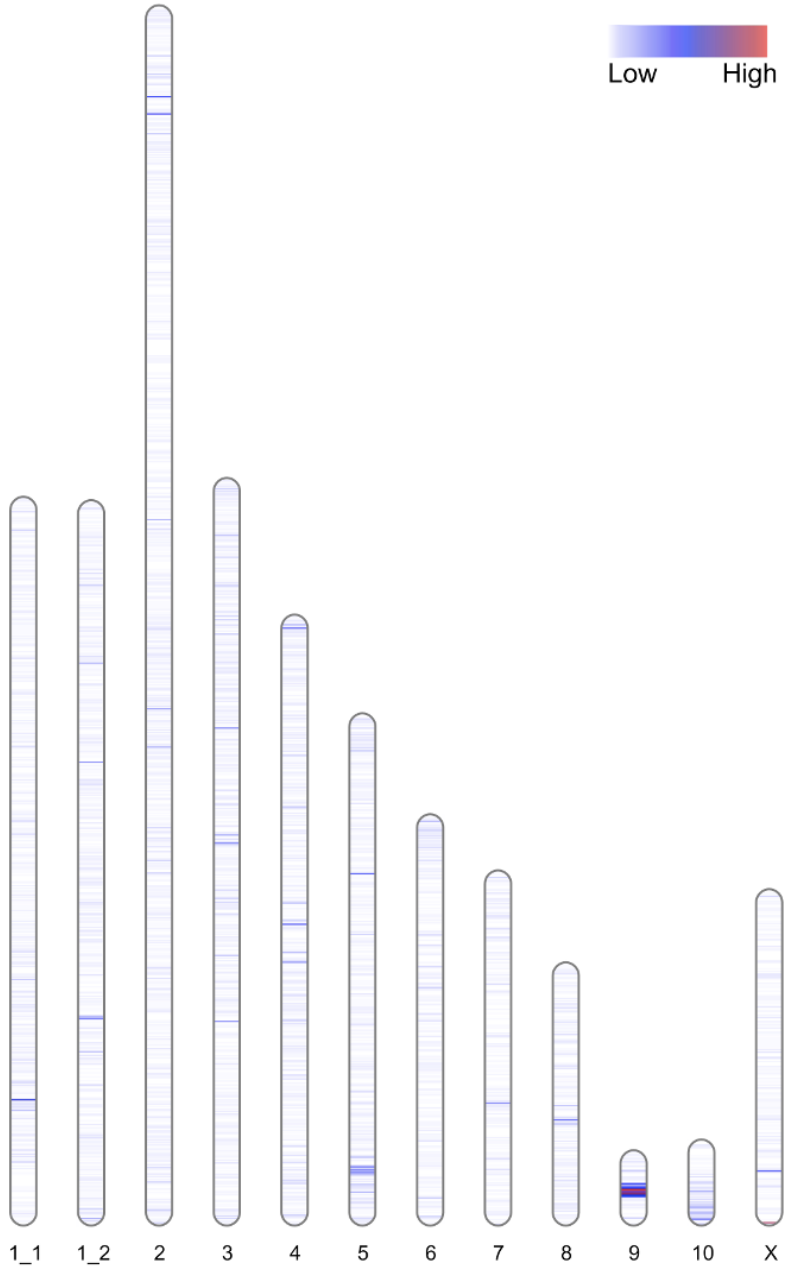


Figure S4.5: CHOZN PDL 20 eccDNA biogenesis map. Chromosome-scale heatmap of eccDNA sequences observed for the CHOZN PDL 20 samples. Frequency of observed eccDNA is shown in color; low (white) to high (red).

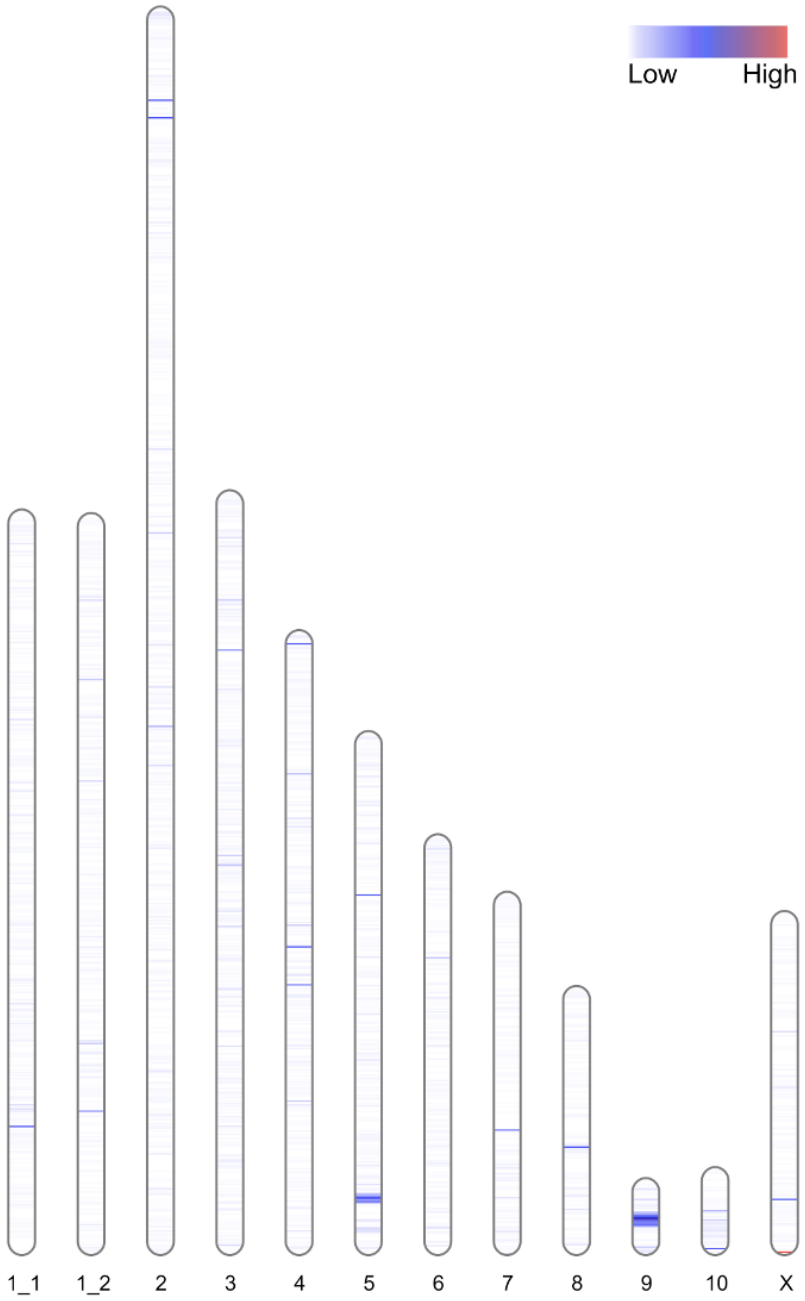


Figure S4.6: CHOZN PDL 35 eccDNA biogenesis map. Chromosome-scale heatmap of eccDNA sequences observed for the CHOZN PDL 35 samples. Frequency of observed eccDNA is shown in color; low (white) to high (red).

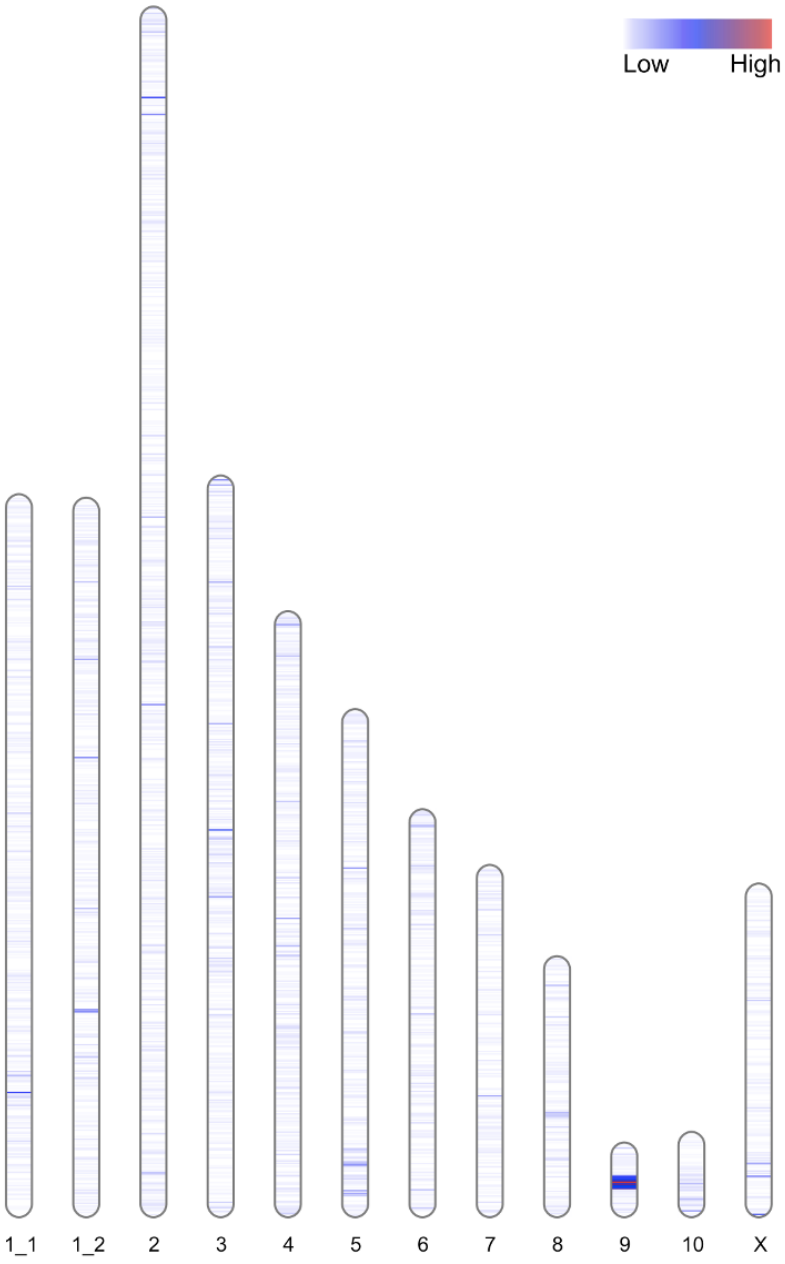


Figure S4.7: CHOZN PDL 90 eccDNA biogenesis map. Chromosome-scale heatmap of eccDNA sequences observed for the CHOZN PDL 90 samples. Frequency of observed eccDNA is shown in color; low (white) to high (red).

Appendix D

Chapter 5 Supplemental Figures

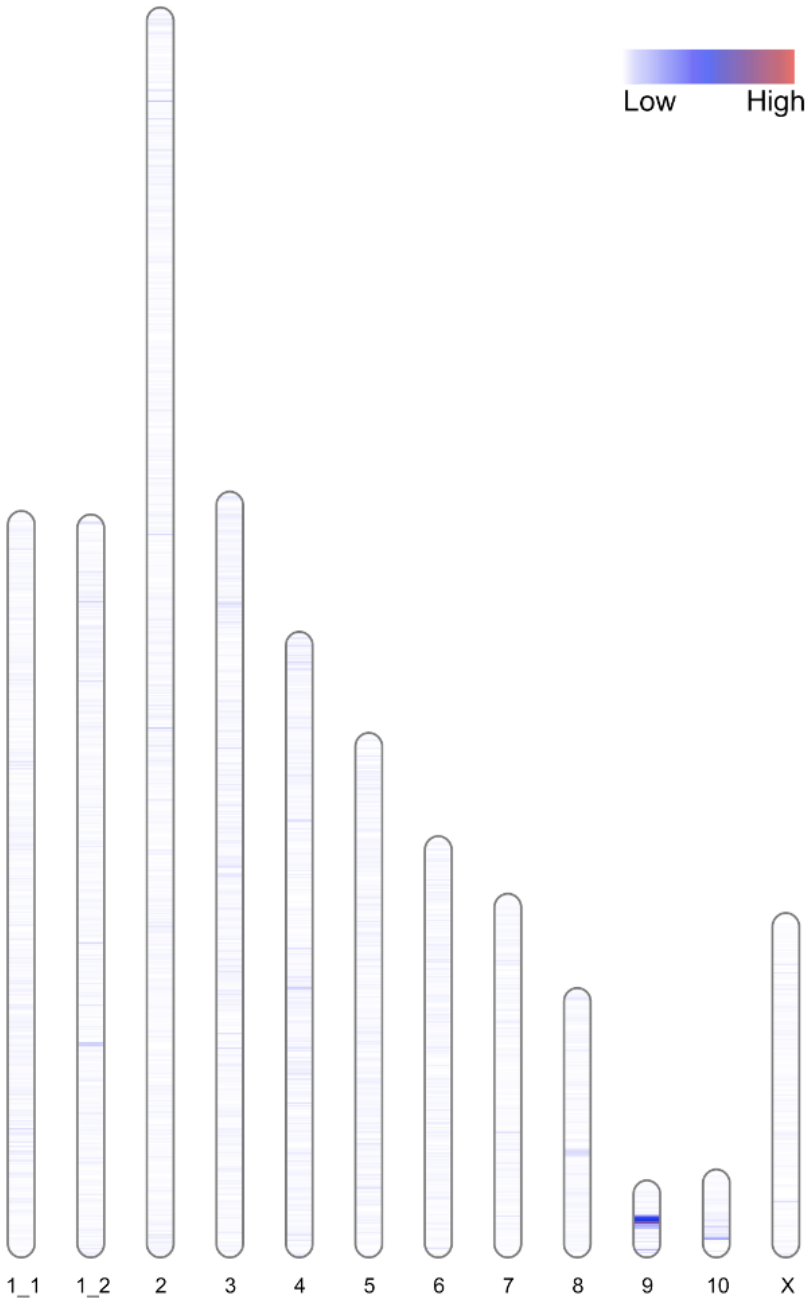


Figure S5.1: Unadapted eccDNA biogenesis map. Chromosome-scale heatmap of eccDNA sequences observed for the unadapted samples. Frequency of observed eccDNA is shown in color; low (white) to high (red).

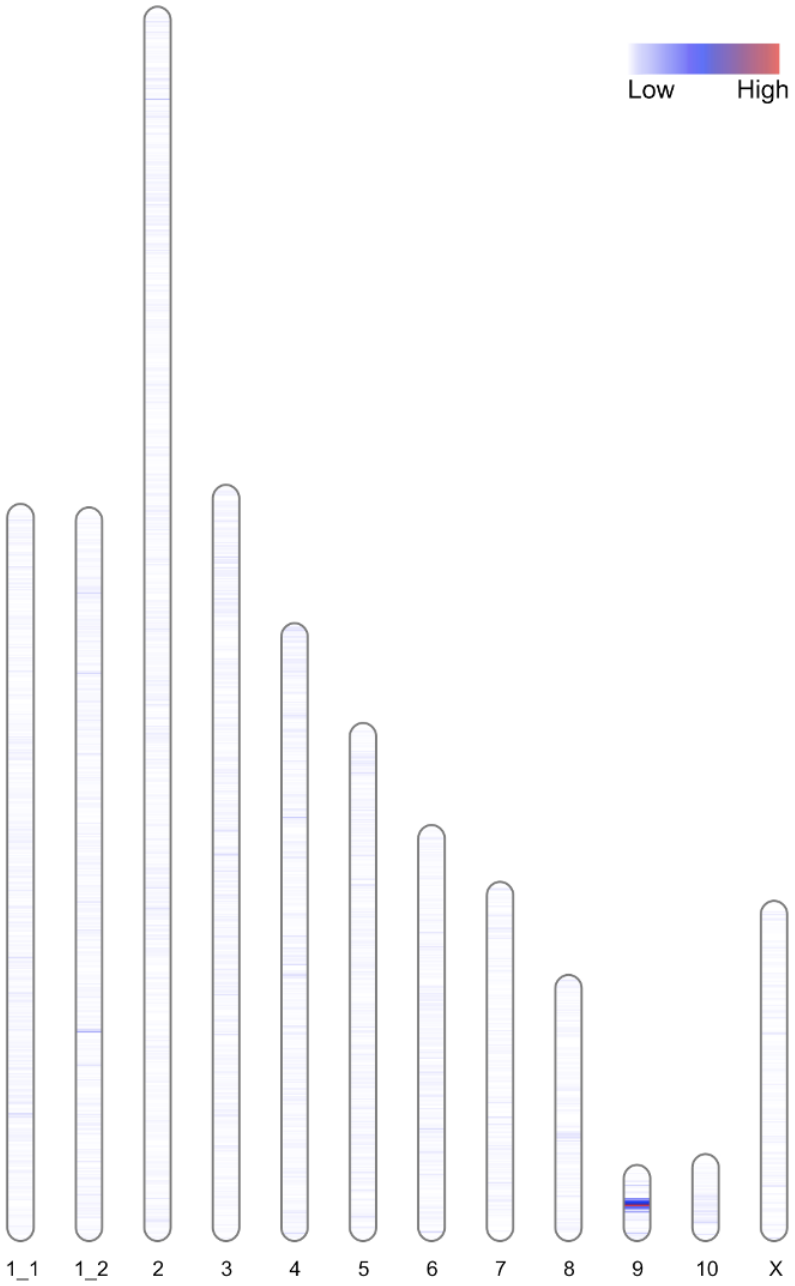


Figure S5.2: 30 mM-adapted eccDNA biogenesis map. Chromosome-scale heatmap of eccDNA sequences observed for the 30 mM-adapted samples. Frequency of observed eccDNA is shown in color; low (white) to high (red).

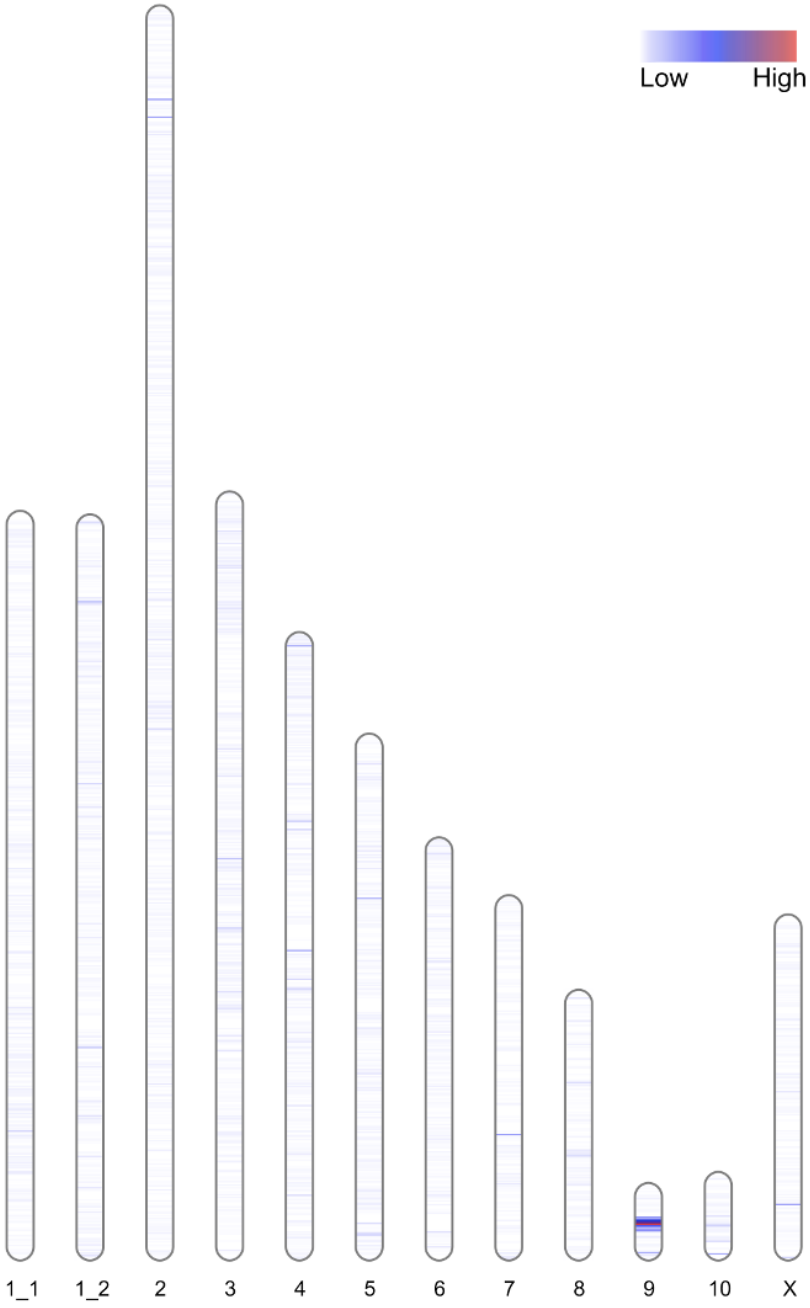


Figure S5.3: 60 mM-adapted eccDNA biogenesis map. Chromosome-scale heatmap of eccDNA sequences observed for the 60 mM-adapted samples. Frequency of observed eccDNA is shown in color; low (white) to high (red).

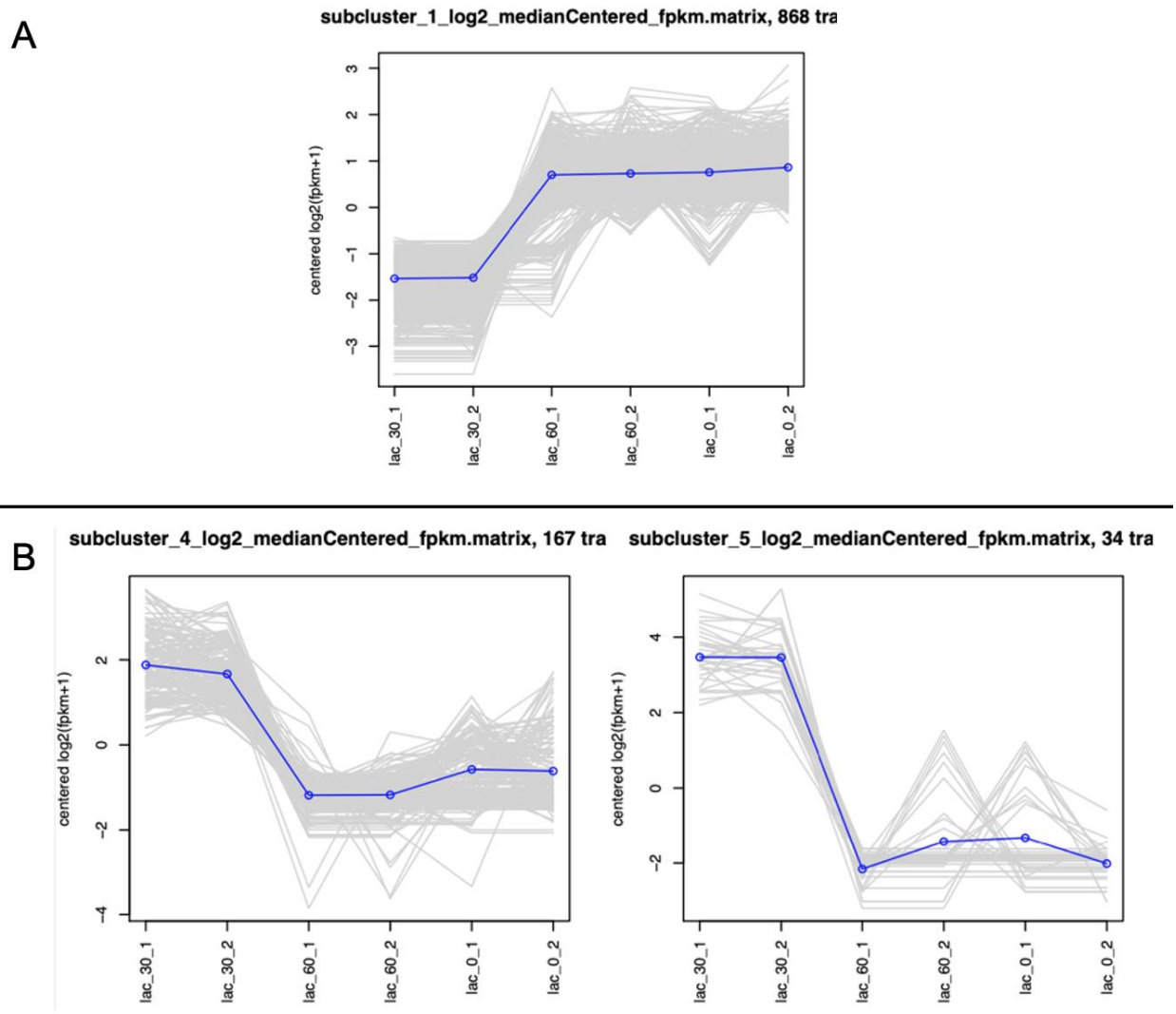


Figure S5.4: edgeR clusters of genes differentially expressed in 30 mM-adapted samples relative to other conditions. A) 868 genes with lower expression in 30 mM-adapgted samples. **B)** 201 genes with greater expression in 30 mM-adapted samples. 30 mM-adapted replicates are shown on the left-most side of the x-axis.

References

1 Walsh, G. & Walsh, E. Biopharmaceutical benchmarks 2022. *Nat Biotechnol* **40**, 1722-1760, doi:10.1038/s41587-022-01582-x (2022).

2 Dahodwala, H. & Lee, K. H. The fickle CHO: a review of the causes, implications, and potential alleviation of the CHO cell line instability problem. *Current opinion in biotechnology* **60**, 128-137, doi:10.1016/j.copbio.2019.01.011 (2019).

3 Wu, S., Bafna, V., Chang, H. Y. & Mischel, P. S. Extrachromosomal DNA: An Emerging Hallmark in Human Cancer. *Annu Rev Pathol* **17**, 367-386, doi:10.1146/annurev-pathmechdis-051821-114223 (2022).

4 Moller, H. D., Parsons, L., Jorgensen, T. S., Botstein, D. & Regenberg, B. Extrachromosomal circular DNA is common in yeast. *Proc Natl Acad Sci U S A* **112**, E3114-3122, doi:10.1073/pnas.1508825112 (2015).

5 Molin, W. T., Yaguchi, A., Blenner, M. & Saski, C. A. The EccDNA Replicon: A Heritable, Extranuclear Vehicle That Enables Gene Amplification and Glyphosate Resistance in Amaranthus palmeri. *Plant Cell* **32**, 2132-2140, doi:10.1105/tpc.20.00099 (2020).

6 Kumar, P. *et al.* Normal and Cancerous Tissues Release Extrachromosomal Circular DNA (eccDNA) into the Circulation. *Mol Cancer Res* **15**, 1197-1205, doi:10.1158/1541-7786.MCR-17-0095 (2017).

7 Stanfield, S. W. & Helinski, D. R. Multiple mechanisms generate extrachromosomal circular DNA in Chinese hamster ovary cells. *Nucleic Acids Res* **14**, 3527-3538, doi:10.1093/nar/14.8.3527 (1986).

8 Zuo, S. *et al.* Extrachromosomal Circular DNA (eccDNA): From Chaos to Function. *Front Cell Dev Biol* **9**, 792555, doi:10.3389/fcell.2021.792555 (2021).

9 Turner, K. M. *et al.* Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. *Nature* **543**, 122-125, doi:10.1038/nature21356 (2017).

10 Mehta, D., Cornet, L., Hirsch-Hoffmann, M., Zaidi, S. S. & Vanderschuren, H. Full-length sequencing of circular DNA viruses and extrachromosomal circular DNA using CIDER-Seq. *Nat Protoc* **15**, 1673-1689, doi:10.1038/s41596-020-0301-0 (2020).

11 Chitwood, D. G. *et al.* Microevolutionary dynamics of eccDNA in Chinese hamster ovary cells grown in fed-batch cultures under control and lactate-stressed conditions. *Sci Rep* **13**, 1200, doi:10.1038/s41598-023-27962-0 (2023).

12 Chitwood, D. G. *et al.* Characterization of metabolic responses, genetic variations, and microsatellite instability in ammonia-stressed CHO cells grown in fed-batch cultures. *BMC Biotechnol* **21**, 4, doi:10.1186/s12896-020-00667-2 (2021).

13 Puck, T. T., Cieciura, S. J. & Robinson, A. Genetics of somatic mammalian cells. III. Long-term cultivation of euploid cells from human and animal subjects. *J Exp Med* **108**, 945-956, doi:10.1084/jem.108.6.945 (1958).

14 Siciliano, M. J., Stallings, R. L., & Adair, G. M. The genetic map of the Chinese hamster and the genetic consequences of chromosomal rearrangements in CHO cells. *Molecular Cell Genetics*, 40 (1985).

15 Wurm, M. J. & Wurm, F. M. Naming CHO cells for bio-manufacturing: Genome plasticity and variant phenotypes of cell populations in bioreactors question the relevance of old names. *Biotechnol J* **16**, e2100165, doi:10.1002/biot.202100165 (2021).

16 Arathoon, W. R. & Birch, J. R. Large-scale cell culture in biotechnology. *Science* **232**, 1390-1395, doi:10.1126/science.2424083 (1986).

17 Capstick, P. B., Telling, R. C., Chapman, W. G. & Stewart, D. L. Growth of a cloned strain of hamster kidney cells in suspended cultures and their susceptibility to the virus of foot-and-mouth disease. *Nature* **195**, 1163-1164, doi:10.1038/1951163a0 (1962).

18 Wurm, F. M. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat Biotechnol* **22**, 1393-1398, doi:10.1038/nbt1026 (2004).

19 Lalonde, M. E. & Durocher, Y. Therapeutic glycoprotein production in mammalian cells. *J Biotechnol* **251**, 128-140, doi:10.1016/j.jbiotec.2017.04.028 (2017).
20 Wong, D. C., Wong, N. S., Goh, J. S., May, L. M. & Yap, M. G. Profiling of N-glycosylation gene expression in CHO cell fed-batch cultures. *Biotechnol Bioeng* **107**, 516-528, doi:10.1002/bit.22828 (2010).
21 Wurm, F. M. & Wurm, M. J. Cloning of CHO Cells, Productivity and Genetic Stability—A Discussion. *Processes* **5**, 20 (2017).
22 Walsh, G. Biopharmaceutical benchmarks 2018. *Nat Biotechnol* **36**, 1136-1145, doi:10.1038/nbt.4305 (2018).
23 Yee, L. & Blanch, H. W. Recombinant protein expression in high cell density fed-batch cultures of Escherichia coli. *Biotechnology (N Y)* **10**, 1550-1556, doi:10.1038/nbt1292-1550 (1992).
24 Schulze, M., Niemann, J., Wijffels, R. H., Matuszczyk, J. & Martens, D. E. Rapid intensification of an established CHO cell fed-batch process. *Biotechnol Prog* **38**, e3213, doi:10.1002/btpr.3213 (2022).
25 Pereira, S., Kildegaard, H. F. & Andersen, M. R. Impact of CHO Metabolism on Cell Growth and Protein Production: An Overview of Toxic and Inhibiting Metabolites and Nutrients. *Biotechnol J* **13**, e1700499, doi:10.1002/biot.201700499 (2018).
26 Dorai, H. *et al.* Early prediction of instability of Chinese hamster ovary cell lines expressing recombinant antibodies and antibody-fusion proteins. *Biotechnol Bioeng* **109**, 1016-1030, doi:10.1002/bit.24367 (2012).
27 McClintock, B. The significance of responses of the genome to challenge. *Science* **226**, 792-801, doi:10.1126/science.15739260 (1984).
28 Aguilera, A. & Garcia-Muse, T. Causes of genome instability. *Annu Rev Genet* **47**, 1-32, doi:10.1146/annurev-genet-111212-133232 (2013).
29 Baik, J. Y. & Lee, K. H. Growth Rate Changes in CHO Host Cells Are Associated with Karyotypic Heterogeneity. *Biotechnol J* **13**, e1700230, doi:10.1002/biot.201700230 (2018).
30 Bandyopadhyay, A. A. *et al.* Recurring genomic structural variation leads to clonal instability and loss of productivity. *Biotechnology and Bioengineering* **116**, 41-53, doi:10.1002/bit.26823 (2019).
31 Bailey, L. A., Hatton, D., Field, R. & Dickson, A. J. Determination of Chinese hamster ovary cell line stability and recombinant antibody expression during long-term culture. *Biotechnol Bioeng* **109**, 2093-2103, doi:10.1002/bit.24485 (2012).
32 Chusainow, J. *et al.* A study of monoclonal antibody-producing CHO cell lines: what makes a stable high producer? *Biotechnol Bioeng* **102**, 1182-1196, doi:10.1002/bit.22158 (2009).
33 Harfe, B. D. & Jinks-Robertson, S. DNA mismatch repair and genetic instability. *Annu Rev Genet* **34**, 359-399, doi:10.1146/annurev.genet.34.1.359 (2000).
34 Fan, L. *et al.* Improving the efficiency of CHO cell line generation using glutamine synthetase gene knockout cells. *Biotechnol Bioeng* **109**, 1007-1015, doi:10.1002/bit.24365 (2012).
35 Kaufman, R. J., Sharp, P. A. & Latt, S. A. Evolution of chromosomal regions containing transfected and amplified dihydrofolate reductase sequences. *Mol Cell Biol* **3**, 699-711, doi:10.1128/mcb.3.4.699-711.1983 (1983).
36 Modrich, P. & Lahue, R. Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annu Rev Biochem* **65**, 101-133, doi:10.1146/annurev.bi.65.070196.000533 (1996).
37 Kim, M., O'Callaghan, P. M., Droms, K. A. & James, D. C. A mechanistic understanding of production instability in CHO cell lines expressing recombinant monoclonal antibodies. *Biotechnol Bioeng* **108**, 2434-2446, doi:10.1002/bit.23189 (2011).

38 Lingg, N., Zhang, P., Song, Z. & Bardor, M. The sweet tooth of biopharmaceuticals: importance of recombinant protein glycosylation analysis. *Biotechnol J* **7**, 1462-1472, doi:10.1002/biot.201200078 (2012).
39 Hogwood, C. E., Bracewell, D. G. & Smales, C. M. Host cell protein dynamics in recombinant CHO cells: impacts from harvest to purification and beyond. *Bioengineered* **4**, 288-291, doi:10.4161/bioe.23382 (2013).
40 Moritz, B., Becker, P. B. & Gopfert, U. CMV promoter mutants with a reduced propensity to productivity loss in CHO cells. *Sci Rep* **5**, 16952, doi:10.1038/srep16952 (2015).
41 Lee, J. S., Kallehauge, T. B., Pedersen, L. E. & Kildegaard, H. F. Site-specific integration in CHO cells mediated by CRISPR/Cas9 and homology-directed DNA repair pathway. *Sci Rep* **5**, 8572, doi:10.1038/srep08572 (2015).
42 Hiller, G. W., Ovalle, A. M., Gagnon, M. P., Curran, M. L. & Wang, W. Cell-controlled hybrid perfusion fed-batch CHO cell process provides significant productivity improvement over conventional fed-batch cultures. *Biotechnol Bioeng* **114**, 1438-1447, doi:10.1002/bit.26259 (2017).
43 Moritz K.F. Wolf, A. P., Veronika Lorenz, Daniel J. Karst, Jonathan Souquet, Hervé Broly, Massimo Morbidelli. A two-step procedure for the design of perfusion bioreactors. *Biochemical Engineering Journal* **151**, doi:https://doi.org/10.1016/j.bej.2019.107295 (2019).
44 Tai, M., Ly, A., Leung, I. & Nayar, G. Efficient high-throughput biological process characterization: Definitive screening design with the ambr250 bioreactor system. *Biotechnol Prog* **31**, 1388-1395, doi:10.1002/btpr.2142 (2015).
45 Langheinrich, C. & Nienow, A. W. Control of pH in large-scale, free suspension animal cell bioreactors: alkali addition and pH excursions. *Biotechnol Bioeng* **66**, 171-179, doi:10.1002/(sici)1097-0290(1999)66:3<171::aid-bit5>3.0.co;2-t (1999).
46 Kim, H. S. & Lee, G. M. Differences in optimal pH and temperature for cell growth and antibody production between two Chinese hamster ovary clones derived from the same parental clone. *J Microbiol Biotechnol* **17**, 712-720 (2007).
47 Handlogten, M. W. *et al.* Intracellular response to process optimization and impact on productivity and product aggregates for a high-titer CHO cell process. *Biotechnol Bioeng* **115**, 126-138, doi:10.1002/bit.26460 (2018).
48 Jiang, R., Chen, H. & Xu, S. pH excursions impact CHO cell culture performance and antibody N-linked glycosylation. *Bioprocess Biosyst Eng* **41**, 1731-1741, doi:10.1007/s00449-018-1996-y (2018).
49 Ivarsson, M., Villiger, T. K., Morbidelli, M. & Soos, M. Evaluating the impact of cell culture process parameters on monoclonal antibody N-glycosylation. *J Biotechnol* **188**, 88-96, doi:10.1016/j.jbiotec.2014.08.026 (2014).
50 Harcum, S. W. *et al.* PID controls: the forgotten bioprocess parameters. *Discover Chemical Engineering* **2**, 1, doi:10.1007/s43938-022-00008-z (2022).
51 Nienow, A. W. Reactor engineering in large scale animal cell culture. *Cytotechnology* **50**, 9-33, doi:10.1007/s10616-006-9005-8 (2006).
52 Becker, M., Junghans, L., Teleki, A., Bechmann, J. & Takors, R. The Less the Better: How Suppressed Base Addition Boosts Production of Monoclonal Antibodies With Chinese Hamster Ovary Cells. *Front Bioeng Biotechnol* **7**, 76, doi:10.3389/fbioe.2019.00076 (2019).
53 Qian, Y. *et al.* Hypoxia influences protein transport and epigenetic repression of CHO cell cultures in shake flasks. *Biotechnol J* **9**, 1413-1424, doi:10.1002/biot.201400315 (2014).
54 A. Converti, C. S., M. Del Borghi & G. Ferraiolo. **The effects of mixing on bioprocesses. Concentration distributions and mechanical shear stress**. *Bioprocess Engineering* **9**, 183-189 (1993).

55 Schneider, M., Marison, I. W. & von Stockar, U. The importance of ammonia in mammalian cell culture. *J Biotechnol* **46**, 161-185, doi:10.1016/0168-1656(95)00196-4 (1996).
56 Chen, P. & Harcum, S. W. Effects of elevated ammonium on glycosylation gene expression in CHO cells. *Metab Eng* **8**, 123-132, doi:10.1016/j.ymben.2005.10.002 (2006).
57 Ozturk, S. S., Riley, M. R. & Palsson, B. O. Effects of ammonia and lactate on hybridoma growth, metabolism, and antibody production. *Biotechnol Bioeng* **39**, 418-431, doi:10.1002/bit.260390408 (1992).
58 Synoground, B. F. *et al.* Transient ammonia stress on Chinese hamster ovary (CHO) cells yield alterations to alanine metabolism and IgG glycosylation profiles. *Biotechnol J* **16**, e2100098, doi:10.1002/biot.202100098 (2021).
59 Wahrheit, J., Nicolae, A. & Heinzle, E. Dynamics of growth and metabolism controlled by glutamine availability in Chinese hamster ovary cells. *Appl Microbiol Biotechnol* **98**, 1771-1783, doi:10.1007/s00253-013-5452-2 (2014).
60 Hartley, F., Walker, T., Chung, V. & Morten, K. Mechanisms driving the lactate switch in Chinese hamster ovary cells. *Biotechnol Bioeng* **115**, 1890-1903, doi:10.1002/bit.26603 (2018).
61 Mulukutla, B. C., Gramer, M. & Hu, W. S. On metabolic shift to lactate consumption in fed-batch culture of mammalian cells. *Metab Eng* **14**, 138-149, doi:10.1016/j.ymben.2011.12.006 (2012).
62 Zagari, F., Jordan, M., Stettler, M., Broly, H. & Wurm, F. M. Lactate metabolism shift in CHO cell culture: the role of mitochondrial oxidative activity. *N Biotechnol* **30**, 238-245, doi:10.1016/j.nbt.2012.05.021 (2013).
63 Altamirano, C., Illanes, A., Becerra, S., Cairo, J. J. & Godia, F. Considerations on the lactate consumption by CHO cells in the presence of galactose. *J Biotechnol* **125**, 547-556, doi:10.1016/j.jbiotec.2006.03.023 (2006).
64 Ghorbaniaghdam, A., Chen, J., Henry, O. & Jolicoeur, M. Analyzing clonal variation of monoclonal antibody-producing CHO cell lines using an in silico metabolomic platform. *PLoS One* **9**, e90832, doi:10.1371/journal.pone.0090832 (2014).
65 Martinez, V. S. *et al.* Flux balance analysis of CHO cells before and after a metabolic switch from lactate production to consumption. *Biotechnol Bioeng* **110**, 660-666, doi:10.1002/bit.24728 (2013).
66 Liste-Calleja, L. *et al.* Lactate and glucose concomitant consumption as a self-regulated pH detoxification mechanism in HEK293 cell cultures. *Appl Microbiol Biotechnol* **99**, 9951-9960, doi:10.1007/s00253-015-6855-z (2015).
67 Zalai, D. *et al.* Combining mechanistic and data-driven approaches to gain process knowledge on the control of the metabolic shift to lactate uptake in a fed-batch CHO process. *Biotechnol Prog* **31**, 1657-1668, doi:10.1002/btpr.2179 (2015).
68 Luo, J. *et al.* Comparative metabolite analysis to understand lactate metabolism shift in Chinese hamster ovary cell culture process. *Biotechnol Bioeng* **109**, 146-156, doi:10.1002/bit.23291 (2012).
69 Cruz, H. J., Freitas, C. M., Alves, P. M., Moreira, J. L. & Carrondo, M. J. Effects of ammonia and lactate on growth, metabolism, and productivity of BHK cells. *Enzyme Microb Technol* **27**, 43-52, doi:10.1016/s0141-0229(00)00151-4 (2000).
70 Lao, M. S. & Toth, D. Effects of ammonium and lactate on growth and metabolism of a recombinant Chinese hamster ovary cell culture. *Biotechnol Prog* **13**, 688-691, doi:10.1021/bp9602360 (1997).
71 Ley, D. *et al.* Reprogramming AA catabolism in CHO cells with CRISPR/Cas9 genome editing improves cell growth and reduces byproduct secretion. *Metab Eng* **56**, 120-129, doi:10.1016/j.ymben.2019.09.005 (2019).

72 Freund, N. W. & Croughan, M. S. A Simple Method to Reduce both Lactic Acid and Ammonium Production in Industrial Animal Cell Culture. *Int J Mol Sci* **19**, doi:10.3390/ijms19020385 (2018).
73 Gagnon, M. *et al.* High-end pH-controlled delivery of glucose effectively suppresses lactate accumulation in CHO fed-batch cultures. *Biotechnol Bioeng* **108**, 1328-1337, doi:10.1002/bit.23072 (2011).
74 Fan, Y. *et al.* Amino acid and glucose metabolism in fed-batch CHO cell culture affects antibody production and glycosylation. *Biotechnol Bioeng* **112**, 521-535, doi:10.1002/bit.25450 (2015).
75 Li, J., Wong, C. L., Vijayasankaran, N., Hudson, T. & Amanullah, A. Feeding lactate for CHO cell culture processes: impact on culture metabolism and performance. *Biotechnol Bioeng* **109**, 1173-1186, doi:10.1002/bit.24389 (2012).
76 Dorner, A. J. *et al.* The stress response in Chinese hamster ovary cells. Regulation of ERp72 and protein disulfide isomerase expression and secretion. *J Biol Chem* **265**, 22029-22034 (1990).
77 Hernandez, I. *et al.* Epigenetic regulation of gene expression in Chinese Hamster Ovary cells in response to the changing environment of a batch culture. *Biotechnol Bioeng* **116**, 677-692, doi:10.1002/bit.26891 (2019).
78 Xu, X. *et al.* The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nat Biotechnol* **29**, 735-741, doi:10.1038/nbt.1932 (2011).
79 Hilliard, W., MacDonald, M. L. & Lee, K. H. Chromosome-scale scaffolds for the Chinese hamster reference genome assembly to facilitate the study of the CHO epigenome. *Biotechnol Bioeng* **117**, 2331-2339, doi:10.1002/bit.27432 (2020).
80 Aguilera, A. & Gomez-Gonzalez, B. Genome instability: a mechanistic view of its causes and consequences. *Nat Rev Genet* **9**, 204-217, doi:10.1038/nrg2268 (2008).
81 Pikor, L., Thu, K., Vucic, E. & Lam, W. The detection and implication of genome instability in cancer. *Cancer Metastasis Rev* **32**, 341-352, doi:10.1007/s10555-013-9429-5 (2013).
82 Nowak, M. A. *et al.* The role of chromosomal instability in tumor initiation. *Proc Natl Acad Sci U S A* **99**, 16226-16231, doi:10.1073/pnas.202617399 (2002).
83 Zheng, J. Oncogenic chromosomal translocations and human cancer (review). *Oncol Rep* **30**, 2011-2019, doi:10.3892/or.2013.2677 (2013).
84 Storchova, Z. & Pellman, D. From polyploidy to aneuploidy, genome instability and cancer. *Nat Rev Mol Cell Biol* **5**, 45-54, doi:10.1038/nrm1276 (2004).
85 Macheret, M. & Halazonetis, T. D. DNA replication stress as a hallmark of cancer. *Annu Rev Pathol* **10**, 425-448, doi:10.1146/annurev-pathol-012414-040424 (2015).
86 Petermann, E., Orta, M. L., Issaeva, N., Schultz, N. & Helleday, T. Hydroxyurea-stalled replication forks become progressively inactivated and require two different RAD51-mediated pathways for restart and repair. *Mol Cell* **37**, 492-502, doi:10.1016/j.molcel.2010.01.021 (2010).
87 Zeman, M. K. & Cimprich, K. A. Causes and consequences of replication stress. *Nat Cell Biol* **16**, 2-9, doi:10.1038/ncb2897 (2014).
88 Hecht, S. S. Tobacco smoke carcinogens and lung cancer. *J Natl Cancer Inst* **91**, 1194-1210, doi:10.1093/jnci/91.14.1194 (1999).
89 Sosa, V. *et al.* Oxidative stress and cancer: an overview. *Ageing Res Rev* **12**, 376-390, doi:10.1016/j.arr.2012.10.004 (2013).
90 Shackelford, R. E., Kaufmann, W. K. & Paules, R. S. Cell cycle control, checkpoint mechanisms, and genotoxic stress. *Environ Health Perspect* **107 Suppl 1**, 5-24, doi:10.1289/ehp.99107s15 (1999).
91 Rodgers, K. & McVey, M. Error-Prone Repair of DNA Double-Strand Breaks. *J Cell Physiol* **231**, 15-24, doi:10.1002/jcp.25053 (2016).

92 Valles, G. J., Bezsonova, I., Woodgate, R. & Ashton, N. W. USP7 Is a Master Regulator of Genome Stability. *Front Cell Dev Biol* **8**, 717, doi:10.3389/fcell.2020.00717 (2020).
93 Przybytkowski, E. *et al.* Chromosome-breakage genomic instability and chromothripsis in breast cancer. *BMC Genomics* **15**, 579, doi:10.1186/1471-2164-15-579 (2014).
94 Narayanan, L., Fritzell, J. A., Baker, S. M., Liskay, R. M. & Glazer, P. M. Elevated levels of mutation in multiple tissues of mice deficient in the DNA mismatch repair gene Pms2. *Proc Natl Acad Sci U S A* **94**, 3122-3127, doi:10.1073/pnas.94.7.3122 (1997).
95 Buermeyer, A. B., Deschenes, S. M., Baker, S. M. & Liskay, R. M. Mammalian DNA mismatch repair. *Annu Rev Genet* **33**, 533-564, doi:10.1146/annurev.genet.33.1.533 (1999).
96 Peltomaki, P. DNA mismatch repair gene mutations in human cancer. *Environ Health Perspect* **105 Suppl 4**, 775-780, doi:10.1289/ehp.105-1470030 (1997).
97 Peltomaki, P. Role of DNA mismatch repair defects in the pathogenesis of human cancer. *J Clin Oncol* **21**, 1174-1179, doi:10.1200/JCO.2003.04.060 (2003).
98 Nakanishi, M., Shimada, M. & Niida, H. Genetic instability in cancer cells by impaired cell cycle checkpoints. *Cancer Sci* **97**, 984-989, doi:10.1111/j.1349-7006.2006.00289.x (2006).
99 Motoyama, N. & Naka, K. DNA damage tumor suppressor genes and genomic instability. *Curr Opin Genet Dev* **14**, 11-16, doi:10.1016/j.gde.2003.12.003 (2004).
100 Yoshida, K. & Miki, Y. Role of BRCA1 and BRCA2 as regulators of DNA repair, transcription, and cell cycle in response to DNA damage. *Cancer Sci* **95**, 866-871, doi:10.1111/j.1349-7006.2004.tb02195.x (2004).
101 Ciccia, A. & Elledge, S. J. The DNA damage response: making it safe to play with knives. *Mol Cell* **40**, 179-204, doi:10.1016/j.molcel.2010.09.019 (2010).
102 Di Micco, R. *et al.* Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication. *Nature* **444**, 638-642, doi:10.1038/nature05327 (2006).
103 Houtgraaf, J. H., Versmissen, J. & van der Giessen, W. J. A concise review of DNA damage checkpoints and repair in mammalian cells. *Cardiovasc Revasc Med* **7**, 165-172, doi:10.1016/j.carrev.2006.02.002 (2006).
104 Vijg, J. & Suh, Y. Genome instability and aging. *Annu Rev Physiol* **75**, 645-668, doi:10.1146/annurev-physiol-030212-183715 (2013).
105 Schmitt, M. W., Prindle, M. J. & Loeb, L. A. Implications of genetic heterogeneity in cancer. *Ann N Y Acad Sci* **1267**, 110-116, doi:10.1111/j.1749-6632.2012.06590.x (2012).
106 Boland, C. R. *et al.* A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res* **58**, 5248-5257 (1998).
107 Vieira, M. L., Santini, L., Diniz, A. L. & Munhoz Cde, F. Microsatellite markers: what they mean and why they are so useful. *Genet Mol Biol* **39**, 312-328, doi:10.1590/1678-4685-GMB-2016-0027 (2016).
108 Gadgil, R., Barthelemy, J., Lewis, T. & Leffak, M. Replication stalling and DNA microsatellite instability. *Biophys Chem* **225**, 38-48, doi:10.1016/j.bpc.2016.11.007 (2017).
109 Kim, T. M. & Park, P. J. A genome-wide view of microsatellite instability: old stories of cancer mutations revisited with new sequencing technologies. *Cancer Res* **74**, 6377-6382, doi:10.1158/0008-5472.CAN-14-1225 (2014).
110 Vilar, E. & Gruber, S. B. Microsatellite instability in colorectal cancer-the stable evidence. *Nat Rev Clin Oncol* **7**, 153-162, doi:10.1038/nrclinonc.2009.237 (2010).

111 Wu, X., Xu, Y., Chai, W. & Her, C. Causal link between microsatellite instability and hMRE11 dysfunction in human cancers. *Mol Cancer Res* **9**, 1443-1448, doi:10.1158/1541-7786.MCR-11-0322 (2011).
112 Muzny, D. M. *et al.* Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330-337, doi:10.1038/nature11252 (2012).
113 Bass, A. J. *et al.* Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513**, 202-209, doi:10.1038/nature13480 (2014).
114 Cancer Genome Atlas Research, N. *et al.* Integrated genomic characterization of endometrial carcinoma. *Nature* **497**, 67-73, doi:10.1038/nature12113 (2013).
115 Li, R., Wang, Y., Li, J. & Zhou, X. Extrachromosomal circular DNA (eccDNA): an emerging star in cancer. *Biomark Res* **10**, 53, doi:10.1186/s40364-022-00399-9 (2022).
116 Cohen, S., Regev, A. & Lavi, S. Small polydispersed circular DNA (spcDNA) in human cells: association with genomic instability. *Oncogene* **14**, 977-985, doi:10.1038/sj.onc.1200917 (1997).
117 Hotta, Y. & Bassel, A. Molecular Size and Circularity of DNA in Cells of Mammals and Higher Plants. *Proc Natl Acad Sci U S A* **53**, 356-362, doi:10.1073/pnas.53.2.356 (1965).
118 Molin, W. T., Yaguchi, A., Blenner, M. & Saski, C. A. Autonomous replication sequences from the Amaranthus palmeri eccDNA replicon enable replication in yeast. *BMC Res Notes* **13**, 330, doi:10.1186/s13104-020-05169-0 (2020).
119 Spier Camposano, H., Molin, W. T. & Saski, C. A. Sequence characterization of eccDNA content in glyphosate sensitive and resistant Palmer amaranth from geographically distant populations. *PLoS One* **17**, e0260906, doi:10.1371/journal.pone.0260906 (2022).
120 Wang, K. *et al.* Deciphering extrachromosomal circular DNA in Arabidopsis. *Comput Struct Biotechnol J* **19**, 1176-1183, doi:10.1016/j.csbj.2021.01.043 (2021).
121 Hull, R. M. *et al.* Transcription-induced formation of extrachromosomal DNA during yeast ageing. *PLoS Biol* **17**, e3000471, doi:10.1371/journal.pbio.3000471 (2019).
122 Cohen, S., Yacobi, K. & Segal, D. Extrachromosomal circular DNA of tandemly repeated genomic sequences in Drosophila. *Genome Res* **13**, 1133-1145, doi:10.1101/gr.907603 (2003).
123 Cohen, Z., Bacharach, E. & Lavi, S. Mouse major satellite DNA is prone to eccDNA formation via DNA Ligase IV-dependent pathway. *Oncogene* **25**, 4515-4524, doi:10.1038/sj.onc.1209485 (2006).
124 Radloff, R., Bauer, W. & Vinograd, J. A dye-buoyant-density method for the detection and isolation of closed circular duplex DNA: the closed circular DNA in HeLa cells. *Proc Natl Acad Sci U S A* **57**, 1514-1521, doi:10.1073/pnas.57.5.1514 (1967).
125 Hull, R. M. & Houseley, J. The adaptive potential of circular DNA accumulation in ageing cells. *Curr Genet* **66**, 889-894, doi:10.1007/s00294-020-01069-9 (2020).
126 Kumar, P. *et al.* ATAC-seq identifies thousands of extrachromosomal circular DNA in cancer and cell lines. *Sci Adv* **6**, eaba2489, doi:10.1126/sciadv.aba2489 (2020).
127 Paulsen, T., Kumar, P., Koseoglu, M. M. & Dutta, A. Discoveries of Extrachromosomal Circles of DNA in Normal and Tumor Cells. *Trends Genet* **34**, 270-278, doi:10.1016/j.tig.2017.12.010 (2018).
128 van Loon, N., Miller, D. & Murnane, J. P. Formation of extrachromosomal circular DNA in HeLa cells by nonhomologous recombination. *Nucleic Acids Res* **22**, 2447-2452, doi:10.1093/nar/22.13.2447 (1994).
129 Zhu, J. *et al.* Molecular characterization of cell-free eccDNAs in human plasma. *Sci Rep* **7**, 10968, doi:10.1038/s41598-017-11368-w (2017).
130 Stanfield, S. W. & Helinski, D. R. Cloning and characterization of small circular DNA from Chinese hamster ovary cells. *Mol Cell Biol* **4**, 173-180, doi:10.1128/mcb.4.1.173 (1984).

131 Neumann, A. A. *et al.* Alternative lengthening of telomeres in normal mammalian somatic cells. *Genes Dev* **27**, 18-23, doi:10.1101/gad.205062.112 (2013).
132 Reon, B. J. & Dutta, A. Biological Processes Discovered by High-Throughput Sequencing. *Am J Pathol* **186**, 722-732, doi:10.1016/j.ajpath.2015.10.033 (2016).
133 Shibata, Y. *et al.* Extrachromosomal microDNAs and chromosomal microdeletions in normal tissues. *Science* **336**, 82-86, doi:10.1126/science.1213307 (2012).
134 Cohen, S., Agmon, N., Sobol, O. & Segal, D. Extrachromosomal circles of satellite repeats and 5S ribosomal DNA in human cells. *Mob DNA* **1**, 11, doi:10.1186/1759-8753-1-11 (2010).
135 Park, P. U., Defossez, P. A. & Guarente, L. Effects of mutations in DNA repair genes on formation of ribosomal DNA circles and life span in Saccharomyces cerevisiae. *Mol Cell Biol* **19**, 3848-3856, doi:10.1128/MCB.19.5.3848 (1999).
136 Cao, X. *et al.* Extrachromosomal Circular DNA: Category, Biogenesis, Recognition, and Functions. *Front Vet Sci* **8**, 693641, doi:10.3389/fvets.2021.693641 (2021).
137 Gibaud, A. *et al.* Extrachromosomal amplification mechanisms in a glioma with amplified sequences from multiple chromosome loci. *Hum Mol Genet* **19**, 1276-1285, doi:10.1093/hmg/ddq004 (2010).
138 Kim, H. *et al.* Extrachromosomal DNA is associated with oncogene amplification and poor outcome across multiple cancers. *Nat Genet*, doi:10.1038/s41588-020-0678-2 (2020).
139 Poirier, E. Z. *et al.* Dicer-2-Dependent Generation of Viral DNA from Defective Genomes of RNA Viruses Modulates Antiviral Immunity in Insects. *Cell Host Microbe* **23**, 353-365 e358, doi:10.1016/j.chom.2018.02.001 (2018).
140 Wang, Y. *et al.* eccDNAs are apoptotic products with high innate immunostimulatory activity. *Nature* **599**, 308-314, doi:10.1038/s41586-021-04009-w (2021).
141 Krolewski, J. J., Schindler, C. W. & Rush, M. G. Structure of extrachromosomal circular DNAs containing both the Alu family of dispersed repetitive sequences and other regions of chromosomal DNA. *J Mol Biol* **174**, 41-54, doi:10.1016/0022-2836(84)90364-4 (1984).
142 Cohen, S. & Segal, D. Extrachromosomal circular DNA in eukaryotes: possible involvement in the plasticity of tandem repeats. *Cytogenet Genome Res* **124**, 327-338, doi:10.1159/000218136 (2009).
143 Lee, J. A., Carvalho, C. M. & Lupski, J. R. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* **131**, 1235-1247, doi:10.1016/j.cell.2007.11.037 (2007).
144 Yang, L. *et al.* Diverse mechanisms of somatic structural variations in human cancer genomes. *Cell* **153**, 919-929, doi:10.1016/j.cell.2013.04.010 (2013).
145 Carroll, S. M. *et al.* Double minute chromosomes can be produced from precursors derived from a chromosomal deletion. *Mol Cell Biol* **8**, 1525-1533, doi:10.1128/mcb.8.4.1525-1533.1988 (1988).
146 Peng, H., Mirouze, M. & Bucher, E. Extrachromosomal circular DNA: A neglected nucleic acid molecule in plants. *Curr Opin Plant Biol* **69**, 102263, doi:10.1016/j.pbi.2022.102263 (2022).
147 Chen, X. *et al.* ATAC-see reveals the accessible genome by transposase-mediated imaging and sequencing. *Nat Methods* **13**, 1013-1020, doi:10.1038/nmeth.4031 (2016).
148 Kazazian, H. H., Jr. & Moran, J. V. Mobile DNA in Health and Disease. *N Engl J Med* **377**, 361-370, doi:10.1056/NEJMra1510092 (2017).
149 Sun, F. J., Fleurdepine, S., Bousquet-Antonelli, C., Caetano-Anolles, G. & Deragon, J. M. Common evolutionary trends for SINE RNA structures. *Trends Genet* **23**, 26-33, doi:10.1016/j.tig.2006.11.005 (2007).

150 Verhaak, R. G. W., Bafna, V. & Mischel, P. S. Extrachromosomal oncogene amplification in tumour pathogenesis and evolution. *Nat Rev Cancer* **19**, 283-288, doi:10.1038/s41568-019-0128-6 (2019).
151 Yan, Y. *et al.* Current understanding of extrachromosomal circular DNA in cancer pathogenesis and therapeutic resistance. *J Hematol Oncol* **13**, 124, doi:10.1186/s13045-020-00960-9 (2020).
152 Von Hoff, D. D., Needham-VanDevanter, D. R., Yucel, J., Windle, B. E. & Wahl, G. M. Amplified human MYC oncogenes localized to replicating submicroscopic circular DNA molecules. *Proc Natl Acad Sci U S A* **85**, 4804-4808, doi:10.1073/pnas.85.13.4804 (1988).
153 Wang, M. *et al.* Extrachromosomal Circular DNAs: Origin, formation and emerging function in Cancer. *Int J Biol Sci* **17**, 1010-1025, doi:10.7150/ijbs.54614 (2021).
154 Qiu, H., Shao, Z. Y., Wen, X. & Zhang, L. Z. New insights of extrachromosomal DNA in tumorigenesis and therapeutic resistance of cancer. *Am J Cancer Res* **10**, 4056-4065 (2020).
155 Wright, S. I., Agrawal, N. & Bureau, T. E. Effects of recombination rate and gene density on transposable element distributions in Arabidopsis thaliana. *Genome Res* **13**, 1897-1903, doi:10.1101/gr.1281503 (2003).
156 Thomas, J. & Pritham, E. J. Helitrons, the Eukaryotic Rolling-circle Transposable Elements. *Microbiol Spectr* **3**, doi:10.1128/microbiolspec.MDNA3-0049-2014 (2015).
157 Kapitonov, V. V. & Jurka, J. Helitrons on a roll: eukaryotic rolling-circle transposons. *Trends Genet* **23**, 521-529, doi:10.1016/j.tig.2007.08.004 (2007).
158 Thomas, J., Phillips, C. D., Baker, R. J. & Pritham, E. J. Rolling-circle transposons catalyze genomic innovation in a mammalian lineage. *Genome Biol Evol* **6**, 2595-2610, doi:10.1093/gbe/evu204 (2014).
159 Jameson, N. *et al.* Helitron mediated amplification of cytochrome P450 monooxygenase gene in maize. *Plant Mol Biol* **67**, 295-304, doi:10.1007/s11103-008-9318-4 (2008).
160 Yona, A. H. *et al.* tRNA genes rapidly change in evolution to meet novel translational demands. *Elife* **2**, e01339, doi:10.7554/eLife.01339 (2013).
161 Torrent, M., Chalancon, G., de Groot, N. S., Wuster, A. & Madan Babu, M. Cells alter their tRNA abundance to selectively regulate protein synthesis during stress conditions. *Sci Signal* **11**, doi:10.1126/scisignal.aat6409 (2018).
162 Bronkhorst, A. J., Ungerer, V. & Holdenrieder, S. The emerging role of cell-free DNA as a molecular marker for cancer management. *Biomol Detect Quantif* **17**, 100087, doi:10.1016/j.bdq.2019.100087 (2019).
163 Peng, L., Zhou, N., Zhang, C. Y., Li, G. C. & Yuan, X. Q. eccDNAdb: a database of extrachromosomal circular DNA profiles in human cancers. *Oncogene* **41**, 2696-2705, doi:10.1038/s41388-022-02286-x (2022).
164 Dong, S., Raffaele, S. & Kamoun, S. The two-speed genomes of filamentous pathogens: waltz with plants. *Curr Opin Genet Dev* **35**, 57-65, doi:10.1016/j.gde.2015.09.001 (2015).
165 Wang, Q. *et al.* Characterization of the Two-Speed Subgenomes of Fusarium graminearum Reveals the Fast-Speed Subgenome Specialized for Adaption and Infection. *Front Plant Sci* **8**, 140, doi:10.3389/fpls.2017.00140 (2017).
166 Joubert, P. M. & Krasileva, K. V. The extrachromosomal circular DNAs of the rice blast pathogen Magnaporthe oryzae contain a wide variety of LTR retrotransposons, genes, and effectors. *BMC Biol* **20**, 260, doi:10.1186/s12915-022-01457-2 (2022).
167 Kasamatsu, H., Robberson, D. L. & Vinograd, J. A novel closed-circular mitochondrial DNA with properties of a replicating intermediate. *Proc Natl Acad Sci U S A* **68**, 2252-2257, doi:10.1073/pnas.68.9.2252 (1971).

168 Taanman, J. W. The mitochondrial genome: structure, transcription, translation and replication. *Biochim Biophys Acta* **1410**, 103-123, doi:10.1016/s0005-2728(98)00161-3 (1999).
169 Feng, W. *et al.* Targeted removal of mitochondrial DNA from mouse and human extrachromosomal circular DNA with CRISPR-Cas9. *Comput Struct Biotechnol J* **20**, 3059-3067, doi:10.1016/j.csbj.2022.06.028 (2022).
170 Moller, H. D. *et al.* CRISPR-C: circularization of genes and chromosome by CRISPR in human cells. *Nucleic Acids Res* **46**, e131, doi:10.1093/nar/gky767 (2018).
171 Zhao, Y., Yu, L., Zhang, S., Su, X. & Zhou, X. Extrachromosomal circular DNA: Current status and future prospects. *Elife* **11**, doi:10.7554/eLife.81412 (2022).
172 Ambinder, R. F., Shah, W. A., Rawlins, D. R., Hayward, G. S. & Hayward, S. D. Definition of the sequence requirements for binding of the EBNA-1 protein to its palindromic target sites in Epstein-Barr virus DNA. *J Virol* **64**, 2369-2379, doi:10.1128/JVI.64.5.2369-2379.1990 (1990).
173 Lee, J. K., Choi, Y. L., Kwon, M. & Park, P. J. Mechanisms and Consequences of Cancer Genome Instability: Lessons from Genome Sequencing Studies. *Annu Rev Pathol* **11**, 283-312, doi:10.1146/annurev-pathol-012615-044446 (2016).
174 Li, H. *et al.* Genetic analysis of the clonal stability of Chinese hamster ovary cells for recombinant protein production. *Mol Biosyst* **12**, 102-109, doi:10.1039/c5mb00627a (2016).
175 Romanova, N. & Noll, T. Engineered and Natural Promoters and Chromatin-Modifying Elements for Recombinant Protein Expression in CHO Cells. *Biotechnol J* **13**, e1700232, doi:10.1002/biot.201700232 (2018).
176 Henson, J. D. *et al.* DNA C-circles are specific and quantifiable markers of alternative-lengthening-of-telomeres activity. *Nat Biotechnol* **27**, 1181-1185, doi:10.1038/nbt.1587 (2009).
177 Fan, Y. *et al.* Frequency of double minute chromosomes and combined cytogenetic abnormalities and their characteristics. *J Appl Genet* **52**, 53-59, doi:10.1007/s13353-010-0007-z (2011).
178 Demeke, M. M., Foulquie-Moreno, M. R., Dumortier, F. & Thevelein, J. M. Rapid evolution of recombinant Saccharomyces cerevisiae for Xylose fermentation through formation of extra-chromosomal circular DNA. *PLoS Genet* **11**, e1005010, doi:10.1371/journal.pgen.1005010 (2015).
179 Cohen, Z., Bacharach, E. & Lavi, S. Mouse major satellite DNA is prone to eccDNA formation via DNA Ligase IV-dependent pathway. *Oncogene* **25**, 4515-4524, doi:10.1038/sj.onc.1209485 (2006).
180 Li, R. M., Wang, Y., Li, J. & Zhou, X. K. Extrachromosomal circular DNA (eccDNA): an emerging star in cancer. *Biomark Res* **10**, doi:ARTN 53
10.1186/s40364-022-00399-9 (2022).
181 Zuo, S. R. *et al.* Extrachromosomal Circular DNA (eccDNA): From Chaos to Function. *Frontiers in Cell and Developmental Biology* **9**, doi:ARTN 792555
10.3389/fcell.2021.792555 (2022).
182 Ain, Q., Schmeer, C., Wengerodt, D., Witte, O. W. & Kretz, A. Extrachromosomal Circular DNA: Current Knowledge and Implications for CNS Aging and Neurodegeneration. *Int J Mol Sci* **21**, doi:10.3390/ijms21072477 (2020).
183 de Nadal, E., Ammerer, G. & Posas, F. Controlling gene expression in response to stress. *Nat Rev Genet* **12**, 833-845, doi:10.1038/nrg3055 (2011).
184 Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658-1659, doi:10.1093/bioinformatics/btl158 (2006).

185 Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150-3152, doi:10.1093/bioinformatics/bts565 (2012).
186 Cantarel, B. L. *et al.* MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res* **18**, 188-196, doi:10.1101/gr.6743907 (2008).
187 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).
188 Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-3100, doi:10.1093/bioinformatics/bty191 (2018).
189 Chan, P. P., Lin, B. Y., Mak, A. J. & Lowe, T. M. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res* **49**, 9077-9096, doi:10.1093/nar/gkab688 (2021).
190 Kanehisa, M. & Sato, Y. KEGG Mapper for inferring cellular functions from protein sequences. *Protein Sci* **29**, 28-35, doi:10.1002/pro.3711 (2020).
191 Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* **2**, 100141, doi:10.1016/j.xinn.2021.100141 (2021).
192 Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284-287, doi:10.1089/omi.2011.0118 (2012).
193 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215**, 403-410, doi:10.1016/S0022-2836(05)80360-2 (1990).
194 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120, doi:10.1093/bioinformatics/btu170 (2014).
195 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).
196 Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323, doi:10.1186/1471-2105-12-323 (2011).
197 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140, doi:10.1093/bioinformatics/btp616 (2010).
198 Koboldt, D. C. *et al.* VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* **25**, 2283-2285, doi:10.1093/bioinformatics/btp373 (2009).
199 Robinson, J. T., Thorvaldsdottir, H., Wenger, A. M., Zehir, A. & Mesirov, J. P. Variant Review with the Integrative Genomics Viewer. *Cancer Res* **77**, e31-e34, doi:10.1158/0008-5472.CAN-17-0337 (2017).
200 Maglott, D., Ostell, J., Pruitt, K. D. & Tatusova, T. Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res* **35**, D26-31, doi:10.1093/nar/gkl993 (2007).
201 Wei, C. H., Allot, A., Leaman, R. & Lu, Z. PubTator central: automated concept annotation for biomedical full text articles. *Nucleic Acids Res* **47**, W587-W593, doi:10.1093/nar/gkz389 (2019).
202 Kwon, D., Kim, S., Wei, C. H., Leaman, R. & Lu, Z. ezTag: tagging biomedical concepts via interactive learning. *Nucleic Acids Res* **46**, W523-W529, doi:10.1093/nar/gky428 (2018).
203 Berardini, T. Z. *et al.* The Gene Ontology in 2010: extensions and refinements The Gene Ontology Consortium. *Nucleic Acids Research* **38**, D331-D335, doi:10.1093/nar/gkp1018 (2010).
204 Delcher, A. L., Salzberg, S. L. & Phillippy, A. M. Using MUMmer to identify similar regions in large sequence sets. *Curr Protoc Bioinformatics* **Chapter 10**, Unit 10 13, doi:10.1002/0471250953.bi1003s00 (2003).

205 Paredes, V., Park, J. S., Jeong, Y., Yoon, J. & Baek, K. Unstable expression of recombinant antibody during long-term culture of CHO cells is accompanied by histone H3 hypoacetylation. *Biotechnol Lett* **35**, 987-993, doi:10.1007/s10529-013-1168-8 (2013).
206 Yang, Y., Mariati, Chusainow, J. & Yap, M. G. DNA methylation contributes to loss in productivity of monoclonal antibody-producing CHO cell lines. *J Biotechnol* **147**, 180-185, doi:10.1016/j.jbiotec.2010.04.004 (2010).
207 Kim, N. S. & Lee, G. M. Response of recombinant Chinese hamster ovary cells to hyperosmotic pressure: effect of Bcl-2 overexpression. *J Biotechnol* **95**, 237-248, doi:10.1016/s0168-1656(02)00011-1 (2002).
208 Ma, N. *et al.* A single nutrient feed supports both chemically defined NS0 and CHO fed-batch processes: Improved productivity and lactate metabolism. *Biotechnol Prog* **25**, 1353-1363, doi:10.1002/btpr.238 (2009).
209 Hauser, H. r. & Wagner, R. *Mammalian cell biotechnology in protein production /editors, Hansjörg Hauser, Roland Wagner*. (Walter de Gruyter, 1997).
210 Klaubert, S. R. *et al.* Method to transfer Chinese hamster ovary (CHO) batch shake flask experiments to large-scale, computer-controlled fed-batch bioreactors. *Methods Enzymol* **660**, 297-320, doi:10.1016/bs.mie.2021.05.005 (2021).
211 Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921, doi:10.1038/35057062 (2001).
212 Charles A. Ishak, D. D. D. C. Reactivation of Endogenous Retroelements in Cancer Development and Therapy. *Annual Review of Cancer Biology* **4**, doi:030419-033525 (2020).
213 Feichtinger, J. *et al.* Comprehensive genome and epigenome characterization of CHO cells in response to evolutionary pressures and over time. *Biotechnol Bioeng* **113**, 2241-2253, doi:10.1002/bit.25990 (2016).
214 Paulsen, T., Malapati, P., Eki, R., Abbas, T. & Dutta, A. EccDNA formation is dependent on MMEJ, repressed by c-NHEJ pathway, and stimulated by DNA double-strand break. *bioRxiv*, 2020.2012.2003.410480, doi:10.1101/2020.12.03.410480 (2020).
215 Stelzer, G. *et al.* The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr Protoc Bioinformatics* **54**, 1 30 31-31 30 33, doi:10.1002/cpbi.5 (2016).
216 Khayami, R., Hashemi, S. R. & Kerachian, M. A. Role of aldo-keto reductase family 1 member B1 (AKR1B1) in the cancer process and its therapeutic potential. *J Cell Mol Med* **24**, 8890-8902, doi:10.1111/jcmm.15581 (2020).
217 Fambrini, M., Usai, G., Vangelisti, A., Mascagni, F. & Pugliesi, C. The plastic genome: The impact of transposable elements on gene functionality and genomic structural variations. *Genesis* **58**, e23399, doi:10.1002/dvg.23399 (2020).
218 Mariati, Koh, E. Y., Yeo, J. H., Ho, S. C. & Yang, Y. Toward stable gene expression in CHO cells. *Bioengineered* **5**, 340-345, doi:10.4161/bioe.32111 (2014).
219 Li, J., Settivari, R. S. & LeBaron, M. J. Genetic instability of in vitro cell lines: Implications for genetic toxicity testing. *Environ Mol Mutagen* **60**, 559-562, doi:10.1002/em.22280 (2019).
220 Frattini, A. *et al.* High variability of genomic instability and gene expression profiling in different HeLa clones. *Sci Rep* **5**, 15377, doi:10.1038/srep15377 (2015).
221 Qian, Y. *et al.* New insights into genetic instability of an industrial CHO cell line by orthogonal omics. *Biochemical Engineering Journal* **164**, 107799, doi:https://doi.org/10.1016/j.bej.2020.107799 (2020).
222 Liao, Z. *et al.* Classification of extrachromosomal circular DNA with a focus on the role of extrachromosomal DNA (ecDNA) in tumor heterogeneity and progression. *Biochim Biophys Acta Rev Cancer* **1874**, 188392, doi:10.1016/j.bbcan.2020.188392 (2020).

223 Wang, T., Zhang, H., Zhou, Y. & Shi, J. Extrachromosomal circular DNA: a new potential role in cancer progression. *J Transl Med* **19**, 257, doi:10.1186/s12967-021-02927-x (2021).
224 Moller, H. D., Ramos-Madrigal, J., Prada-Luengo, I., Gilbert, M. T. P. & Regenberg, B. Near-Random Distribution of Chromosome-Derived Circular DNA in the Condensed Genome of Pigeons and the Larger, More Repeat-Rich Human Genome. *Genome Biol Evol* **12**, 3762-3777, doi:10.1093/gbe/evz281 (2020).
225 Chiu, R. W. K. *et al.* What Is Extrachromosomal Circular DNA and What Does It Do? *Clin Chem* **66**, 754-759, doi:10.1093/clinchem/hvaa096 (2020).
226 Mazzucco, G. *et al.* Telomere damage induces internal loops that generate telomeric circles. *Nat Commun* **11**, 5297, doi:10.1038/s41467-020-19139-4 (2020).
227 Tomaska, L., Nosek, J., Kramara, J. & Griffith, J. D. Telomeric circles: universal players in telomere maintenance? *Nat Struct Mol Biol* **16**, 1010-1015, doi:10.1038/nsmb.1660 (2009).
228 Prada-Luengo, I. *et al.* Replicative aging is associated with loss of genetic heterogeneity from extrachromosomal circular DNA in Saccharomyces cerevisiae. *Nucleic Acids Res* **48**, 7883-7898, doi:10.1093/nar/gkaa545 (2020).
229 Sinclair, D. A. & Guarente, L. Extrachromosomal rDNA circles--a cause of aging in yeast. *Cell* **91**, 1033-1042, doi:10.1016/s0092-8674(00)80493-6 (1997).
230 Cordova, L. T. *et al.* Generation of Reference Cell Lines, Media, and a Process Platform for CHO Cell Biomanufacturing. *Biotechnol Bioeng*, doi:10.1002/bit.28290 (2022).
231 Hamaker, N. K., Min, L. & Lee, K. H. Comprehensive assessment of host cell protein expression after extended culture and bioreactor production of CHO cell lines. *Biotechnol Bioeng* **119**, 2221-2238, doi:10.1002/bit.28128 (2022).
232 Takamatsu, H., Hamamoto, K., Ishimaru, K., Yokoyama, S. & Tokashiki, M. Large-scale perfusion culture process for suspended mammalian cells that uses a centrifuge with multiple settling zones. *Appl Microbiol Biotechnol* **45**, 454-457, doi:10.1007/BF00578455 (1996).
233 Wen, Z. Y., Teng, X. W. & Chen, F. A novel perfusion system for animal cell cultures by two step sequential sedimentation. *J Biotechnol* **79**, 1-11, doi:10.1016/s0168-1656(00)00219-4 (2000).
234 Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674, doi:10.1016/j.cell.2011.02.013 (2011).
235 Preston, B. D., Albertson, T. M. & Herr, A. J. DNA replication fidelity and cancer. *Semin Cancer Biol* **20**, 281-293, doi:10.1016/j.semcancer.2010.10.009 (2010).
236 Hayflick, L. The Limited in Vitro Lifetime of Human Diploid Cell Strains. *Exp Cell Res* **37**, 614-636, doi:10.1016/0014-4827(65)90211-9 (1965).
237 Hayflick, L. & Moorhead, P. S. The serial cultivation of human diploid cell strains. *Exp Cell Res* **25**, 585-621, doi:10.1016/0014-4827(61)90192-6 (1961).
238 Kunisada, T., Yamagishi, H., Ogita, Z., Kirakawa, T. & Mitsui, Y. Appearance of extrachromosomal circular DNAs during in vivo and in vitro ageing of mammalian cells. *Mech Ageing Dev* **29**, 89-99, doi:10.1016/0047-6374(85)90050-8 (1985).
239 Macieira-Coelho, A. Implications of the reorganization of the cell genome for aging or immortalization of dividing cells in vitro. *Gerontology* **26**, 276-282, doi:10.1159/000212428 (1980).
240 Calabretta, B., Robberson, D. L., Barrera-Saldana, H. A., Lambrou, T. P. & Saunders, G. F. Genome instability in a region of human DNA enriched in Alu repeat sequences. *Nature* **296**, 219-225, doi:10.1038/296219a0 (1982).
241 Johnson, R. & Strehler, B. L. Loss of genes coding for ribosomal RNA in ageing brain cells. *Nature* **240**, 412-414, doi:10.1038/240412a0 (1972).

242 Shmookler Reis, R. J. & Goldstein, S. Loss of reiterated DNA sequences during serial passage of human diploid fibroblasts. *Cell* **21**, 739-749, doi:10.1016/0092-8674(80)90437-7 (1980).
243 Shmookler Reis, R. J., Lumpkin, C. K., Jr., McGill, J. R., Riabowol, K. T. & Goldstein, S. Extrachromosomal circular copies of an 'inter-Alu' unstable sequence in human DNA are amplified during in vitro and in vivo ageing. *Nature* **301**, 394-398, doi:10.1038/301394a0 (1983).
244 Maciejowski, J. & de Lange, T. Telomeres in cancer: tumour suppression and genome instability. *Nat Rev Mol Cell Biol* **18**, 175-186, doi:10.1038/nrm.2016.171 (2017).
245 Pickett, H. A., Cesare, A. J., Johnston, R. L., Neumann, A. A. & Reddel, R. R. Control of telomere length by a trimming mechanism that involves generation of t-circles. *EMBO J* **28**, 799-809, doi:10.1038/emboj.2009.42 (2009).
246 Cesare, A. J. & Griffith, J. D. Telomeric DNA in ALT cells is characterized by free telomeric circles and heterogeneous t-loops. *Mol Cell Biol* **24**, 9948-9957, doi:10.1128/MCB.24.22.9948-9957.2004 (2004).
247 Gray, M. D. *et al.* The Werner syndrome protein is a DNA helicase. *Nat Genet* **17**, 100-103, doi:10.1038/ng0997-100 (1997).
248 Frantzeskakis, L., Kusch, S. & Panstruga, R. The need for speed: compartmentalized genome evolution in filamentous phytopathogens. *Mol Plant Pathol* **20**, 3-7, doi:10.1111/mpp.12738 (2019).
249 Wacker, T. *et al.* Two-speed genome evolution drives pathogenicity in fungal pathogens of animals. *Proc Natl Acad Sci U S A* **120**, e2212633120, doi:10.1073/pnas.2212633120 (2023).
250 Nesta, A. V., Tafur, D. & Beck, C. R. Hotspots of Human Mutation. *Trends Genet* **37**, 717-729, doi:10.1016/j.tig.2020.10.003 (2021).
251 Bzymek, M. & Lovett, S. T. Instability of repetitive DNA sequences: the role of replication in multiple mechanisms. *Proc Natl Acad Sci U S A* **98**, 8319-8325, doi:10.1073/pnas.111008398 (2001).
252 Kosiol, C. *et al.* Patterns of positive selection in six Mammalian genomes. *PLoS Genet* **4**, e1000144, doi:10.1371/journal.pgen.1000144 (2008).
253 Schweizer, G. *et al.* Positively Selected Effector Genes and Their Contribution to Virulence in the Smut Fungus Sporisorium reilianum. *Genome Biol Evol* **10**, 629-645, doi:10.1093/gbe/evy023 (2018).
254 Raffaele, S. & Kamoun, S. Genome evolution in filamentous plant pathogens: why bigger can be better. *Nat Rev Microbiol* **10**, 417-430, doi:10.1038/nrmicro2790 (2012).
255 Yousefzadeh, M. *et al.* DNA damage-how and why we age? *Elife* **10**, doi:10.7554/eLife.62852 (2021).
256 D'Angiolella, V. *et al.* Cyclin F-mediated degradation of ribonucleotide reductase M2 controls genome integrity and DNA repair. *Cell* **149**, 1023-1034, doi:10.1016/j.cell.2012.03.043 (2012).
257 Hussain, S. *et al.* F-box only protein 9 and its role in cancer. *Mol Biol Rep* **49**, 1537-1544, doi:10.1007/s11033-021-07057-7 (2022).
258 Fleming, J. D. *et al.* NF-Y coassociates with FOS at promoters, enhancers, repetitive elements, and inactive chromatin regions, and is stereo-positioned with growth-controlling transcription factors. *Genome Res* **23**, 1195-1209, doi:10.1101/gr.148080.112 (2013).
259 Elgaaen, B. V. *et al.* POLD2 and KSP37 (FGFBP2) correlate strongly with histology, stage and outcome in ovarian carcinomas. *PLoS One* **5**, e13837, doi:10.1371/journal.pone.0013837 (2010).

260 Morimachi, M. *et al.* Low expression of DDX5 is associated with poor prognosis in patients with pancreatic ductal adenocarcinoma. *J Clin Pathol* **74**, 741-745, doi:10.1136/jclinpath-2020-207002 (2021).
261 Paulsen, T., Shibata, Y., Kumar, P., Dillon, L. & Dutta, A. Small extrachromosomal circular DNAs, microDNA, produce short regulatory RNAs that suppress gene expression independent of canonical promoters. *Nucleic Acids Res* **47**, 4586-4596, doi:10.1093/nar/gkz155 (2019).
262 Jimenez, N., Martinez, V. S. & Gerdtzen, Z. P. Engineering CHO cells galactose metabolism to reduce lactate synthesis. *Biotechnol Lett* **41**, 779-788, doi:10.1007/s10529-019-02680-8 (2019).
263 Altamirano, C., Paredes, C., Illanes, A., Cairo, J. J. & Godia, F. Strategies for fed-batch cultivation of t-PA producing CHO cells: substitution of glucose and glutamine and rational design of culture medium. *J Biotechnol* **110**, 171-179, doi:10.1016/j.jbiotec.2004.02.004 (2004).
264 Fomina-Yadlin, D. *et al.* Cellular responses to individual amino-acid depletion in antibody-expressing and parental CHO cell lines. *Biotechnol Bioeng* **111**, 965-979, doi:10.1002/bit.25155 (2014).
265 Tandon, I., Pal, R., Pal, J. K. & Sharma, N. K. Extrachromosomal circular DNAs: an extra piece of evidence to depict tumor heterogeneity. *Future Sci OA* **5**, FSO390, doi:10.2144/fsoa-2019-0024 (2019).
266 Hao, Z. *et al.* RIdeogram: drawing SVG graphics to visualize and map genome-wide data on the idiograms. *PeerJ Comput Sci* **6**, e251, doi:10.7717/peerj-cs.251 (2020).
267 Rorbach, J., Nicholls, T. J. & Minczuk, M. PDE12 removes mitochondrial RNA poly(A) tails and controls translation in human mitochondria. *Nucleic Acids Res* **39**, 7750-7763, doi:10.1093/nar/gkr470 (2011).
268 Poulsen, J. B. *et al.* Human 2'-phosphodiesterase localizes to the mitochondrial matrix with a putative function in mitochondrial RNA turnover. *Nucleic Acids Res* **39**, 3754-3770, doi:10.1093/nar/gkq1282 (2011).
269 Ohta, S. *et al.* The protein composition of mitotic chromosomes determined using multiclassifier combinatorial proteomics. *Cell* **142**, 810-821, doi:10.1016/j.cell.2010.07.047 (2010).
270 Dong, C. *et al.* PTBP3 mediates TGF-beta-induced EMT and metastasis of lung adenocarcinoma. *Cell Cycle* **21**, 1406-1421, doi:10.1080/15384101.2022.2052530 (2022).
271 Hou, P. *et al.* PTBP3 contributes to colorectal cancer growth and metastasis via translational activation of HIF-1alpha. *J Exp Clin Cancer Res* **38**, 301, doi:10.1186/s13046-019-1312-y (2019).
272 de Sury, R., Martinez, P., Procaccio, V., Lunardi, J. & Issartel, J. P. Genomic structure of the human NDUFS8 gene coding for the iron-sulfur TYKY subunit of the mitochondrial NADH:ubiquinone oxidoreductase. *Gene* **215**, 1-10, doi:10.1016/s0378-1119(98)00275-3 (1998).
273 Taniue, K. *et al.* RNA Exosome Component EXOSC4 Amplified in Multiple Cancer Types Is Required for the Cancer Cell Survival. *Int J Mol Sci* **23**, doi:10.3390/ijms23010496 (2022).
274 Iness, A. N. *et al.* The cell cycle regulatory DREAM complex is disrupted by high expression of oncogenic B-Myb. *Oncogene* **38**, 1080-1092, doi:10.1038/s41388-018-0490-y (2019).
275 Halestrap, A. P. & Meredith, D. The SLC16 gene family-from monocarboxylate transporters (MCTs) to aromatic amino acid transporters and beyond. *Pflugers Arch* **447**, 619-628, doi:10.1007/s00424-003-1067-2 (2004).

276 Adeva, M., Gonzalez-Lucan, M., Seco, M. & Donapetry, C. Enzymes involved in l-lactate metabolism in humans. *Mitochondrion* **13**, 615-629, doi:10.1016/j.mito.2013.08.011 (2013).
277 Wright, E. M. & Turk, E. The sodium/glucose cotransport family SLC5. *Pflugers Arch* **447**, 510-518, doi:10.1007/s00424-003-1063-6 (2004).
278 Chen, N. H., Reith, M. E. & Quick, M. W. Synaptic uptake and beyond: the sodium- and chloride-dependent neurotransmitter transporter family SLC6. *Pflugers Arch* **447**, 519-531, doi:10.1007/s00424-003-1064-5 (2004).
279 Grewer, C., Gameiro, A. & Rauen, T. SLC1 glutamate transporters. *Pflugers Arch* **466**, 3-24, doi:10.1007/s00424-013-1397-7 (2014).
280 Rincon-Arano, H., Rosales, R., Mora, N., Rodriguez-Castaneda, A. & Rosales, C. R-Ras promotes tumor growth of cervical epithelial cells. *Cancer* **97**, 575-585, doi:10.1002/cncr.11093 (2003).
281 Szanto, I. NADPH Oxidase 4 (NOX4) in Cancer: Linking Redox Signals to Oncogenic Metabolic Adaptation. *Int J Mol Sci* **23**, doi:10.3390/ijms23052702 (2022).
282 Kontorovich, T., Cohen, Y., Nir, U. & Friedman, E. Promoter methylation patterns of ATM, ATR, BRCA1, BRCA2 and p53 as putative cancer risk modifiers in Jewish BRCA1/BRCA2 mutation carriers. *Breast Cancer Res Treat* **116**, 195-200, doi:10.1007/s10549-008-0121-3 (2009).
283 Wu, Y. *et al.* Function of HNRNPC in breast cancer cells by controlling the dsRNA-induced interferon response. *EMBO J* **37**, doi:10.15252/embj.201899017 (2018).
284 Xia, N. *et al.* HNRNPC regulates RhoA to induce DNA damage repair and cancer-associated fibroblast activation causing radiation resistance in pancreatic cancer. *J Cell Mol Med* **26**, 2322-2336, doi:10.1111/jcmm.17254 (2022).
285 Glacken, M. W., Fleischaker, R. J. & Sinskey, A. J. Reduction of waste product excretion via nutrient control: Possible strategies for maximizing product and cell yields on serum in cultures of mammalian cells. *Biotechnol Bioeng* **28**, 1376-1389, doi:10.1002/bit.260280912 (1986).
286 Chang, R. S. & Geyer, R. P. Propagation of conjunctival and HeLa cells in various carbohydrate media. *Proc Soc Exp Biol Med* **96**, 336-340, doi:10.3181/00379727-96-23471 (1957).
287 Altamirano, C., Paredes, C., Cairo, J. J. & Godia, F. Improvement of CHO cell culture medium formulation: simultaneous substitution of glucose and glutamine. *Biotechnol Prog* **16**, 69-75, doi:10.1021/bp990124j (2000).
288 Europa, A. F., Gambhir, A., Fu, P. C. & Hu, W. S. Multiple steady states with distinct cellular metabolism in continuous culture of mammalian cells. *Biotechnol Bioeng* **67**, 25-34, doi:10.1002/(sici)1097-0290(20000105)67:1<25::aid-bit4>3.0.co;2-k (2000).
289 Zhou, W., Rehm, J. & Hu, W. S. High viable cell concentration fed-batch cultures of hybridoma cells through on-line nutrient feeding. *Biotechnol Bioeng* **46**, 579-587, doi:10.1002/bit.260460611 (1995).
290 Yoon, S. K., Choi, S. L., Song, J. Y. & Lee, G. M. Effect of culture pH on erythropoietin production by Chinese hamster ovary cells grown in suspension at 32.5 and 37.0 degrees C. *Biotechnol Bioeng* **89**, 345-356, doi:10.1002/bit.20353 (2005).
291 Li, F., Vijayasankaran, N., Shen, A. Y., Kiss, R. & Amanullah, A. Cell culture processes for monoclonal antibody production. *MAbs* **2**, 466-479, doi:10.4161/mabs.2.5.12720 (2010).
292 Le, H. *et al.* Multivariate analysis of cell culture bioprocess data--lactate consumption as process indicator. *J Biotechnol* **162**, 210-223, doi:10.1016/j.jbiotec.2012.08.021 (2012).
293 Toussaint, C., Henry, O. & Durocher, Y. Metabolic engineering of CHO cells to alter lactate metabolism during fed-batch cultures. *J Biotechnol* **217**, 122-131, doi:10.1016/j.jbiotec.2015.11.010 (2016).

294 Gupta, S. K. *et al.* Metabolic engineering of CHO cells for the development of a robust protein production platform. *PLoS One* **12**, e0181455, doi:10.1371/journal.pone.0181455 (2017).
295 Zhou, M. *et al.* Decreasing lactate level and increasing antibody production in Chinese Hamster Ovary cells (CHO) by reducing the expression of lactate dehydrogenase and pyruvate dehydrogenase kinases. *J Biotechnol* **153**, 27-34, doi:10.1016/j.jbiotec.2011.03.003 (2011).
296 Kim, S. H. & Lee, G. M. Down-regulation of lactate dehydrogenase-A by siRNAs for reduced lactic acid formation of Chinese hamster ovary cells producing thrombopoietin. *Appl Microbiol Biotechnol* **74**, 152-159, doi:10.1007/s00253-006-0654-5 (2007).
297 Kim, S. H. & Lee, G. M. Functional expression of human pyruvate carboxylase for reduced lactic acid formation of Chinese hamster ovary cells (DG44). *Appl Microbiol Biotechnol* **76**, 659-665, doi:10.1007/s00253-007-1041-6 (2007).
298 Noh, S. M., Park, J. H., Lim, M. S., Kim, J. W. & Lee, G. M. Reduction of ammonia and lactate through the coupling of glutamine synthetase selection and downregulation of lactate dehydrogenase-A in CHO cells. *Appl Microbiol Biotechnol* **101**, 1035-1045, doi:10.1007/s00253-016-7876-y (2017).
299 Sadasivam, S. & DeCaprio, J. A. The DREAM complex: master coordinator of cell cycle-dependent gene expression. *Nat Rev Cancer* **13**, 585-595, doi:10.1038/nrc3556 (2013).
300 Ma, J. *et al.* PTBP3 promotes malignancy and hypoxia-induced chemoresistance in pancreatic cancer cells by ATG12 up-regulation. *J Cell Mol Med* **24**, 2917-2930, doi:10.1111/jcmm.14896 (2020).
301 Xie, C. *et al.* PTBP3 modulates P53 expression and promotes colorectal cancer cell proliferation by maintaining UBE4A mRNA stability. *Cell Death Dis* **13**, 128, doi:10.1038/s41419-022-04564-8 (2022).
302 Liang, X. *et al.* Inhibition of polypyrimidine tract-binding protein 3 induces apoptosis and cell cycle arrest, and enhances the cytotoxicity of 5- fluorouracil in gastric cancer cells. *Br J Cancer* **116**, 903-911, doi:10.1038/bjc.2017.32 (2017).
303 Lopez-Maury, L., Marguerat, S. & Bahler, J. Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nat Rev Genet* **9**, 583-593, doi:10.1038/nrg2398 (2008).
304 Macia, J. *et al.* Dynamic signaling in the Hog1 MAPK pathway relies on high basal signal transduction. *Sci Signal* **2**, ra13, doi:10.1126/scisignal.2000056 (2009).
305 Acar, M., Mettetal, J. T. & van Oudenaarden, A. Stochastic switching as a survival strategy in fluctuating environments. *Nat Genet* **40**, 471-475, doi:10.1038/ng.110 (2008).
306 Swindell, W. R., Huebner, M. & Weber, A. P. Plastic and adaptive gene expression patterns associated with temperature stress in Arabidopsis thaliana. *Heredity (Edinb)* **99**, 143-150, doi:10.1038/sj.hdy.6800975 (2007).
307 Carlson, R. Biodesic 2011 Bioeconomy Update. *BioDesic* (2011). <www.biodesic.com/library/Biodesic_2011_Bioeconomy_Update.pdf >.
308 Walsh, G. Biopharmaceutical benchmarks 2014. *Nat Biotechnol* **32**, 992-1000, doi:10.1038/nbt.3040 (2014).
309 Commerce. *The Pharmaceutical and Biotech Industries in the United States*, <http://selectusa.commerce.gov/industry-snapshots/pharmaceutical-and-biotech-industries-united-states> (2014).
310 Walsh, G. Post-translational modifications of protein biopharmaceuticals. *Drug Discov Today* **15**, 773-780, doi:10.1016/j.drudis.2010.06.009 (2010).
311 Betts, Z. & Dickson, A. J. Assessment of UCOE on Recombinant EPO Production and Expression Stability in Amplified Chinese Hamster Ovary Cells. *Mol Biotechnol* **57**, 846-858, doi:10.1007/s12033-015-9877-y (2015).

312 Veith, N., Ziehr, H., MacLeod, R. A. & Reamon-Buettner, S. M. Mechanisms underlying epigenetic and transcriptional heterogeneity in Chinese hamster ovary (CHO) cell lines. *BMC Biotechnol* **16**, 6, doi:10.1186/s12896-016-0238-0 (2016).
313 Kaas, C. S., Kristensen, C., Betenbaugh, M. J. & Andersen, M. R. Sequencing the CHO DXB11 genome reveals regional variations in genomic stability and haploidy. *BMC Genomics* **16**, 160, doi:10.1186/s12864-015-1391-x (2015).
314 Lewis, N. E. *et al.* Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome. *Nat Biotechnol* **31**, 759-765, doi:10.1038/nbt.2624 (2013).
315 Arnold, L., Lee, K., Rucker-Pezzini, J. & Lee, J. H. Implementation of Fully Integrated Continuous Antibody Processing: Effects on Productivity and COGm. *Biotechnol J* **14**, e1800061, doi:10.1002/biot.201800061 (2019).
316 Ha, T. K., Kim, Y. G. & Lee, G. M. Understanding of altered N-glycosylation-related gene expression in recombinant Chinese hamster ovary cells subjected to elevated ammonium concentration by digital mRNA counting. *Biotechnol Bioeng* **112**, 1583-1593, doi:10.1002/bit.25568 (2015).
317 Kim, D. Y. *et al.* Fed-batch CHO cell t-PA production and feed glutamine replacement to reduce ammonia production. *Biotechnol. Prog.* **29**, 165-175, doi:10.1002/btpr.1658 (2013).
318 Pereira, A. G. M., Walther, J. L., Hollenbach, M. & Young, J. D. C-13 Flux Analysis Reveals that Rebalancing Medium Amino Acid Composition can Reduce Ammonia Production while Preserving Central Carbon Metabolism of CHO Cell Cultures. *Biotechnol J.* **13**, doi:10.1002/biot.201700518 (2018).
319 Hause, R. J., Pritchard, C. C., Shendure, J. & Salipante, S. J. Classification and characterization of microsatellite instability across 18 cancer types. *Nat Med* **22**, 1342-1350, doi:10.1038/nm.4191 (2016).
320 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
321 Poplin, R. *et al.* Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*, 201178, doi:10.1101/201178 (2018).
322 Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80-92, doi:10.4161/fly.19695 (2012).
323 Thiel, T., Michalek, W., Varshney, R. K. & Graner, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). *Theor Appl Genet* **106**, 411-422, doi:10.1007/s00122-002-1031-0 (2003).
324 Wei, C. H., Kao, H. Y. & Lu, Z. PubTator: a web-based text mining tool for assisting biocuration. *Nucleic Acids Res* **41**, W518-522, doi:10.1093/nar/gkt441 (2013).
325 Yu, G. clusterProfiler: An universal enrichment tool for functional and comparative study. *bioRxiv*, 256784, doi:10.1101/256784 (2018).
326 Elliott, K. *et al.* Spent media analysis with an integrated CE-MS analyzer of Chinese hamster ovary cells grown in an ammonia-stressed parallel microbioreactor platform. *BioProcess Journal* **19**, doi:doi.org/10.12655 (2020).
327 Fan, Y. *et al.* Amino Acid and Glucose Metabolism in Fed-Batch CHO Cell Culture Affects Antibody Production and Glycosylation. *Biotechnol. Bioeng.* **112**, 521-535, doi:10.1002/bit.25450 (2015).
328 Pereira, S., Kildegaard, H. F. & Andersen, M. R. Impact of CHO Metabolism on Cell Growth and Protein Production: An Overview of Toxic and Inhibiting Metabolites and Nutrients. *Biotechnol J.* **13**, doi:10.1002/biot.201700499 (2018).

329 Brodsky, A. N., Caldwell, M., Bae, S. & Harcum, S. W. Glycosylation-related genes in NS0 cells are insensitive to moderately elevated ammonium concentrations. *J Biotechnol* **187**, 78-86, doi:10.1016/j.jbiotec.2014.07.018 (2014).
330 Freund, N. W. & Croughan, M. S. A Simple Method to Reduce both Lactic Acid and Ammonium Production in Industrial Animal Cell Culture. *Int. J. Mol. Sci.* **19**, doi:10.3390/ijms19020385 (2018).
331 M. Yang, M. B. Effects of Ammonia and Glucosamine on the Heterogeneity of Erythropoietin Glycoforms. *Biotechnol. Prog.* **18**, 129-138 (2002).
332 Genzel, Y., Ritter, J. B., Konig, S., Alt, R. & Reichl, U. Substitution of glutamine by pyruvate to reduce ammonia formation and growth inhibition of mammalian cells. *Biotechnol. Prog.* **21**, 58-69, doi:10.1021/bp049827d (2005).
333 Bartkova, J. *et al.* Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints. *Nature* **444**, 633-637, doi:10.1038/nature05268 (2006).
334 Saitoh, T. & Hiraga, S. Studies on Molecular Mechanism of DNA-Replication in Escherichia-Coli .3. Genetic-Analysis of Mutation Causing Resumption of DNA-Replication Sensitive to Rifampicin, Which Exists in DNA Mutant Defective in Initiation of Replication. *Jpn J Genet* **49**, 320-320 (1975).
335 Ponder, R. G., Fonville, N. C. & Rosenberg, S. M. A switch from high-fidelity to error-prone DNA double-strand break repair underlies stress-induced mutation. *Molecular Cell* **19**, 791-804, doi:10.1016/j.molcel.2005.07.025 (2005).
336 van Deursen, J. M. The role of senescent cells in ageing. *Nature* **509**, 439-446, doi:10.1038/nature13193 (2014).
337 Thompson, L. H. & Schild, D. Homologous recombinational repair of DNA ensures mammalian chromosome stability. *Mutat Res* **477**, 131-153, doi:10.1016/s0027-5107(01)00115-4 (2001).
338 Modrich, P. Mechanisms and biological effects of mismatch repair. *Annu Rev Genet* **25**, 229-253, doi:10.1146/annurev.ge.25.120191.001305 (1991).
339 Sancar, A. & Hearst, J. E. Molecular matchmakers. *Science* **259**, 1415-1420, doi:10.1126/science.8451638 (1993).