AI-Driven Insights into the complexities of Chinese hamster ovary cells death in order to optimize production processes.

Mohammad Moshfeghnia [1] , Hasan Jalili [2] , Sayed-Amir Marashi [3]

1. Department of Life Science Engineering, Faculty of New Sciences and Technologies, University of Tehran, Tehran, Iran. E.mail : m.moshfeghnia77@ut.ac.ir
2. Department of Life Science Engineering, Faculty of New Sciences and Technologies, University of Tehran, Tehran, Iran. E.mail : hjalili@ut.ac.ir
3. Department of Biotechnology, College of Science, University of Tehran, Tehran, Iran. E.mail : marashi@ut.ac.ir

Abstract

Chinese hamster ovary (CHO) cells are a multipurpose and high-performance cell line for recombinant protein production in biopharmaceutical industry. They have proven their ability to produce a wide range of therapeutic proteins with high efficiency and quality. Design novel and high-performance CHO (Chinese Hamster Ovary) cell lines have incredible impact in biopharmaceutical industry that cause reduce prices and increase product efficiency. One of the best ways is to prevent CHO cells death during Bioprocessing. Apoptosis is the most shape of CHO cells death during Bioprocessing. Apoptosis and cell-cycle complex signaling pathways are necessary for the control of cell growth, efficiency, and death in CHO cells. Therefore, understanding the protein interactions of these pathways and their interactions with other cellular processes can help to optimize the performance and quality of CHO cell lines. AI-Driven insights solutions and Advanced machine learning algorithms like GAT (Graph Attention Network) that use in this project indicate most important Targets in these complex signaling pathways. Also Viruses like : Hepatitis C, HIV1 and Bacteria like : Salmonella have High intersection size and Low P-value with these complex signaling pathways. These Microorganisms should be use to design high-performance CHO cell lines because they are master in these pathways. This Method can be use for find novel and high efficiency targets for cure Cancer in Humans.

Introduction

Chinese hamster ovary (CHO) cells are the most widely used mammalian cell line for the production of recombinant protein therapeutics, especially monoclonal antibodies (mAbs). CHO cells have several advantages over other cell lines, such as their ability to grow in suspension culture at high densities, their adaptability to various growth conditions, and their capacity to perform proper folding and post-translational modifications of the expressed proteins, including glycosylation [1]. Glycosylation is a critical quality attribute of many biopharmaceuticals, as it affects their stability, efficacy, and immunogenicity [2]. Therefore, understanding and controlling the glycosylation profile of recombinant proteins produced in CHO cells is essential for ensuring their safety and functionality.

CHO cells have a long history of use in biopharmaceutical industry, dating back to 1987 when the first recombinant therapeutic protein expressed in CHO cells, tissue plasminogen activator (tPA), was approved for clinical use [1]. Since then, many other recombinant proteins have been successfully produced in CHO cells, such as erythropoietin, interferons, growth factors, coagulation factors, and hormones [1]. However, the most prominent application of CHO cells is in the production of mAbs, which account for more than half of all biopharmaceuticals on the market [3]. CHO cells are able to produce mAbs with high yield and quality, as well as with human-like glycosylation patterns [4].

The development of a CHO cell line for recombinant protein production involves several steps, such as transfection of the gene encoding the protein of interest into a host cell line, selection and screening of expressing clones, characterization and optimization of cell culture parameters, scale-up and validation of the production process, and analysis of the product quality [2]. Each step requires careful consideration and optimization to achieve the desired productivity and quality of the recombinant protein. Moreover, each step may introduce genetic and phenotypic variations in the CHO cell line, which may affect its performance and stability over time [2]. Therefore, it is important to monitor and control the genetic and environmental factors that influence the CHO cell line during its development and manufacturing.

CHO cells are not only a valuable tool for biopharmaceutical production, but also a fascinating model system for studying cellular biology and physiology. The genome sequence of CHO cells was published in 2011, revealing insights into their genomic features and evolution [5]. Since then, various omics technologies have been applied to study the transcriptome, proteome, metabolome, and glycome of CHO cells, providing a comprehensive understanding of their molecular mechanisms and pathways [6]. These studies have also enabled the identification of key genes and factors that regulate the expression and glycosylation of recombinant proteins in CHO cells [6]. Furthermore, gene editing technologies such as CRISPR/Cas9 have been used to manipulate the genome of CHO cells to improve their characteristics and performance as biopharmaceutical hosts [7].

In conclusion, CHO cells are a versatile and robust cell line for recombinant protein production in biopharmaceutical industry. They have proven their ability to produce a wide range of therapeutic proteins with high yield and quality. They have also been extensively characterized and engineered to enhance their productivity and quality attributes. As biopharmaceuticals become more complex and diverse, CHO cells will continue to evolve and adapt to meet the challenges and demands of biomanufacturing.

Chinese hamster ovary (CHO) cells are the most widely used mammalian host for the production of recombinant protein therapeutics, accounting for nearly 70% of the market share [8]. CHO cells have several advantages, such as high growth rate, adaptability to various culture conditions, and ability to perform complex post-translational modifications, especially glycosylation [9]. The biopharmaceutical industry is constantly seeking to improve the productivity and quality of CHO-derived products, as well as to reduce the cost and time of development and manufacturing. One of the strategies to achieve these goals is to use continuous culture systems, which offer several benefits over the conventional fed-batch mode, such as higher cell density, lower nutrient and metabolite fluctuations, and reduced process variability [10].

The global market for biopharmaceuticals produced using CHO cells is expected to grow steadily in the next few years, driven by the increasing demand for monoclonal antibodies, biosimilars, and novel biologics for various indications. According to GlobalData, the top 20 biopharmaceutical companies reported a marginal increase in their market capitalization from $3.56 trillion in Q2 2023 to $3.57 trillion in Q3 2023 [11].

CHO (Chinese hamster ovary) cell lines are widely used for the production of recombinant proteins and monoclonal antibodies in biopharmaceutical industry [12]. However, the productivity and quality of these products are often limited by the intrinsic characteristics of CHO cells, such as low specific productivity, inefficient protein folding and secretion, and susceptibility to apoptosis [13]. Therefore, there is a great need to design novel and high-performance CHO cell lines with genetic and protein engineering. Genetic engineering can introduce or modify genes that enhance the expression, folding, secretion, or stability of the target proteins [14]. Protein engineering can modify the structure and function of the target proteins to improve their affinity, specificity, or activity [15]. By combining these two approaches, it is possible to create CHO cell lines that can produce high-quality and high-value biopharmaceuticals with improved efficacy and safety [16].

One of the challenges in biotechnology is to produce high performance CHO (Chinese Hamster Ovary) cell lines that can withstand various stress conditions and maintain high productivity [17]. Apoptosis, or programmed cell death, is a major factor that affects the viability and productivity of CHO cells [18]. Apoptosis can be triggered by two main pathways: the extrinsic pathway and the intrinsic pathway [19]. The extrinsic pathway involves the activation of death receptors on the cell surface, such as Fas, TNFR1, DR3, DR4, and DR5, which belong to the TNF (Tumor Necrosis Factor) gene superfamily [20]. These receptors recruit different signaling molecules that activate caspase-8, a key initiator of apoptosis [21]. The intrinsic pathway is activated by intracellular stress signals, such as growth factor withdrawal, toxin accumulation, or Ca2+ flux, that cause mitochondrial dysfunction and release of pro-apoptotic proteins, such as cytochrome c and Smac/DIABLO [22]. These proteins activate caspase-9, another initiator of apoptosis [23]. Both caspase-8 and caspase-9 then cleave and activate effector caspases, such as caspase-3, -6, and -7, which execute apoptosis by cleaving various cellular substrates [24].

Understanding the molecular mechanisms of apoptosis in CHO cells is crucial for developing strategies to enhance their performance and productivity. By modulating the expression or activity of death receptors, signaling molecules, or caspases, it may be possible to increase the resistance of CHO cells to apoptosis and improve their survival and yield [25]. Therefore, we should focus on death signaling pathways, especially apoptosis, to produce high performance CHO cell lines.

Apoptosis can reduce cell viability, productivity, and product quality, as well as increase the complexity and cost of downstream processing. Therefore, controlling apoptosis is essential for designing high performance CHO cell lines for biopharmaceutical production.[26]

Apoptosis can be triggered by various stresses that cells encounter during bioprocessing, such as hypoxia, shear, nutrient depletion, pH, osmolality, and protein misfolding. These stresses can activate different apoptotic pathways, involving caspases, Bcl-2 family proteins, and other molecules, that lead to cell death. Understanding the mechanisms and regulation of apoptosis in CHO cells can help to identify the key factors that influence cell survival and function.[26]

Several strategies have been proposed to control apoptosis in CHO cell lines, such as optimizing culture media and operating conditions, engineering cell lines to overexpress anti-apoptotic genes or underexpress pro-apoptotic genes, and developing mathematical models to simulate and optimize bioprocesses. These strategies aim to minimize the apoptotic stresses, enhance the cell resistance, and improve the culture performance.[26]

Controlling apoptosis is important not only for increasing the quantity, but also the quality of the bioproducts. Apoptosis can affect the glycosylation, folding, and aggregation of recombinant proteins, which can have implications for their efficacy, safety, and immunogenicity. Therefore, controlling apoptosis can ensure the consistency and quality of the bioproducts, as well as facilitate the quality-by-design approach in bioprocess development.[26]

In conclusion, apoptosis is a critical factor that determines the performance of CHO cell lines in bioprocessing. Controlling apoptosis can improve cell viability, productivity, and product quality, as well as reduce the bioprocess complexity and cost. Therefore, controlling apoptosis is essential for designing high performance CHO cell lines for biopharmaceutical production.[26]

Cell-cycle signaling pathways regulate the growth, proliferation and differentiation of cells, while apoptosis signaling pathways control the programmed cell death or elimination of unwanted cells [27]. Both types of pathways are important for maintaining the homeostasis and quality of CHO cell lines, which are widely used for the production of recombinant proteins and biopharmaceuticals [28]. By analyzing both cell-cycle and apoptosis signaling pathways, we can gain a better understanding of the molecular mechanisms that affect the productivity, stability and robustness of CHO cell lines [29]. Moreover, we can identify potential targets or modulators that can enhance the performance of CHO cell lines by increasing their growth rate, viability, specific productivity and resistance to stress [30]. This can lead to the development of more efficient and cost-effective bioprocesses for the manufacturing of therapeutic proteins and other bioproducts [31].

Apoptosis and cell-cycle signaling pathways are crucial for the regulation of cell growth, differentiation, and death in mammalian cells [32]. These pathways are also involved in the production of recombinant proteins using Chinese hamster ovary (CHO) cells, which are the most widely used host cells for biopharmaceutical manufacturing [33]. Therefore, understanding the molecular mechanisms of these pathways and their interactions with other cellular processes can help to optimize the performance and quality of CHO cell lines. Advanced machine learning algorithms and AI solutions can provide powerful tools to analyze large-scale data sets from omics technologies, such as transcriptomics, proteomics, and metabolomics, that capture the dynamic changes of apoptosis and cell-cycle signaling pathways in response to various stimuli and conditions [34]. By applying these methods, we can gain insights into the complex regulatory networks that govern these pathways and identify key factors that influence their activity and stability. Moreover, we can use these insights to design high-performance CHO cell lines that have enhanced productivity, robustness, and consistency [35].

Machine learning algorithms, especially graph neural networks (GNNs), have shown great potential for analyzing complex biological networks, such as apoptosis and cell-cycle protein-protein interactions (PPIs) [36]. These networks are crucial for understanding the molecular mechanisms of cell growth, survival, and death, which are relevant for designing high-performance CHO cell lines for recombinant protein production [37]. In this paper, we propose GNN-based model that can analyze PPIs from graph structure information. Our model leverages the advantages of representation learning and graph-based deep learning to capture the features and patterns of proteins and their interactions . We show that our model achieves superior performance and can provide insights into the biological roles of proteins and their interactions. This GNN-based model also can be use for recognize most important targets that lead cancer in humans. Actually we can experiments our targets in CHO cell line for find novel Bioengineering method for cure cancer in humans.

Methods

Graph neural networks (GNNs) are a powerful machine learning technique that can learn from both the sequence and the structure of protein-protein interactions (PPIs). GNNs can encode the features of each protein node and its neighbors in a graph, and use graph convolutions to aggregate and propagate information across the network. This way, GNNs can capture the complex and dynamic patterns of PPIs, especially for novel proteins that have not been seen before [38] [39].

Traditional machine learning algorithms, such as support vector machines (SVMs) or random forests (RFs), rely on hand-crafted features or sequence-based embeddings to represent proteins. However, these methods often ignore the structural information of PPI networks, such as their degree, position, and neighboring nodes in a graph. Moreover, these methods may not generalize well to novel proteins that have different features or sequences from the training data [38] [40].

Therefore, we argue that GNNs are a better choice for PPI prediction, especially for inter-novel-protein interactions that involve proteins that are not in the training set. GNNs can learn unknown from correlations, that is, they can infer the interactions of novel proteins based on their similarities and connections with known proteins. GNNs can also handle different types and scales of PPI networks, such as apoptosis and cell-cycle signaling pathways, by adapting their graph convolutions and pooling layers [39].

In conclusion, GNNs are a promising machine learning technique for PPI prediction, as they can leverage both the sequence and the structure of PPI networks, and achieve state-of-the-art performance for inter-novel-protein interactions. We recommend using GNNs instead of traditional machine learning algorithms for PPI analysis.

Graph Neural Networks (GNNs) are a class of artificial neural networks that can process data represented as graphs. Graphs are data structures that consist of nodes (or vertices) and edges (or links) that connect them. Nodes can represent any entity and edges can represent any relationship. Graphs are useful for modeling complex systems with interactions and dependencies.

There are many types of GNNs, but they all share some common characteristics. They all use a process called **message passing**, which means that each node updates its representation (or feature vector) by aggregating information from its neighboring nodes. This way, nodes can learn from their local context and capture the structure of the graph. GNNs also use **graph convolutional layers**, which are analogous to convolutional layers in image processing. Graph convolutional layers apply a shared weight matrix to each node's feature vector and its neighbors' feature vectors, and then apply a non-linear activation function. This allows GNNs to learn node embeddings that are invariant to graph permutations and generalizable to unseen graphs.

One of the most popular types of GNNs is the **Graph Attention Network (GAT)**, which was proposed by Veličković et al. in 2018 [1]. GAT is based on the idea of **self-attention**, which means that each node learns how to assign different importance weights to its neighbors. This way, GAT can focus on the most relevant nodes for each node's representation, and also handle graphs with varying degrees of nodes. GAT uses a multi-head attention mechanism, which means that each node computes multiple attention coefficients for each neighbor, and then concatenates or averages them to form the final node embedding. GAT has been shown to outperform other GNNs on various tasks, such as node classification, link prediction, and graph classification [41].

One of the potential applications of GAT is in bioinformatics, where graphs can be used to model biological molecules or pathways. For example, GAT can be used to analyze apoptosis and cell-cycle signaling pathways, which are networks of proteins that regulate cell death and division. By applying GAT to these pathways, we can identify the most important proteins (or targets) that influence the outcome of these processes. This can help us understand the mechanisms of apoptosis and cell-cycle to control them, and also design new drugs or therapies [42].

A Graph Attention Network (GAT) is a machine learning algorithm that operates on graph-structured data, such as social networks or molecule structures. It uses masked self-attentional layers to learn different weights for different nodes in a neighborhood, without requiring any matrix operations or prior knowledge of the graph structure [43]. The architecture of a GAT consists of stacking multiple attention layers, where each layer computes a linear transformation of the node features, followed by a softmax function to obtain the attention coefficients. The output of each layer is then a weighted sum of the transformed features of the neighboring nodes [43].

One of the strengths of GAT is that it can handle both homogeneous and heterogeneous graphs, meaning that the nodes and edges can have different types and attributes. Another strength is that it can learn interpretable attention patterns, which can reveal how the model assigns importance to different nodes and edges [44]. GAT has been shown to achieve state-of-the-art results on various graph-based tasks, such as node classification, link prediction, graph classification, and knowledge graph completion [43].

GAT can be applied to the problem of finding the most important proteins and targets in apoptosis and cell-cycle signaling pathways, which are two biological processes that regulate cell death and division. Apoptosis and cell-cycle are represented by graphs, where the nodes are proteins and the edges are interactions between them. By using GAT, we can learn the attention coefficients for each node and edge, which indicate how relevant they are for the signaling pathways. This can help us

identify the key proteins and targets that are involved in these processes, and potentially discover new ones [45]. Moreover, GAT can also handle heterogeneous graphs, where the nodes and edges have different types and attributes, such as gene expression levels, protein domains, or interaction types [46]. This can provide more information and insights into the mechanisms and functions of apoptosis and cell-cycle.

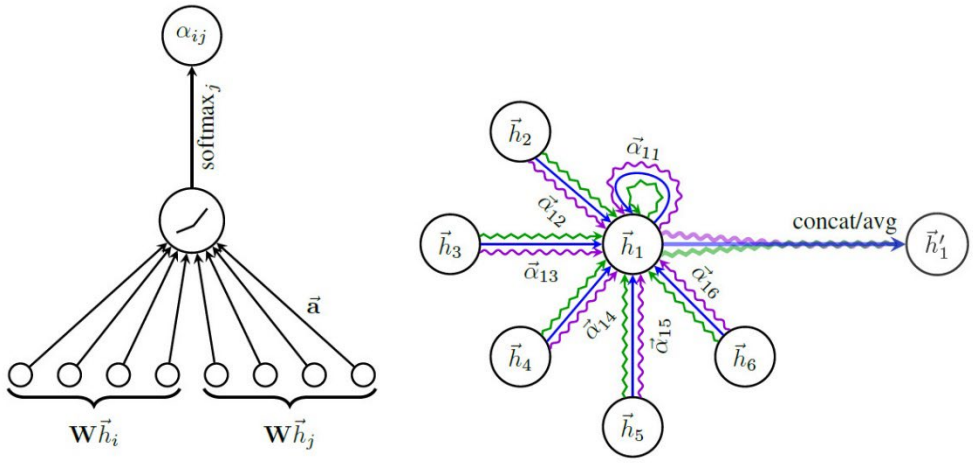
The architecture and attention mechanism of GAT algorithm has been shown in figure1[Veličković et al. (2018). Graph Attention Networks. https://arxiv.org/abs/1710.10903] :

Figure1[Veličković et al. (2018). Graph Attention Networks. https://arxiv.org/abs/1710.10903]

$$e_{ij} = a(\mathbf{W}\vec{h}_i, \mathbf{W}\vec{h}_j) \tag{1}$$

$$\alpha_{ij} = \text{softmax}_j(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k\in\mathcal{N}_i}\exp(e_{ik})}. \tag{2}$$

$$\alpha_{ij} = \frac{\exp\left(\text{LeakyReLU}\left(\vec{\mathbf{a}}^T[\mathbf{W}\vec{h}_i \| \mathbf{W}\vec{h}_j]\right)\right)}{\sum_{k\in\mathcal{N}_i}\exp\left(\text{LeakyReLU}\left(\vec{\mathbf{a}}^T[\mathbf{W}\vec{h}_i \| \mathbf{W}\vec{h}_k]\right)\right)} \tag{3}$$



$$\vec{h}'_i = \sigma\left(\sum_{j\in\mathcal{N}_i}\alpha_{ij}\mathbf{W}\vec{h}_j\right). \tag{4}$$

$$\vec{h}'_i = \mathop{\Big\|}_{k=1}^{K}\sigma\left(\sum_{j\in\mathcal{N}_i}\alpha_{ij}^k\mathbf{W}^k\vec{h}_j\right) \tag{5}$$

$$\vec{h}'_i = \sigma\left(\frac{1}{K}\sum_{k=1}^{K}\sum_{j\in\mathcal{N}_i}\alpha_{ij}^k\mathbf{W}^k\vec{h}_j\right) \tag{6}$$

We use GAT machine learning algorithm for first time to find most important protein targets for design novel CHO cell lines that cause optimize production processes and reduce costs.

Now, Describe main components of this Algorithm:

1. Packages and Libraries:

```
import pandas as pd

import numpy as np

import torch

from torch_geometric.data import Data

from torch.nn import functional as F

from torch_geometric.nn import GATConv

from sklearn.model_selection import LeaveOneOut

import networkx as nx

from torch.optim.lr_scheduler import StepLR
```

Pandas is a popular Python library for data analysis and manipulation, which provides high-performance data structures and operations for working with tabular and time series data [47]. Numpy is another Python library that supports multidimensional arrays and matrices, as well as mathematical functions for linear algebra, statistics, and Fourier transform [48]. Torch is a Python package that provides tensor computations and automatic differentiation for deep learning [49]. Torch-geometric is an extension library for PyTorch that implements geometric deep learning methods for graphs, point clouds, and manifolds [50]. Sklearn is a Python module that offers various tools for machine learning, such as classification, regression, clustering, dimensionality reduction, and model selection [51]. Networkx is a Python package that enables the creation, manipulation, and analysis of complex networks, such as social networks, biological networks, and transportation networks [52].

2. Datasets, Machine and software:

2.1 Softwares : Jetbrains Dataspell [JetBrains DataSpell: The IDE for data analysis] for writing and editing codes / Cytoscape for visualize Algorithm [https://cytoscape.org/] / Sublime text editor for editing codes [Sublime Text - Text Editing, Done Right] / Anaconda platform for perform machine learning Algorithm [Anaconda | The World's Most Popular Data Science Platform].

Datasets : string-db version 12.0 [https://string-db.org/] Apoptosis and Cell-cycle of CHO datasets / KEGG Apoptosis and Cell-cycle pathways, Also use KEGG pathways for enrichment and validation of Algorithm [KEGG: Kyoto Encyclopedia of Genes and Genomes] / Scientific articles about Apoptosis and Cell-cycle in CHO cell line that mention next page.

Machine : Training of Algorithm on Nvidia GeForce RTX 3060 GPU.

3. Main parts of Algorithm :

Importing necessary libraries / load the dataset / get unique node names and their labels / initialize node features with zero arrays / compute edge features and aggregate them to node features / normalize the aggregated features by the node degree (number of edges per node) / Leave-one-out cross-validation / Step the learning rate scheduler / Print the type of the attn tensor / Convert attn to a floating point tensor / Store attention weights to each edge / Taking the mean of all attention weights.

Note : This GAT algorithm is type of Node classification due to the attention weights of each edge.

A Graph Attention Network (GAT) is a type of graph neural network that can perform node classification tasks by learning the attention weights of each edge in the graph [53]. The attention weights indicate how important the features of a neighboring node are for updating the state of a target node. GAT takes as input a graph (namely an edge tensor and a node feature tensor) and outputs updated node states [54]. The node states are, for each target node, neighborhood aggregated information of N-hops (where N is decided by the number of layers of the GAT) [54].

Nodes (Proteins) have three labels: 1.Anti (Anti Apoptosis) 2.Pro (Pro Apoptotic) 3.Unknown (Need Experimental Validation).

scientific articles about apoptosis in CHO cell line that published previously:

1.Modeling apoptosis resistance in CHO cells with CRISPR-mediated knockouts of Bak1, Bax, and Bok [55] / 2. BAK and BAX Deletion Using Zinc-Finger Nucleases Yields Apoptosis-Resistant CHO Cells[56] / 3. Attenuating apoptosis in Chinese hamster ovary cells for improved biopharmaceutical production[57] / 4. Omics driven discoveries of gene targets for apoptosis attenuation in CHO cells[58] / 5. Anti-Cell Death Engineering of CHO Cells: Co-Overexpression of Bcl-2 for Apoptosis Inhibition, Beclin-1 for Autophagy Induction[59] / 6. Apoptosis: A mammalian cell bioprocessing perspective[60] / 7. Cell death in mammalian cell culture: molecular mechanisms and cell line engineering strategies[61] / 8. Life and death in mammalian cell culture: strategies for apoptosis inhibition[62] / 9. Regulating apoptosis in mammalian cell cultures[63] / 10. Bax and Bak knockout apoptosis-resistant Chinese hamster ovary cell lines significantly improve culture viability and titer in intensified fed-batch culture process[64] / 11. A comparison of the properties of a Bcl-xL variant to the wild-type anti-apoptosis inhibitor in mammalian cell cultures[65] / 12. Aven and Bcl-xL Enhance Protection Against Apoptosis for Mammalian Cells Exposed to Various Culture Conditions[66] / 13. Combining Caspase and Mitochondrial Dysfunction Inhibitors of Apoptosis to Limit Cell Death in Mammalian Cell Cultures[67] / 14. Proteomics in biomanufacturing control: Protein dynamics of CHO-K1 cells and conditioned media during apoptosis and necrosis[68] / 15. Inhibiting the Apoptosis Pathway Using MDM2[69] / 16. Molecular Regulation of Cell-Cycle Progression and Apoptosis in Mammalian Cells: Implications for Biotechnology[70] / 17. Overexpression of survivin and cyclin D1 in CHO cells confers apoptosis resistance and enhances growth in serum-free suspension culture[71] / 18. Over-expression of the X-linked inhibitor of apoptosis protein (XIAP) delays serum deprivation-induced apoptosis in CHO-K1 cells[72] / 19. Overexpression of tissue factor pathway inhibitor in CHO-K1 cells results in increased activation of NF-jB and apoptosis mediated by a caspase-3 independent pathway[73] / 20. Overexpression of TLR2 and TLR4 susceptibility to serum deprivation-induced apoptosis in CHO cells[74] / 21. Regulation of Cell Proliferation and Apoptosis in CHO-K1 Cells by the Coexpression of c-Myc and Bcl-2[75] / 22. Transcriptional Profiling of Apoptotic Pathways in Batch and Fed-Batch CHO Cell Cultures[76] / 23. A reference genome of the Chinese hamster based on a hybrid assembly strategy[77] / 24. Proteomic Analysis of Chinese Hamster Ovary Cells[78].

Leave-one-out cross-validation: Leave-one-out cross-validation (LOOCV) is a technique for evaluating the performance of a machine learning model by using each observation in the dataset as a test set once and the rest of the observations as a training set. LOOCV has several advantages over other methods of cross-validation, such as k-fold or validation-set methods.

First, LOOCV provides a less biased estimate of the test mean squared error (MSE) because it uses almost all the available data for training, thus reducing the impact of randomness in the data partitioning [79]. Second, LOOCV tends to not overestimate the test MSE as much as other methods, especially when the dataset is small or when the model is complex [80]. Third, LOOCV can be useful when there is limited labeled data available, as it does not require a separate validation set and uses all the data points for both training and testing [81].

Results

Apoptosis Mean accuracy : 87% / Cell-cycle Mean accuracy : 97%

Algorithm Visualizations: Cytoscape use [https://cytoscape.org/] for visualize algorithm. Visualization have two parts: 1. Apoptosis 2. Cell-cycle . yFiles Hierarchic Layout [yFiles Layout Algorithms for Cytoscape (yworks.com)] is use for Layout.

1.Apoptosis

. Summary statistics

Number of nodes was 92 and Number of edges was 617. Also Avg.number of neighbors was 13.4.

Figure 2. Most important nodes (Proteins) in Apoptosis pathway

1	name	label	Degree	Radiality	Stress
2	Tp53	Unknown	40	0.98543956	6628
3	Bcl2	Anti	37	0.982967033	5994
4	Akt1	Anti	34	0.983516484	5682
5	Casp7	Pro	34	0.983516484	5550
6	Bcl2l1	Unknown	37	0.983241758	5466
7	Xiap	Anti	35	0.982142857	4238
8	Tnf	Unknown	33	0.983241758	3864
9	Casp8	Pro	34	0.983241758	3532
10	Bid	Pro	32	0.981043956	3250
11	Traf2	Unknown	28	0.98021978	2712
12	Casp2	Unknown	27	0.978846154	2692
13	Actb	Unknown	16	0.975	2630
14	I79_008519	Unknown	17	0.972802198	2116
15	Hras	Unknown	19	0.977472527	1920
16	Akt2	Anti	20	0.977747253	1894
17	Cflar	Unknown	25	0.978296703	1668
18	Mapk3	Unknown	20	0.978021978	1532
19	Akt3	Anti	18	0.976923077	1528
20	Fadd	Anti	27	0.97967033	1386

Figure 3. Most important edges (Links) in Apoptosis pathway.

1	name	weight
2	Septin4 (-) Xiap	0.53135145
3	I79_005999 (-) Sptan1	0.5290915
4	I79_004741 (-) I79_008519	0.33600506
5	Dab2ip (-) Traf2	0.26518193
6	Prf1 (-) Tnf	0.2547996
7	Tnfrsf1a (-) Gadd45b	0.2543058
8	Kras (-) I79_025313	0.25319237
9	Gadd45b (-) Gadd45g	0.25160906
10	I79_025313 (-) Nras	0.24902438
11	Eif2s1 (-) I79_009953	0.24457493
12	Tp53 (-) Gadd45g	0.24390636
13	Tp53 (-) Gadd45b	0.24191119
14	Dab2ip (-) Pik3ca	0.22843286
15	Rela (-) Ngf	0.21615309
16	Pik3ca (-) Ngf	0.18924475
17	Pik3r1 (-) Ngf	0.18743424
18	Ddit3 (-) Eif2s1	0.17340614
19	Fos (-) Ddit3	0.16812572
20	Parp3 (-) Tp53	0.16573964

Note: g_scs Significance Threshold Method is use for functional enrichment analysis.

The g_scs Significance Threshold Method is a method for correcting multiple testing errors in functional enrichment analysis [82]. It is based on the Strong Control of the Generalized Family-wise Error Rate (gFWER), which is the probability of finding at least one false positive result among all significant results [83]. The g_scs method adjusts the p-values of each functional term by taking into account the hierarchical structure of the term ontology and the correlation between terms [84].

One of the strengths of this method is that it provides a better balance between sensitivity and specificity than other methods such as Bonferroni correction or false discovery rate (FDR) [84]. It also avoids over-penalizing terms that are highly dependent on each other, such as parent-child terms in the Gene Ontology [84]. Another strength of this method is that it can handle both flat and ordered gene lists, which allows for more flexible and informative analysis of gene expression data [85].

Figure 4. Enrichment analysis of Apoptosis pathway

term id	term name	intersection size	p-value
KEGG:04210	Apoptosis	55	5.68E-109
KEGG:04668	TNF signaling pathway	21	9.17E-25
KEGG:05160	Hepatitis C	24	1.16E-24
KEGG:05170	Human immunodeficiency virus 1 infection	24	1.12E-22
KEGG:05163	Human cytomegalovirus infection	23	2.45E-20
KEGG:05161	Hepatitis B	20	3.91E-19
KEGG:05167	Kaposi sarcoma-associated herpesvirus infection	21	5.79E-19
KEGG:05132	Salmonella infection	22	6.46E-18
KEGG:05169	Epstein-Barr virus infection	20	1.24E-17
KEGG:05164	Influenza A	19	2.05E-17
KEGG:04071	Sphingolipid signaling pathway	16	6.04E-16
KEGG:04620	Toll-like receptor signaling pathway	15	7.26E-16
KEGG:05166	Human T-cell leukemia virus 1 infection	19	2.54E-15
KEGG:04064	NF-kappa B signaling pathway	14	1.88E-14
KEGG:04140	Autophagy - animal	16	3.40E-14
KEGG:04010	MAPK signaling pathway	19	1.21E-13
KEGG:05200	Pathways in cancer	24	1.44E-13
KEGG:04215	Apoptosis - multiple species	10	3.14E-13
KEGG:01521	EGFR tyrosine kinase inhibitor resistance	12	1.46E-12
KEGG:05165	Human papillomavirus infection	19	3.27E-12
KEGG:04068	FoxO signaling pathway	14	6.88E-12
KEGG:05168	Herpes simplex virus 1 infection	17	1.64E-11
KEGG:05203	Viral carcinogenesis	15	3.64E-11
KEGG:04370	VEGF signaling pathway	10	4.16E-11
KEGG:05135	Yersinia infection	13	1.22E-10
GO:0031264	death-inducing signaling complex	5	1.95E-10
GO:0005829	cytosol	36	2.04E-10
KEGG:04218	Cellular senescence	13	6.26E-10
KEGG:04217	Necroptosis	13	9.92E-10
GO:0005737	cytoplasm	58	1.51E-09
KEGG:04014	Ras signaling pathway	14	8.74E-09
GO:0097136	Bcl-2 family protein complex	5	1.08E-08
KEGG:04150	mTOR signaling pathway	12	4.28E-08
GO:0097342	ripoptosome	4	2.56E-07
KEGG:04151	PI3K-Akt signaling pathway	15	2.97E-07
KEGG:04211	Longevity regulating pathway	9	3.57E-07
KEGG:04540	Gap junction	8	1.41E-06
KEGG:04912	GnRH signaling pathway	8	1.90E-06
GO:0043231	intracellular membrane-bounded organelle	61	2.51E-06

KEGG:04015	Rap1 signaling pathway	11	4.30E-06
KEGG:05206	MicroRNAs in cancer	11	5.28E-06
KEGG:04213	Longevity regulating pathway - multiple species	7	5.98E-06
GO:0031265	CD95 death-inducing signaling complex	3	1.33E-05
GO:0043227	membrane-bounded organelle	61	1.51E-05
KEGG:04115	p53 signaling pathway	7	3.47E-05
GO:0005943	phosphatidylinositol 3-kinase complex, class IA	3	1.32E-04
GO:0043229	intracellular organelle	63	2.18E-04
GO:0005622	intracellular anatomical structure	66	3.08E-04
KEGG:04630	JAK-STAT signaling pathway	8	4.63E-04
KEGG:04024	cAMP signaling pathway	9	4.77E-04
GO:0097651	phosphatidylinositol 3-kinase complex, class I	3	7.34E-04
GO:0043226	organelle	63	7.39E-04
KEGG:04623	Cytosolic DNA-sensing pathway	6	9.33E-04
GO:0035631	CD40 receptor complex	3	0.001098508
KEGG:04141	Protein processing in endoplasmic reticulum	8	0.001100745
KEGG:04371	Apelin signaling pathway	7	0.001879085
GO:0098796	membrane protein complex	14	0.00225202
KEGG:04142	Lysosome	7	0.002931519
KEGG:04066	HIF-1 signaling pathway	6	0.003202476
GO:0097145	BAK complex	2	0.003401888
GO:0002947	tumor necrosis factor receptor superfamily complex	2	0.003401888
GO:0005942	phosphatidylinositol 3-kinase complex	3	0.005847945
GO:0098552	side of membrane	8	0.008587009
KEGG:05171	Coronavirus disease - COVID-19	9	0.009495592
GO:0005739	mitochondrion	15	0.012167616
GO:0012505	endomembrane system	24	0.012185461
GO:0046930	pore complex	3	0.014441926
GO:0005741	mitochondrial outer membrane	5	0.017177606
GO:0043235	receptor complex	7	0.017611821
GO:0031968	organelle outer membrane	5	0.029126218
GO:0019867	outer membrane	5	0.029126218
GO:0005635	nuclear envelope	7	0.044481413
GO:0005634	nucleus	42	0.045632414
GO:0031975	envelope	11	0.047091571
GO:0031967	organelle envelope	11	0.047091571

Figure 5. Hierarchic Layout of Apoptosis pathway

Note: Edges with higher weights have higher Thickness and higher Color contrast.

Nodes have three kind of color that indicate Enrichment analysis:

1. Apoptosis : #a6cee3 (Color ID)

2. TNF signaling pathway : #1f78b4 (Color ID)

3. Hepatitis C : #b2df8a (Color ID)

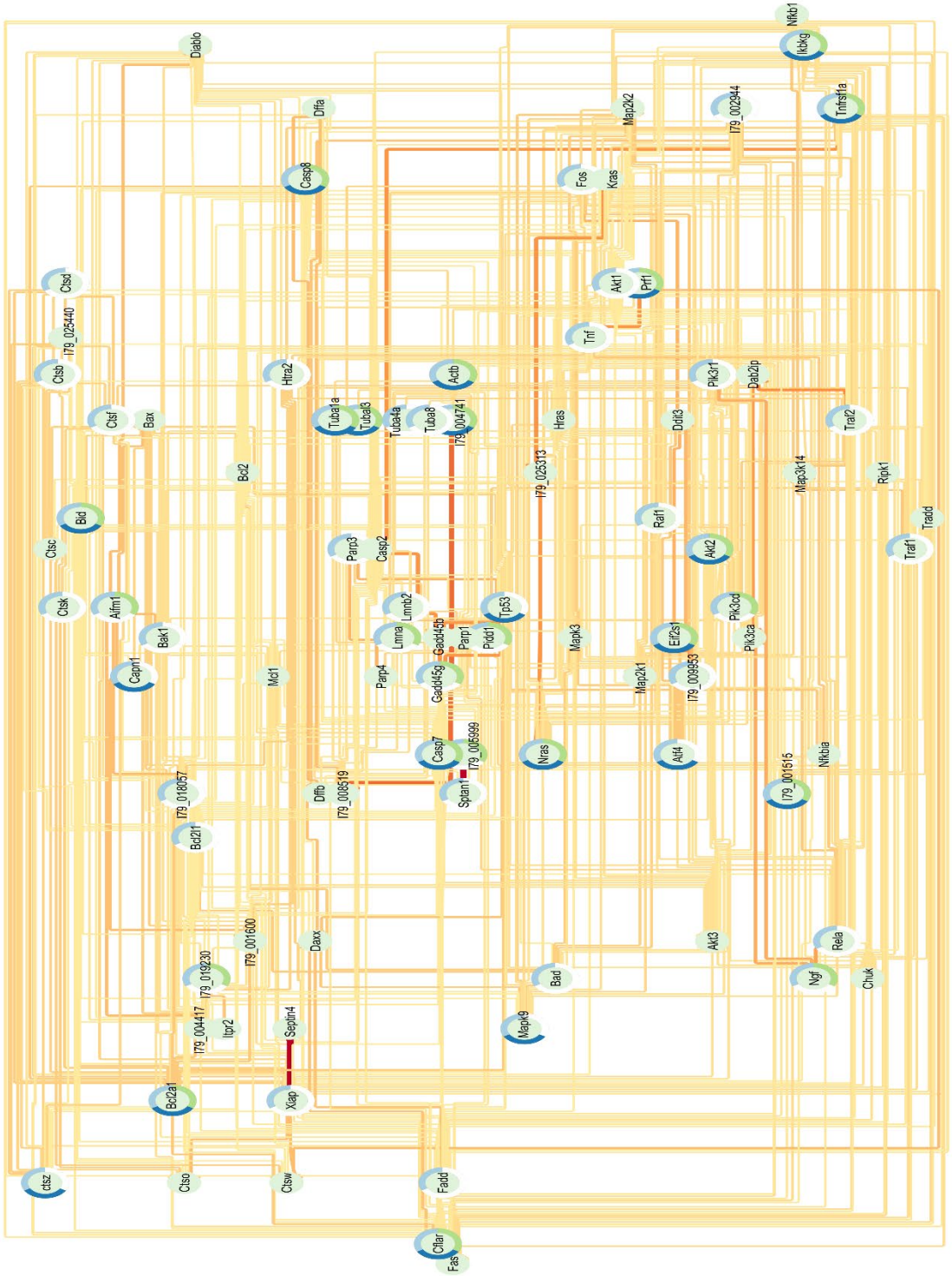
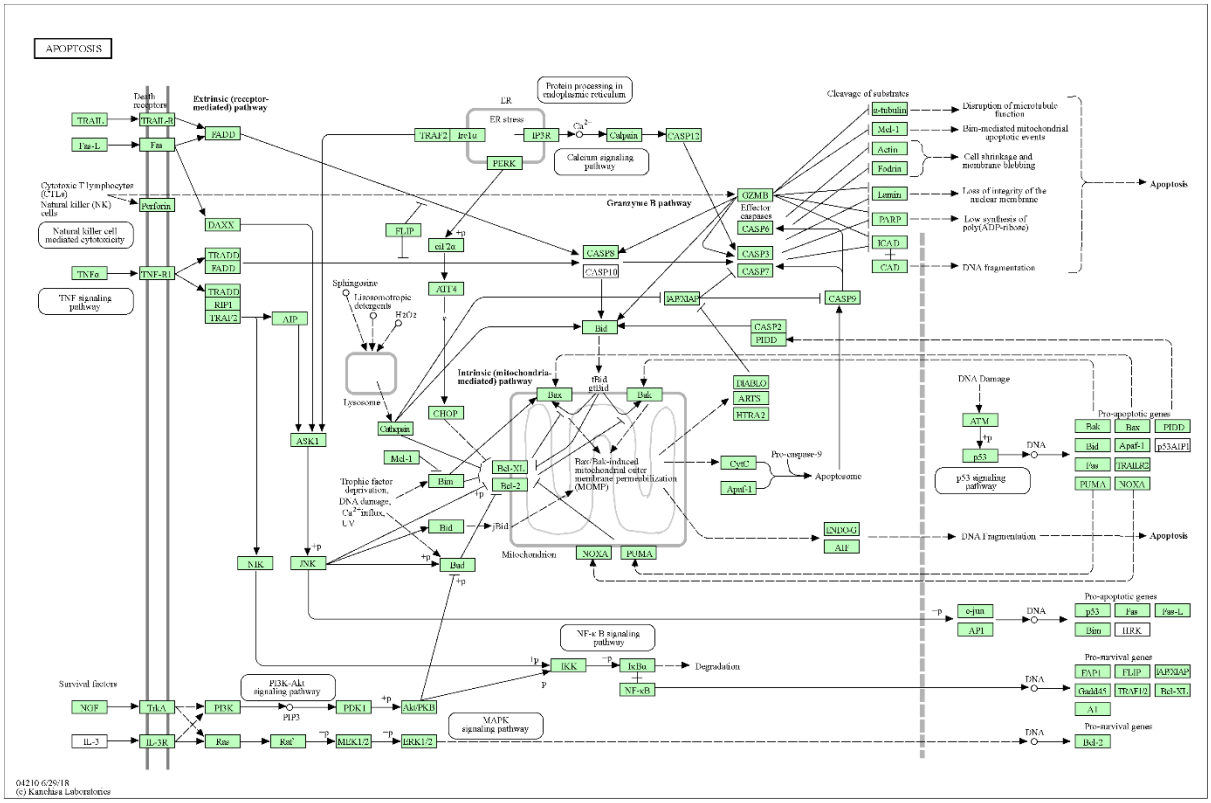

Diablo
Nfkb1
Ikbkg
Dffa
Map2k2
I79_002944
Tnfrsf1a
Casp8
Fos
Kras
Ctsd
I79_025440
Akt1
Prf1
Tnf
Ctsb
Htra2
Actb
Plk3r1
Dab2ip
Ctsf
Bax
Tuba1a
Tuba1l3
Tuba4a
Tuba8
I79_004741
Hras
Ddit3
Traf2
Bcl2
I79_025313
Map3k14
Ripk1
Bid
Raf1
Tradd
Ctsc
Parp3
Casp2
Akt2
Traf1
Ctsk
Aifm1
Lmnb2
Tp53
Plk3cd
Bak1
Lmna
Gadd45b
Parp1
Pidd1
Mapk3
Eif2s1
Plk3ca
Capn1
Mcl1
Parp4
Gadd45g
Map2k1
I79_009953
Casp7
I79_005999
Nras
Aif4
Nfkbia
I79_018057
Dffb
I79_008519
Sptan1
I79_001515
Bcl2l1
I79_001600
Daxx
Akt3
Rela
I79_019230
Bad
Ngf
Chuk
I79_004417
Itpr2
Septin4
Mapk9
Bcl2a1
Xiap
ctsz
Ctso
Ctsw
Fadd
Cflar
Fas

Figure 6. KEGG Apoptosis pathway (Map 04210)



2.Cell-cycle

. Summary statistics

Number of nodes was 99 and Number of edges was 1436. Also Avg.number of neighbors was 29.

Figure 7. Most important Nodes (Proteins) in Cell-cycle pathway

1	name	label	Degree	Radiality	Stress
2	Cdk1	Unknown	79	0.997416688	9008
3	Tp53	Unknown	50	0.993800052	6714
4	Cdk6	Unknown	67	0.995866701	5954
5	Ccna1	Unknown	66	0.99560837	5382
6	Ccnb1	Unknown	63	0.995350039	4716
7	Plk1	Unknown	48	0.993283389	4040
8	Skp1	Unknown	45	0.993154224	3874
9	Ccnb2	Unknown	60	0.994833376	3664
10	Cdkn1a	Unknown	47	0.993283389	3552
11	Cdc20	Unknown	51	0.993670886	3472
12	CDC6	Unknown	56	0.994316714	3436
13	Ccnd1	Unknown	42	0.992766727	3398
14	Ccne1	Unknown	48	0.993283389	3326
15	Rbl2	Unknown	38	0.991991733	3210
16	Rbl1	Unknown	38	0.991991733	3182
17	Smad3	Unknown	20	0.988116766	3082
18	Cdkn1b	Unknown	43	0.992766727	2788
19	Smad4	Unknown	18	0.98682511	2516
20	Myc	Anti	27	0.990570912	2512
21	Cdkn1c	Unknown	41	0.99237923	2364
22	Smad2	Unknown	17	0.987600103	2350
23	Ccne2	Unknown	45	0.992766727	2334
24	E2f1	Anti	42	0.992766727	2298

Figure 8. Most important Edges (Links) in Cell-cycle pathway

1	name	weight
2	Abl1 (-) Tp53	0.13138385
3	Ccnd1 (-) Abl1	0.12664908
4	Cdkn1b (-) Abl1	0.124226265
5	Abl1 (-) Cdkn1c	0.12370234
6	Tp53 (-) Gadd45b	0.12292514
7	Gadd45g (-) Tp53	0.122925125
8	Abl1 (-) Cul1	0.12259631
9	Cdkn1a (-) Abl1	0.12250597
10	Rb1 (-) Abl1	0.119755805
11	Tgfb1 (-) Tgfb3	0.11426465
12	Gadd45g (-) Gadd45b	0.112149365
13	Tgfb1 (-) Tp53	0.110379204
14	Pcna (-) Gadd45b	0.10818198
15	Pcna (-) Gadd45g	0.10818197
16	I79_003855 (-) I79_007265	0.09995147
17	I79_003855 (-) Ywhaz	0.09881722
18	Ywhaz (-) Ywhaq	0.098817
19	I79_003855 (-) Ywhab	0.09820014
20	I79_003855 (-) Ywhah	0.09730821
21	Rbx1 (-) Smad2	0.09705053
22	Smad3 (-) Rbx1	0.09690494
23	Cdkn2d (-) Tp53	0.09617581
24	I79_003855 (-) I79_026021	0.09345734

Figure 9. Hierarchic Layout of Cell-cycle pathway

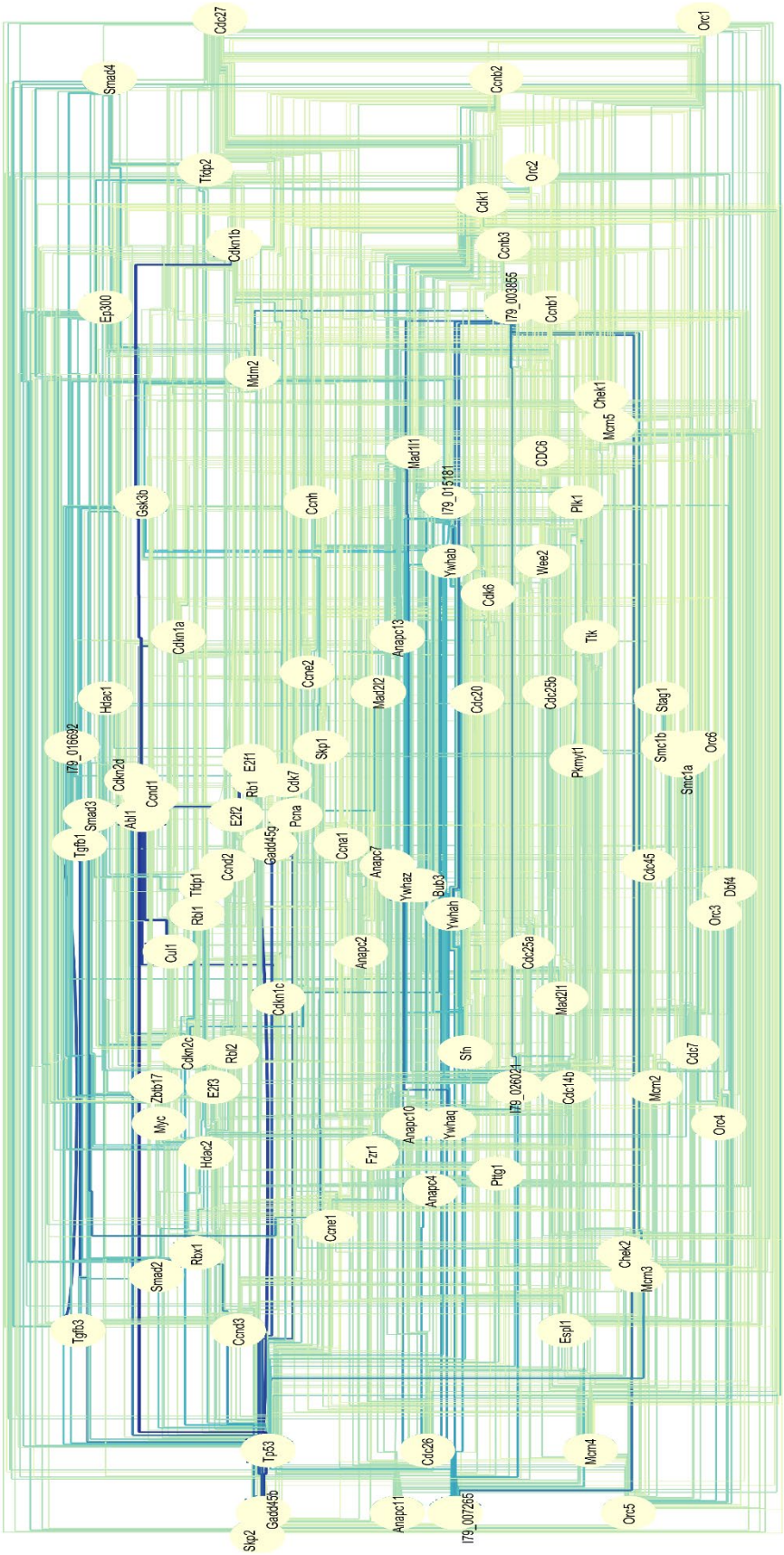
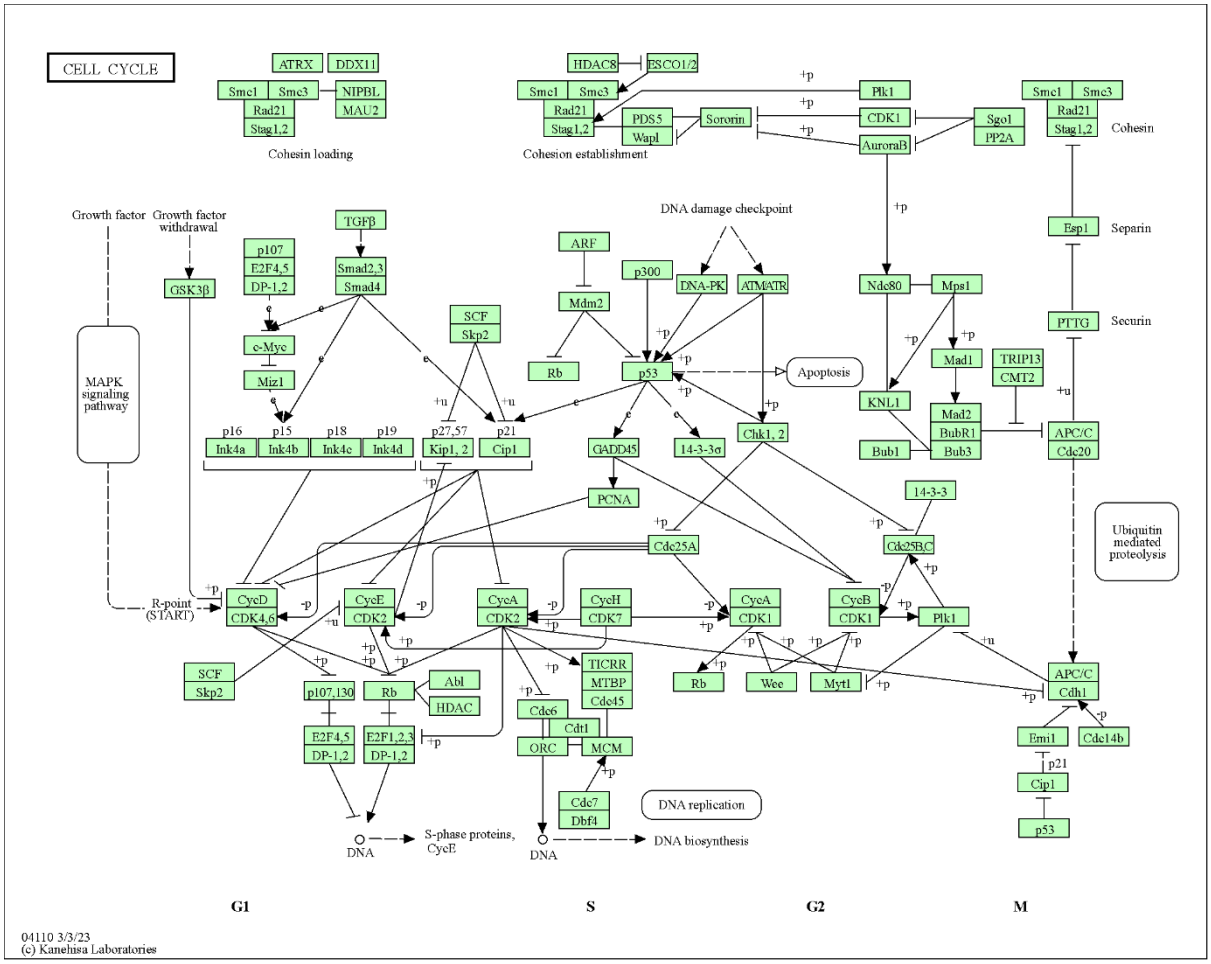


Figure 10. KEGG Cell-cycle pathway (Map 04110)



References:

[1] Jayapal KP et al. Recombinant Protein Therapeutics from CHO Cells — 20 Years and Counting. Chem Eng Prog. 2007;103(12):40-47. https://www.aiche.org/sites/default/files/docs/pages/CHO.pdf

[2] Pereira M. CHO in biomanufacturing: Past, present and future. Eur Pharm Manuf. 2021; https://pharmaceuticalmanufacturer.media/pharmaceutical-industry-insights/biopharma-news/cho-in-biomanufacturing-past-present-and-future/

[3] Wurm FM et al. HEK cells vs. CHO cells in recombinant antibody production: What's the better choice? Pharmiweb.com. 2022; https://www.pharmiweb.com/article/hek-cells-vs-cho-cells-in-recombinant-antibody-production-whats-the-better-choice

[4] Klauser B et al. Exploring the molecular content of CHO exosomes during bioproduction processes using RNA sequencing analysis. Appl Microbiol Biotechnol. 2021;105(17):7319-7330. https://link.springer.com/article/10.1007/s00253-021-11309-8

[5] Xu X et al. The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. Nat Biotechnol. 2011;29(8):735-741. https://www.nature.com/articles/nbt.1932

[6] Lewis NE et al. Omics analysis of Chinese hamster ovary cells. Curr Opin Biotechnol. 2013;24(4):604-610. https://www.sciencedirect.com/science/article/pii/S0958166913000550

[7] Ronda C et al. Crispr-Cas9-mediated genome engineering of CHO cells for production of diverse biopharmaceuticals. Nat Biotechnol. 2019;37(7):694-700. https://www.nature.com/articles/s41587-019-0104-6

[8] Karthik P. Jayapal et al., "Recombinant Protein Therapeutics from CHO Cells — 20 Years and Counting", AIChE Journal, 2007.

[9] Mario Pereira, "CHO in biomanufacturing: Past, present and future", European Pharmaceutical Manufacturer, 2021.

[10] Niamh O'Sullivan et al., "Enhanced recombinant protein production in CHO cell continuous culture using CRISPR-Cas9 genome editing", PLOS ONE, 2021.

[11] GlobalData, "Top 20 global biopharmaceutical companies report marginal growth in market cap during Q3 2023", 2023.

[12] Wurm FM. Production of recombinant protein therapeutics in cultivated mammalian cells. Nat Biotechnol. 2004;22(11):1393-8.
[13] Kim JY, Kim YG, Lee GM. CHO cells in biotechnology for production of recombinant proteins: current state and further potential. Appl Microbiol Biotechnol. 2012;93(3):917-30.
[14] Fussenegger M, Bailey JE. Molecular regulation of recombinant protein production in mammalian cells. Trends Biotechnol. 1998;16(10):431-7.
[15] Arnold FH. Design by directed evolution. Acc Chem Res. 1998;31(3):125-31.
[16] Walsh G. Biopharmaceutical benchmarks 2018. Nat Biotechnol. 2018;36(12):1136-45.

[17] Wurm FM. Production of recombinant protein therapeutics in cultivated mammalian cells. Nat Biotechnol. 2004 Nov;22(11):1393-8.

[18] Majors BS, Chiang GG, Betenbaugh MJ. Links between metabolism and apoptosis in mammalian cells: applications for anti-apoptosis engineering. Metab Eng. 2007 Jan;9(1):317-26.

[19] Elmore S. Apoptosis: a review of programmed cell death. Toxicol Pathol. 2007 Jun;35(4):495-516.

[20] Peter ME, Krammer PH. The CD95(APO-1/Fas) DISC and beyond. Cell Death Differ. 2003 Jan;10(1):26-35.

[21] Lavrik IN, Golks A, Krammer PH. Caspases: pharmacological manipulation of cell death. J Clin Invest. 2005 Oct;115(10):2665-72.

[22] Green DR, Kroemer G. The pathophysiology of mitochondrial cell death. Science. 2004 Jul 30;305(5684):626-9.

[23] Li P, Nijhawan D, Budihardjo I, Srinivasula SM, Ahmad M, Alnemri ES, Wang X. Cytochrome c and dATP-dependent formation of Apaf-1/caspase-9 complex initiates an apoptotic protease cascade. Cell. 1997 Nov 14;91(4):479-89.

[24] Pop C, Salvesen GS. Human caspases: activation, specificity, and regulation. J Biol Chem. 2009 Aug 14;284(33):21777-81.

[25] Kim NS, Lee GM. Down-regulation of Fas-associated death domain (FADD) expression by RNA interference inhibits radiation-induced apoptosis in CHO cells. Biotechnol Bioeng. 2005 May 20;90(4):404-11.

[26] Grilo, Antonio L., and Athanasios Mantalaris. "Apoptosis: A mammalian cell bioprocessing perspective." *Biotechnology advances* 37.3 (2019): 459-475.

[27] Evan GI, Vousden KH. Proliferation, cell cycle and apoptosis in cancer. Nature. 2001;411(6835):342-8.

[28] Wurm FM. Production of recombinant protein therapeutics in cultivated mammalian cells. Nat Biotechnol. 2004;22(11):1393-8.

[29] Kim JY, Kim YG, Lee GM. CHO cells in biotechnology for production of recombinant proteins: current state and further potential. Appl Microbiol Biotechnol. 2012;93(3):917-30.

[30] Doolan P, Barron N, Meleady P, O'Sullivan F, Clarke C, Dowling P, et al. Cell engineering strategies to enhance CHO cell robustness and productivity during bioprocessing. Curr Opin Biotechnol. 2019;60:110-6.

[31] Walsh G. Biopharmaceutical benchmarks 2018. Nat Biotechnol. 2018;36(12):1136-45.

[32] Elmore S. Apoptosis: a review of programmed cell death. Toxicol Pathol. 2007;35(4):495-516. doi:10.1080/01926230701320337

[33] Kim JY, Kim YG, Lee GM. CHO cells in biotechnology for production of recombinant proteins: current state and further potential. Appl Microbiol Biotechnol. 2012;93(3):917-930. doi:10.1007/s00253-011-3758-5

[34] Zhang L, Chen L, Sahin N, et al. Machine learning reveals complex interplay between bioprocess conditions and CHO cell phenotypes. Biotechnol Bioeng. 2019;116(9):2248-2260. doi:10.1002/bit.27027

[35] Lewis NE, Liu X, Li Y, et al. Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome. Nat Biotechnol. 2013;31(8):759-765. doi:10.1038/nbt.2624

[36] Graph-based prediction of Protein-protein interactions with attributed networks. BMC Bioinformatics, 2020.

[37] Advances in Mammalian Cell Line Development Technologies for Recombinant Protein Production. Pharmaceuticals, 2013.

[38] Yang F., Fan K., Song D., Lin H. (2020) Graph-based prediction of Protein-protein interactions with attributed signed graph embedding. BMC Bioinformatics 21:323. https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-020-03646-8

[39] Lv G., Hu Z., Bi Y., Zhang S. (2021) Learning Unknown from Correlations: Graph Neural Network for Inter-novel-protein Interaction Prediction. arXiv:2105.06709. https://arxiv.org/abs/2105.06709

[40] DataCamp (2021) A Comprehensive Introduction to Graph Neural Networks (GNNs). https://www.datacamp.com/tutorial/comprehensive-introduction-graph-neural-networks-gnns-tutorial

[41] Veličković P., Cucurull G., Casanova A., Romero A., Lio P., Bengio Y. (2018). Graph Attention Networks. In: International Conference on Learning Representations (ICLR). https://arxiv.org/abs/1710.10903

[42] Lee J., Yoon J., Kim E., Jeon S., Kang J., Kim S., Xing E.P., Lee S. (2020). Self-Attention Based Graph Neural Network for Drug-Target Interaction Prediction Using Gene Expression Data. In: Pacific Symposium on Biocomputing (PSB). https://psb.stanford.edu/psb-online/proceedings/psb20/lee-j.pdf

[43] Veličković et al. (2018). Graph Attention Networks. https://arxiv.org/abs/1710.10903

[44] DGL (2020). Graph Attention Networks. https://docs.dgl.ai/en/0.4.x/tutorials/models/1_gnn/9_gat.html

[45] Lee et al. (2019). Self-Attention Graph Pooling. https://arxiv.org/abs/1904.08082

[46] Hu et al. (2020). Heterogeneous Graph Transformer. https://arxiv.org/abs/2003.01332

[47] https://pandas.pydata.org/

[48] https://numpy.org/

[49] https://pytorch.org/

[50] https://pytorch-geometric.readthedocs.io/en/latest/

[51] https://scikit-learn.org/stable/

[52] https://networkx.org/

[53] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2018). Graph attention networks. arXiv preprint arXiv:1710.10903.

[54] https://colab.research.google.com/github/keras-team/keras-io/blob/master/examples/graph/ipynb/gat_node_classification.ipynb

[55] DOI: 10.1002/bit.28062

[56] DOI 10.1002/bit.22541

[57] DOI: 10.1002/bit.27269

[58] DOI: 10.1002/bit.27548

[59] DOI 10.1002/bit.24879

[60] https://doi.org/10.1016/j.biotechadv.2019.02.012

[61] DOI 10.1007/s10616-010-9274-0

[62] doi:10.1016/j.tibtech.2004.02.004

[63] DOI 10.1007/s10616-006-9008-5

[64] DOI: 10.1002/btpr.3228

[65] doi:10.1016/S1096-7176(03)00044-2

[66] DOI: 10.1002/bit.10913

[67] DOI: 10.1002/bit.20874

[68] DOI 10.1002/bit.26563

[69] DOI 10.1002/bit.21254

[70] Molecular Regulation of Cell-Cycle Progression and Apoptosis in

Mammalian Cells: Implications for Biotechnology , Biotechnol. Prog. 1998, 14, 807-833

[71] DOI 10.1007/s10529-011-0577-9

[72] doi:10.1016/j.jbiosc.2010.02.017

[73] DOI 10.1007/s11033-012-1882-7

[74] doi:10.1016/j.bbrc.2005.09.123

[75] Regulation of Cell Proliferation and Apoptosis in CHO-K1 Cells by

the Coexpression of c-Myc and Bcl-2, Biotechnol. Prog. 2005, 21, 671-677

[76] DOI: 10.1002/bit.20872

[77] A reference genome of the Chinese hamster based on a hybrid assembly strategy, doi: 10.1002/bit.26722.

[78] Proteomic Analysis of Chinese Hamster Ovary Cells, dx.doi.org/10.1021/pr300476w | J. Proteome Res. 2012, 11, 5265−5276.

[79] https://www.baeldung.com/cs/cross-validation-k-fold-loo

[80] https://www.statology.org/leave-one-out-cross-validation/

[81] https://dataaspirant.com/leave-one-out-cross-validation-loocv/

[82] https://biit.cs.ut.ee/gprofiler/page/apis

[83] Reimand J, Kull M, Peterson H, Hansen J, Vilo J. g:Profiler--a web-based toolset for functional profiling of gene lists from large-scale experiments. Nucleic Acids Res. 2007 Jul;35(Web Server issue):W193-200.

[84] https://www.biit.cs.ut.ee/gprofiler_archive2/r1760_e93_eg40/web/

[85] Reimand J, Arak T, Adler P, Kolberg L, Reisberg S, Peterson H, Vilo J. g:Profiler-a web server for functional interpretation of gene lists (2016 update). Nucleic Acids Res. 2016 Jul 8;44(W1):W83-9.