

AI-based fluorescent labeling for cell line development

Hanna Pankova

Master thesis

| | |
|---------------------|--|
| Date of issue: | 01. April 2022 |
| Date of submission: | 29. August 2022 |
| Reviewers: | Prof. Dr. Markus Kollmann Dr. Wolfgang Halter |

Erklärung

Hiermit versichere ich, dass ich diese Master thesis selbstständig verfasst habe. Ich habe dazu keine anderen als die angegebenen Quellen und Hilfsmittel verwendet.

Düsseldorf, den 29. August 2022

Hanna Pankova

Abstract

Cell line development is an expensive and time-consuming process, however that is the most modern approach for producing the proteins needed in pharmaceuticals.

Contents

| | | |
|----------|--------------------------------------|----------|
| 1 | Introduction | 1 |
| 1.1 | Motivation | 1 |
| 1.2 | Notation | 1 |
| 2 | Background | 1 |
| 2.1 | Biology | 1 |
| 2.2 | ML | 2 |
| 3 | Prediction of cell organelles | 3 |
| 4 | Summary | 4 |
| | References | 5 |
| | List of Figures | 6 |
| | List of Tables | 6 |

1 Introduction

1.1 Motivation

1.2 Notation

2 Background

2.1 Biology

CLD Cell line fluorescence DIC vector CHO cell proteins transfection

Explain CLD process Its need for producing proteins Explain how cells are chosen and that it long Explain the role of fluorescence Explaing its pro and cons

Chinese hamster ovary (CHO) cells are ones of the most popular mammalian cells used for protein production. (doi:10.1016/B978-0-08-100623-8.00007-4)

First step of cell line development (CLD) is the introduction of the gene of interest (GOI) to CHO cells, this process is called a transfection. It is important to transfect a GOI (or a vector) into an optimal site of genome to secure a high protein expression over time during protein production, but transfection mostly happens into a random place. In case if the gene was transfected in the inactive site of genome (and the majority of genome is not transcriptionally active) the cell likely will not express this gene. (doi:10.1016/B978-0-08-100623-8.00007-4) (doi:10.1016/j.coche.2018.08.002) However nowadays it is also possible to insert a GOI into a specific genomic location via CRISPR/Cas9 technology.

The second step is to select cell pools that have successful and stable gene integrations. The reasond why not all of them are suitable for cloning is the following: during the transfection only 80% of cells recieve the vector of GOI (doi:10.1016/B978-0-08-100623-8.00007-4), only the small percent of them actually integrate a vector into the genome (as mentioned above) and only a fraction of those cells are able to stably produce a protein. (Reference needed). This could be done with bulk sorting algorithm. (doi:10.1016/B978-0-08-100623-8.00007-4)

The third step in CLD is to clone the cells. The chosed stable pools of cells are phenotypically and genetically diverse - have different growth rates, metabolic profile and etc. This is not ideal for industrial production ([25] here doi:10.1016/B978-0-08-100623-8.00007-4). Fluorescent Activated Cell Sorting (FACS) technology is used to choose single best cells for further cloning based on several parameters like cell size, granularity, cell surface protein expression and etc. (<https://doi.org/10.1517/14712598.4.11.1821>). Unfortunately fluorescence labeling is expensive and may ruin the cell, so there is a need to label cell parts in silico without intervening into the cell with the chemicals that are used for the fluorescence labeling. There is also a limited number of available fluorescent channels in microscopes, phototoxicity. (<https://doi.org/10.1371/journal.pone.0007497>). Such labels can also be inconsistent, depend a lot on reagent quality, and requiere human lab work hours.

2.2 ML

Background on Unet and ML in general Conrad (1997) hat ein Buch geschrieben. Es gibt auch andere Arbeiten (Patterson and Hennessy, 2004) die referenziert sind. In Abbildung 1 ist ein Sachverhalt dargestellt.

1 Autor: Conrad (1997) (Conrad, 1997)

2 Autoren: Liebeck and Conrad (2015) (Liebeck and Conrad, 2015)

3 Autoren: Liebeck et al. (2016) (Liebeck et al., 2016)

Online resource: ILSVRC2016 (2017)



Figure 1: Gallien zur Zeit Caesars

| Jahr | Erster Consul | Zweiter Consul |
|-------|----------------------|-----------------------|
| 1 | C. Caesar | L. Aemilius Paullus |
| 2 | P. Vinicius | P. Alfenus Varus |
| 3 | L. Aelius Lamia | M. Servilius |
| 4 | Sex. Aelius Catus | C. Sentius Saturninus |
| 5 | L. Valerius Messalla | Cn. Cornelius Cinna |
| suff. | C. Vibius Postumus | C. Ateius Capito |
| 6 | M. Aemilius Lepidus | L. Arruntius |

Table 1: Römische Konsulen

3 Prediction of cell organelles

4 Summary

References

- Stefan Conrad (1997). *Föderierte Datenbanksysteme: Konzepte der Datenintegration*. Springer Verlag.
- ILSVRC2016 (2017). *ILSVRC 2016*. URL: <http://image-net.org> (visited on 07/17/2017).
- Matthias Liebeck and Stefan Conrad (July 2015). “IWNLP: Inverse Wiktionary for Natural Language Processing”. In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. Association for Computational Linguistics, pp. 414–418.
- Matthias Liebeck, Katharina Esau, and Stefan Conrad (Aug. 2016). “What to Do with an Airport? Mining Arguments in the German Online Participation Project Tempelhofer Feld”. In: *Proceedings of the Third Workshop on Argument Mining (ArgMining2016)*. Association for Computational Linguistics, pp. 144–153.
- David A. Patterson and John L. Hennessy (2004). *Computer Organization and Design*. Morgan Kaufmann Publishers.

List of Figures

| | | |
|---|------------------------------------|---|
| 1 | Gallien zur Zeit Caesars | 2 |
|---|------------------------------------|---|

List of Tables

| | | |
|---|-----------------------------|---|
| 1 | Römische Konsulen | 2 |
|---|-----------------------------|---|