

# Master Thesis Proposal: Pathology Foundation Model for Immunohistochemistry (IHC)

Schueffler Lab

## 1 Background and Methodology

Pathologists commonly use immunohistochemistry (IHC) for advanced tissue evaluation. This staining technique employs colorized antigen-specific antibodies to visualize the presence or absence of particular proteins. For example, HER2 staining visualizes the presence and spatial distribution of the human epidermal growth factor receptor 2 (HER2) in breast and gastric cancers (Fig 1), an important marker for breast cancer treatment decisions [LZX<sup>+</sup>22, HLW<sup>+</sup>22].

Hundreds of different IHC stains exist and manual IHC assessment is prone to subjectivity. Therefore, objective and automated analysis is favorable. To cover a wide range of different IHC stains, an IHC-based foundation model is needed. Existing pathology **foundation models (FMs)** [CDL<sup>+</sup>24, ZVV<sup>+</sup>24, WZM<sup>+</sup>24] are predominantly trained on slides stained with hematoxylin and eosin (HE) and therefore cannot easily be applied to IHC-based diagnostic tasks.

In this proposal, our objective is to develop a **foundation model trained directly on IHC slides** to enable more accurate automated analysis, and to facilitate downstream tasks such as staining estimation and cancer subtype classification, ultimately improving clinical decision making.

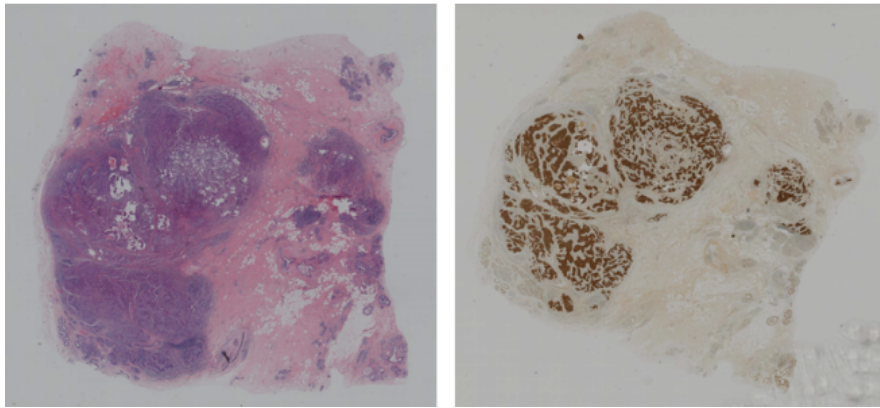


Figure 1: HE (left) and HER2 (right) staining on a breast cancer slide. While current models in pathology focus on HE slides as the most common staining technique, this project aims to build a FM for the IHC-stained images.

## 2 Dataset and Resources

The TUM Pathology produces over 30000 IHC slides per year that can potentially be used to build a foundation model. All slides since 2020 are digitized. The student will get access to the institutional compute cluster that is sufficient to train and evaluate the model. Ethical approval of the project is available.

### 3 Tasks

- Review of the literature on automatic IHC analysis.
- Data compilation and curation for IHC from the TUM pathology slide repository. Curation refers to filtering relevant stainings.
- Training of an IHC FM. The self-supervised training framework (DINOv2 [ODM+23]) is available.
- Implementation and evaluation of at least two downstream IHC analysis tasks (e.g., staining estimation pipelines for MIB-1 [SFO+13], ER [NME+20], PR [GJW+21] and/or HER2 [BLP24])
- Draft a manuscript for publication.

### 4 Requirements

Basic knowledge in at least two of the following areas:

- Deep Learning, Vision Transformers, Machine Learning
- Linux, Python, Pytorch, and distributed GPU training
- Medical images, especially pathology images
- Master seminar CuToMeMaLeCoPa(IN2107, IN4514) or I2DL(IN2346) or ML(IN2064)
- Your desire to pursue a PhD degree is a plus
- Handling sensitive (clinical) data requires

### 5 Supervision and Contact

The successful candidate will be supervised by Jingsong Liu and Prof. Peter Schüffler. If you are interested, please send us a brief description of your relevant experiences together with your grade transcript to [jingsong.liu@tum.de](mailto:jingsong.liu@tum.de).

### References

- [BLP24] Marie Brevet, Zaibo Li, and Anil Parwani. Computational pathology in the identification of her2-low breast cancer: Opportunities and challenges. *Journal of Pathology Informatics*, 15:100343, 2024.
- [CDL+24] Richard J Chen, Tong Ding, Ming Y Lu, Drew FK Williamson, Guillaume Jaume, Andrew H Song, Bowen Chen, Andrew Zhang, Daniel Shao, Muhammad Shaban, et al. Towards a general-purpose foundation model for computational pathology. *Nature Medicine*, 30(3):850–862, 2024.
- [GJW+21] Paul Gamble, Ronnchai Jaroensri, Hongwu Wang, Fraser Tan, Melissa Moran, Trissia Brown, Isabelle Flament-Auvigne, Emad A Rakha, Michael Toss, David J Dabbs, et al. Determining breast cancer biomarker status and associated morphological features using deep learning. *Communications medicine*, 1(1):14, 2021.
- [HLW+22] Zixin Han, Junlin Lan, Tao Wang, Ziwei Hu, Yuxiu Huang, Yanglin Deng, Hejun Zhang, Jianchao Wang, Musheng Chen, Haiyan Jiang, et al. A deep learning quantification algorithm for her2 scoring of gastric cancer. *Frontiers in Neuroscience*, 16:877229, 2022.
- [LZX+22] Shengjie Liu, Chuang Zhu, Feng Xu, Xinyu Jia, Zhongyue Shi, and Mulan Jin. Bci: Breast cancer immunohistochemical image generation through pyramid pix2pix. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1815–1824, 2022.

- [NME<sup>+</sup>20] Nikhil Naik, Ali Madani, Andre Esteva, Nitish Shirish Keskar, Michael F Press, Daniel Ruderman, David B Agus, and Richard Socher. Deep learning-enabled breast cancer hormonal receptor status determination from base-level h&e stains. *Nature communications*, 11(1):5727, 2020.
- [ODM<sup>+</sup>23] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023.
- [SFO<sup>+</sup>13] Peter J Schüffler, Thomas J Fuchs, Cheng Soon Ong, Peter J Wild, Niels J Rupp, and Joachim M Buhmann. Tmarker: A free software toolkit for histopathological cell counting and staining estimation. *Journal of pathology informatics*, 4(2):2, 2013.
- [WZM<sup>+</sup>24] Xiyue Wang, Junhan Zhao, Eliana Marostica, Wei Yuan, Jietian Jin, Jiayu Zhang, Ruijiang Li, Hongping Tang, Kanran Wang, Yu Li, et al. A pathology foundation model for cancer diagnosis and prognosis prediction. *Nature*, 634(8035):970–978, 2024.
- [ZVV<sup>+</sup>24] Eric Zimmermann, Eugene Vorontsov, Julian Viret, Adam Casson, Michal Zelechowski, George Shaikovski, Neil Tenenholtz, James Hall, David Klimstra, Razik Yousfi, et al. Virchow2: Scaling self-supervised mixed magnification models in pathology. *arXiv preprint arXiv:2408.00738*, 2024.