



**Owkin publishes its novel predictive tool for gene expression (HE2RNA)
in 'Nature Communications'**
***A deep learning model to predict RNA-seq expression of tumors
from whole slide images***

New York - USA, August 3, 2020 – [Owkin](#), the healthcare technology company applying [federated learning](#) to medical research, today announces the publication of its HE2RNA [paper in 'Nature Communications'](#) showcasing its novel tool for genomic analysis. The paper, entitled 'Transcriptomic Learning for Digital Pathology' describes how Owkin has developed a detailed and accurate deep learning model to predict RNA-seq expression of tumors from histology images of digitized biopsies.

Gilles Wainrib, Chief Scientific Officer and Co-Founder of Owkin said:

“Understanding the relationship between genotype and phenotype is one of the biggest 21st century challenges in biology. Our research opens a new path to better connect information at the genomic, cellular and tissue levels, and this would not have been possible without recent advances in artificial intelligence.”

Massive changes in gene expression are known to occur in many cancers. Understanding and characterizing these disease-related gene signatures can help to clarify disease mechanisms and prioritize targets for novel personalized therapeutic approaches.

Traditionally, the only available option for identifying gene expression during carcinogenesis has been to use whole transcriptome sequencing techniques (RNA-Seq) and dedicated bioinformatic tools. However, these analyses are costly and time consuming. As a result, medical centers do not routinely use them.

In oncology, tumor biopsies [or histology whole slide images ('WSIs')] are routinely collected in hospitals and research centers as a first step in the diagnostic and treatment pathway. The ready availability of these digitized slides in all research centers makes them a perfect data source for Machine Learning ('ML') models.

In recent years, deep ML has had a tremendous impact on various fields in science such as improvements in speech recognition and image recognition. Recently ML models have been applied to histology WSIs to improve the performance of pathologists in determining the diagnosis and grade of cancer patients. While it is becoming clear that the application of

such models to tissue-based pathology can be very useful, few attempts have been made to connect specific molecular signatures directly to gene expression patterns within histology slides.

A ML model that can use these ubiquitously available histology slides to determine gene expression without the need for expensive sequencing techniques has the potential to be an incredibly useful clinical tool.

Owkin HE2RNA model is named after its capability to predict gene expression (RNA) of numerous tumor genes in 28 different cancers from Hematoxylin-Eosin (HE) stained biopsy slides. The model was also able to highlight (via gene expression) the exact location of each mutation on each WSI, hence creating a Virtual Spatial Transcriptomics map. This interpretability feature, combined with the Model's ability to detect such a broad scope of mutations, offers huge potential to aid patient diagnosis and improve prediction of response to treatment and survival outcome. In this paper, Owkin describes how the model works. The paper also successfully explores the application of HE2RNA to predict genes involved in cancer development and to predict tumor status and response to therapies.

Elodie Pronier, Translational Research Scientist at Owkin and co-author of the paper commented on the success of the model, *"Our efforts were rewarded by the model's ability not only to correctly predict the location of a variety of gene expression signatures in each image but also to transfer the knowledge it learned on bulk data to smaller independent datasets to accurately answer specific clinically relevant molecular questions such as the identification of tumors with microsatellite instability. We are now excited to explore how HE2RNA learning of gene expression can help improve the prediction of other clinical targets on new datasets within our partner research centers to expand our scope and improve the performance of this model."*

Owkin specializes in AI for medical research. Through the application of its technology, the company enables researchers to build ML models on fit-for-AI cohorts, highly curated, multimodal, research-grade longitudinal data, while keeping patient information preserved safely within the hospital's local infrastructure. Ultimately, this method can result in an acceleration of the clinical research process that offers protected data for patients, exhaustive traceability of computations for institutions, and maximum collaboration for researchers.

Owkin's proprietary platform, Owkin Studio, integrates these biomedical images, genomics, and clinical data to discover biomarkers and mechanisms associated with disease evolution and treatment outcomes that will propel the next generation of treatment plans and drugs. Owkin's novel HE2RNA model is available to researchers to apply it to their datasets via [Owkin Studio](#) and the results published in this paper can be visualized and explored in [this demo](#).

.....

About Owkin:

The French-American startup, which was co-founded in 2016 by Dr. Thomas Clozel, a clinical research doctor and former assistant professor in clinical hematology and Gilles Wainrib, Ph.D., a pioneer in the field of artificial intelligence in biology, has raised \$70 million in venture capital.

Owkin connects several of the largest medical research centers and pharmaceutical companies in Europe and the U.S. within a federated research ecosystem. Owkin has developed four key components to build this ecosystem: Owkin Loop (the network), Owkin Connect (the technology infrastructure), Owkin Studio (the AI software tool) and Owkin Lab (the expertise).

Owkin Connect is a privacy-preserving, traceable, secure technology that allows the company to connect with research centers in the Owkin Loop network. Using Owkin Connect's federated learning approach, the data do not move, only algorithms travel. This enables insights from the data to be collectively shared while guaranteeing privacy for patients and compliance with data ownership.

In October 2019, Owkin published [in Nature Medicine](#) its breakthrough analysis of tumor biology using an interpretable deep-learning model, called MesoNet. In February 2020, [Hepatology](#) published Owkin's novel deep learning models to predict survival after hepatocellular carcinoma resection from histology slides. Most recently, in May 2020, following a winning entry to the data challenge organized last October by the [Société Française de Radiologie et d'imagerie médicale \(SFR\)](#), Owkin published its methodology to automatically measure muscular area from CT scans to assess sarcopenia in [Diagnostic and Interventional Imaging](#). In August 2020, Owkin published its novel genomic analysis tool (HE2RNA) in [Nature Communications](#).

For more information, please visit www.owkin.com, follow @OWKINscience on Twitter, contact Anna Huyghues-Despointes: anna.hd@owkin.com