Yale and WuXi NextCODE Use AI to Predict Phenotypes *In Vivo*

Study experimentally validates pioneering artificial intelligence strategy for the life sciences, accurately projecting complex disease processes using modest sets of gene expression data

- *The genetic knockout of ERK1/2 signaling results in the breakdown of proper function of endothelial cells and a range of major vascular pathologies*
- *An innovative AI approach elucidates the drivers and causal structure of the network that leads from genotype to disease*
- *Built on RNAseq from only three-dozen samples, the model was validated* in vitro *and in mice, accurately predicting disease progression mediated by the network*

13 June 2019 –Scientists at [Yale School of Medicine](#) working with [WuXi NextCODE](#)'s Advanced Artificial Intelligence Research Laboratory today published a landmark paper providing experimental validation for the application of artificial intelligence approaches to real-life challenges in medical and life science.

The study, which appears today in the online edition of the [Journal of Experimental Medicine,](#) uses a highly novel approach to shed new light on the known but little understood role of the ERK1/2 pathway in maintaining the healthy state of endothelial cells. These cells provide the critical inner lining of blood and lymph vessels. Yale professor Michael Simons and his team demonstrated that knocking out the ERK 1 and 2 genes in mice led to an increase of TGFß signaling and the conversion of endothelial cells to mesenchymal-like cells (a process known as endothelial-to-mesenchymal transition, or EndMT), ultimately leading to renal failure, hypertension and sudden death due to myocardial fibrosis.

This was classic medical biology and an important novel observation in itself. But the investigators wanted to understand how the disruption of ERK1/2 expression caused the activation of TGFß and led to the end phenotypes they observed. That was a challenge that required mapping out complex and largely unknown pathways through large sets of data. It was a question too big and too open-ended for hypothesis driven biology, but one potentially well suited to an artificial intelligence approach.

To answer it, WuXi NextCODE Chief AI Scientist Tom Chittenden and his colleagues generated RNAseq data from three dozen human umbilical endothelial cell samples, half with ERK1/2 silenced and half untreated. They compared the expression patterns in nearly 14,000 genes between the two groups, using a deep artificial neural network to identify the 1000 genes that were the most differentially expressed. They then used probabilistic programming to test the causal dependencies of millions of possible interactions between these proteins, and identified a core network driven by TGFß2 resulting in EndMT. The causal gene network accurately predicted all observed phenotypes, including a specific cause of the observed systemic hypertension phenotype and renal dysfunction thereby validating the ensemble AI strategy.

"It is a real milestone to be able to draw out and validate a causal biological network using such an efficient and replicable AI approach," said Dr Simons, Professor of Medicine and Cell Biology at Yale and senior author on the paper. "We have become quite good at making observations that correlate a genotype and a phenotype, but tracing the biology that lies between has always been elusive because it is so complex. The promise of AI is that it is powerful enough to bring together all the biology and genetic data we now have to unravel this complexity, and Tom's group is leading the way in showing how this can be done."

"Today we are providing a first concrete look at what we call phenotype projection: an efficient AI-driven approach that can predict complex phenotypes by teasing out the causal molecular underpinnings of disease," said Dr Chittenden, PhD, DPhil, co-senior author. "By furthering our collective understanding of biology, such approaches hold the potential to be truly transformative. It means that we can understand virtually any disease

in much greater detail using cost-effective experimental designs, a fundamental capability for creating precision medicine. The result is a range of validated potential points for developing therapeutic interventions; validated markers for designing smaller clinical trials with a greater chance of success; and a wealth of information for identifying patients likely to respond to approved compounds."

All press inquiries can go to info@wuxinextcode.com.