A088 · Kidzoi Enables Cell-Type-Specific Regulatory Variant Effect Prediction in the Kidney

Presenter: Arif Ahmad Rather — Boston Children's Hospital

Authors: Arif Ahmad Rather, Dongwon Lee

Session: Day 1 · Thu Oct 1, 2026 · 4:15–5:15 PM

Large-scale deep learning models trained on extensive genomic datasets across different modalities can accurately predict genomic signals from DNA sequences, providing valuable insights into gene regulation and the functional impact of genetic variants. However, it is infeasible to encompass all cell types or experimental conditions when training such models. Consequently, studying new biological contexts often requires training additional models, which is computationally intensive. To address this gap, we developed Kidzoi by adapting Borzoi, a pretrained multi-modal genomic foundation model, to single-nucleus chromatin accessibility data from human kidney tissues across 10 distinct cell types. Our results demonstrate that Kidzoi effectively learns cell-type-specific accessibility changes, achieving a Pearson correlation of 0.37-0.67 compared 0.22-0.47 for ChromKid (specialized model for kidney) when predicting chromatin accessibility signals from DNA-sequences. Furthermore, Kidzoi outperforms existing state-of-the-art general purpose and specialized models in predicting regulatory variant effects. For instance, measured by AUROC, Kidzoi achieves a 10% relative improvement over AlphaGenome, 18.5% over Enformer, 32.3% over Borzoi, 23% over Sei, and a 14.2% improvement over ChromKid in predicting cell-type-specific regulatory variant effects in proximal tubule cells. Kidzoi also successfully prioritized and localized GWAS fine-mapped variants for a major kidney phenotype, estimated glomerular filtration rate (eGFR). We found that 29.68% of such variants are predicted to have a significant impact in at least one kidney cell-type. Additionally, these variants, on average, are predicted to impact only 3 out of the 10 distinct cell types, suggesting that they exert their functions through highly specific cellular contexts. Through systematic sequence context ablation, we find that an input window of approximately 32-64 kbp is sufficient for optimal regulatory variant effect prediction, capturing the essential local regulatory context within just 6-12% of the full model input. We further benchmarked single-task against multi-task learning, demonstrating that the multi-task paradigm achieves comparable predictive performance at a substantially reduced computational cost. Together, our findings show that transfer learning from foundation genomic models yields accurate, cell-type-resolved predictors of regulatory variant effects, with Kidzoi surpassing general and specialized baselines and mapping eGFR GWAS variants to a small set of kidney cell types.

Kidzoi Enables Cell-Type-Specific Regulatory Variant Effect Prediction in the Kidney

Arif A. Rather,[1,2,3*] Dongwon Lee,[1,2,3]

[1]Department of Pediatrics, Division of Nephrology, Boston Children's Hospital, Boston, MA, USA; [2]Harvard Medical School, Boston, MA, USA; [3]Broad Institute, Cambridge, MA, USA

Large-scale deep learning models trained on extensive genomic datasets across different modalities can accurately predict genomic signals from DNA sequences, providing valuable insights into gene regulation and the functional impact of genetic variants. However, it is infeasible to encompass all cell types or experimental conditions when training such models. Consequently, studying new biological contexts often requires training additional models, which is computationally intensive. To address this gap, we developed Kidzoi by adapting Borzoi, a pretrained multi-modal genomic foundation model, to single-nucleus chromatin accessibility data from human kidney tissues across 10 distinct cell types.

Our results demonstrate that Kidzoi effectively learns cell-type-specific accessibility changes, achieving a Pearson correlation of 0.37-0.67 compared 0.22-0.47 for ChromKid (specialized model for kidney) when predicting chromatin accessibility signals from DNA-sequences. Furthermore, Kidzoi outperforms existing state-of-the-art general purpose and specialized models in predicting regulatory variant effects. For instance, measured by AUROC, Kidzoi achieves a 10% relative improvement over AlphaGenome, 18.5% over Enformer, 32.3% over Borzoi, 23% over Sei, and a 14.2% improvement over ChromKid in predicting cell-type-specific regulatory variant effects in proximal tubule cells.

Kidzoi also successfully prioritized and localized GWAS fine-mapped variants for a major kidney phenotype, estimated glomerular filtration rate (eGFR). We found that 29.68% of such variants are predicted to have a significant impact in at least one kidney cell-type. Additionally, these variants, on average, are predicted to impact only 3 out of the 10 distinct cell types, suggesting that they exert their functions through highly specific cellular contexts. Through systematic sequence context ablation, we find that an input window of approximately 32-64 kbp is sufficient for optimal regulatory variant effect prediction, capturing the essential local regulatory context within just 6-12% of the full model input. We further benchmarked single-task against multi-task learning, demonstrating that the multi-task paradigm achieves comparable predictive performance at a substantially reduced computational cost. Together, our findings show that transfer learning from foundation genomic models yields accurate, cell-type-resolved predictors of regulatory variant effects, with Kidzoi surpassing general and specialized baselines and mapping eGFR GWAS variants to a small set of kidney cell types.