Projects with videos

#1: **CBIO056 - Novel GenAI to Improve Alzheimer's Diagnosis**

Alzheimer’s disease (AD) is a progressive neurodegenerative disorder characterized by cognitive decline severe enough to interfere with daily activities. Current diagnostic methods, such as neuropsychological assessments, achieve an accuracy below 80%, indicating significant room for improvement. Magnetic resonance imaging (MRI), which reveals characteristic brain shrinkage in AD patients, presents a promising opportunity for machine learning to enhance diagnostic accuracy. This study leverages a ResNet50-based convolutional neural network (CNN) for MRI-based AD diagnosis, initially achieving an F1 score of 89%. More importantly, this research introduces a novel generative adversarial network (GAN)-based framework for more accurate AD diagnosis. Following data pre-processing, AG-GAN selected the largest imaging site as a reference domain, conforming images from three other sites to similar image parameters and effectively reducing scanner-specific variability. Subsequently, the CycleGAN generated additional synthetic images to address data scarcity and enhance training robustness. This systematic GAN-based integration significantly improved diagnostic performance, yielding an F1 score of 96.1%. The proposed methodology addresses critical challenges in multi-site MRI datasets, promoting equitable diagnosis by reducing performance disparities caused by scanner variability, limited data availability, and imbalanced datasets. As the first research project to propose a GAN-based framework to enhance AD diagnosis, these findings underscore the transformative potential of integrating deep learning with GAN-based harmonization and augmentation, improving AD diagnostic reliability and accessibility, particularly benefiting underfunded medical clinics and research facilities.

#2CBIO058 - Modeling Bird Flu Mutations and Effects With AI

Avian influenza (Bird Flu) is a highly transmissive pathogen affecting over 166 million poultry and posing significant health risks to humans. Despite progress on surface proteins, differences between human and avian flu limit vaccine crossover. Internal proteins like the conserved NS1, which binds host proteins and increases viral replication, are promising targets for broader, cross-species flu intervention. Additionally, the 3D structure of viral-NS1 in complex with host-STAU2, an important feature of viral replication, remains unresolved. This study employs a two-part approach: (1) comparative analysis of 63 NS1 proteins to identify bird- and human-specific mutations and (2) development of a Contrastive Learning with Influence-Tree Graph Attention Network (CLIGAT) model to interpret how these mutations influence viral-host interactions. This study finds and maps 9 amino acid differences between bird and human NS1 variants and identifies R193Q as the

strongest mutation (enhancing affinity), while S165F and I129T weakened viral affinity. CLIGAT achieved 98.9–99.2% accuracy in predicting these effects. In conclusion, three therapeutic targets are proposed, including residues 189–201 (linked to increased affinity), residues 162–168 (mutations that reduce affinity), and a potential allosteric site on Host-STAU2 (residues 446–470) that may destabilize viral-host interactions. Using contrastive learning and influence trees, this study improves the interpretability of 3D graph models by identifying direct (contact) and indirect interactions within the NS1-STAU2 complex. By constructing a novel 3D NS1-STAU2-RNA complex and understanding the effects of current mutations, these findings offer three potential intervention strategies to contain bird flu transmission.

#3**CBIO006 - Biclonal Antibodies to Prevent Ferroptosis in AD**

Alzheimer's Disease (AD), a neurodegenerative disorder linked to pathologies characterized with neural cognitive decline, has recently been connected to iron dysregulation and ferroptosis in its pathology. Amyloid-beta has also been shown to aggregate from different forms of iron inclusion due to disturbance of iron homeostasis, thus proving the significance of ferroptosis in AD pathology. This study presents a novel therapeutic approach targeting the hepcidin-ferroportin interaction to mitigate iron accumulation in AD. This study utilizes a framework to develop accurate monoclonal antibodies (mAb) that target and inhibit ferroportin-hepcidin interactions, preventing ubiquitination of the complex and restoring iron efflux.\nThis computational framework initially employs a multi-iterative modeling process based on MODELLER frameworks; DOPE scores and RMSF values are relayed through an algorithm that selects residues and correlative PDB files for the generation of the most accurate protein-structure of a target sequence. Molecular Dynamics was implemented to optimize such models, improving dihedral angle populations with Ramchandran calculations coupled with organization of residual stability. Low-resolution docking of candidate antibodies using the ProABC-2 led to the identification of a high-affinity mAb targeting the ferroportin-binding site, providing an initial model for antibody development.\nPreliminary results demonstrated effective disruption of hepcidin-ferroportin interactions and restoration of ferroportin functions showcased by HADDOCK docking formulas. Such studies are effective in proving that the amyloid hypothesis is not a singular option for AD pharmaceutical development. Future work involves implementing QMD and in-vitro antibody development.\n

#4
CZE001
CBIO001 - Retroelement Regulatory Hypothesis in ASD

Autism Spectrum Disorder (ASD) is a heterogeneous neurological disorder affecting approximately 1 in 100 children today. The diagnosis and treatment of ASD are complex, partly due to its obscure genetic basis. Transposable elements (TEs) are DNA sequences capable of changing their position in the genome. They comprise approximately 50% of the human genome. Although they have largely lost their ability to transpose, they acquired regulatory functions, such as host gene expression control, due to their enrichment with CpG sites and transcription factor binding sites. Notably, it has already been established that certain TE families are enriched near ASD-implicated genes, suggesting potential genomic and regulatory instability. This study utilized public RNA-seq data from post-mortem brain tissue of ASD and non-ASD control donors to investigate the impact of TE epigenetic modifications. We sought to identify common differentially expressed genes (DEGs) between the ASD and non-ASD groups and characterize their functional roles in ASD pathology through gene ontology and pathway enrichment analyses. Additionally, we aimed to identify TE families near these DEGs to explore their potential regulatory role in ASD gene expression. Our analyses revealed a consistent immune-related pattern among DEGs and identified specific TE families (e.g., L2a/b/c, MER4B, and MLT1J) enriched in proximity to DEGs. This suggests that TEs may contribute to immune dysregulation in ASD and potentially influence the autism phenotype. Our research proposes a functional hypothesis for ASD manifestation, identifying TEs as a novel therapeutic target to correct immune dysregulation, pending further experimental validation.\n

#5**CBIO011 - Deep Learning to Diagnose Leukemia Subtypes**

B-cell acute lymphoblastic leukemia (B-ALL) and acute myeloid leukemia (AML) are the 2 most common leukemias. The American Cancer Society estimates that there will be 59,610 leukemia cases and 23,710 deaths in 2023. Early detection of leukemia subtypes is critical in prescribing effective treatment strategies to maximize survival rates. Past studies using machine learning (ML) for diagnosing leukemia have achieved significant results, but very few studies were extended to cover the detection of specific leukemia subtypes. B-ALL and AML subtypes are difficult to distinguish because they have similar morphology. The objective of this study is to develop a novel dual-input model to distinguish subtypes of B-ALL and AML. Transfer learning will be utilized, and different classification algorithms: Fully-connected (FC), FC + Batch Normalization (BN), Global average pooling (GAP) will be tested on top of 3 pre-trained CNNs (VGG16, DenseNet201, InceptionResNetv2). Image segmentation will identify malignant cells, and data augmentation will apply transformations to increase the number of training samples. The model's hyperparameters will be finetuned while checkpoints and early stopping callbacks track loss and mitigate overfitting. Of all CNNs tested, DenseNet201

with the GAP classifier achieved optimal test accuracy of 96.62% with an AUC score of 0.9989. An app called LST Detector will be developed for public use and to create a community-supported repository of blood smear images for the model to undergo further training. This model is the first to successfully identify all morphologically similar subtypes of B-ALL and AML leukemias.

#6
USCA07
CBIO041 - Blood-Based AD Detection With Biotype-Guided GNN

Alzheimer's disease (AD) affects 1 in 9 individuals aged 65+ in the U.S., with females twice as likely as males. As AD pathology begins decades before symptoms appear, early detection and intervention are critical. I developed NeuroPlasmaNet, an AI system that identifies early-stage blood-based biomarkers using graph neural networks (GNNs) to simulate the sex-specific brain-blood multi-omics gene networks guided by neural pathology. Since AD is defined in the brain and to be detected in the blood, I first constructed biotype-stratified brain gene graphs based on cell types and layers, controlling for confounding biases such as age, APOE genotype, and education. I deployed reinforcement fine-tuning with cosine-based similarity reward function to optimize each learning iteration and align predicted genes with AD biological relevancy. This approach revealed distinct male vs. female molecular signatures. I then refined non-invasive blood-based gene selection and identified NeuroPlasma12 gene panel (e.g., C1QB, TXNIP, TREM2, GFAP, PLCG2, CD163, CAMK1D, LRP10). This brain-first analysis grounded blood markers in AD neural pathology, avoiding systemic noise. I introduced the NeuroPlasma Score (NPS) to quantify the gene panel profile, achieving 92.03% accuracy (AUC=0.9410, $p < 0.001$) for early AD detection. I also identified potential therapeutic targets (e.g., GRIN2B, PLCG2, GRM5, CAMK1D), some novel and others published in Nature and JAMA. Beyond gene markers, the results highlight modifiable cognitive risk factors, such as substance use, infections, circadian disruption, and nutrition deficiencies, and promote personalized, preventive brain health strategies to the largely overlooked preclinical population.

GeoMimic-Net: Project Summary GeoMimic-Net is a deep learning system for detecting molecular mimicry—where viral proteins structurally imitate human proteins to hijack cellular machinery. The pipeline combines Equivariant Graph Neural Networks (EGNNs) for learning 3D geometric features with ESM-2 protein language model embeddings for sequence-level understanding, fused via cross-attention. Users provide two PDB files (a viral protein and a human protein), and the model outputs a similarity score (0-1) indicating whether the viral protein mimics the human target. The model was trained using contrastive learning (InfoNCE loss) on known mimicry pairs, then fine-tuned with

supervised triplet loss to reduce false positives from structurally similar but functionally unrelated proteins. Validation on held-out pairs (Zika envelope/AXL receptor, Chikungunya E1/MxRA8) achieved similarity scores >0.84, demonstrating generalization to unseen pathogens. The ablation study confirms that geometry is the primary driver of detection (F1=0.31 alone) while sequence features enhance precision when combined (F1=0.57). Additional analyses include mutation mapping to verify that predicted binding sites don't overlap with known immune-escape variants.