# General Artificial Intelligence (1)

## SAIR-2-07 :Multi-modal Multi-task Masked Autoencoders and Generative AI as A General Framework for Genetic Studies of Complex Diseases

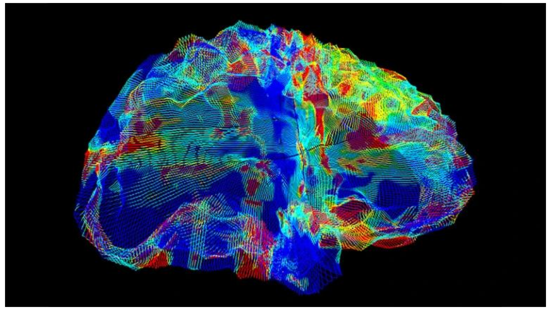
Momiao Xiong

Society of Artificial Intelligence Research

## AI that reads brain scans shows promise for finding Alzheimer's genes

**Machine-learning approach detects Alzheimer's disease with an accuracy of more than 90% — a potential boon for clinicians and scientists developing treatments.**

Max Kozlov, November 10, 2023. Nature



## Alzheimer's drug trials plagued by lack of racial diversity

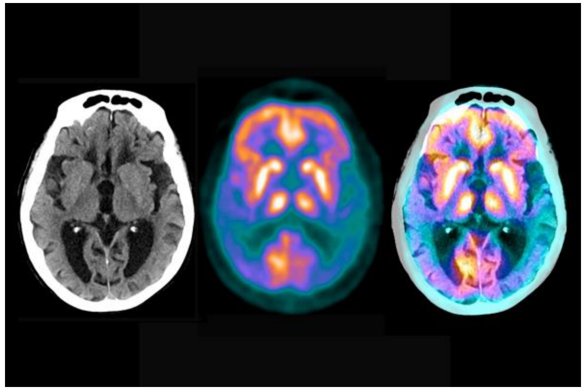
**Under-representation of people of colour sparks concerns over the safety and efficacy of drugs in diverse populations.**

Sara Reardon, August 2, 2023. Nature

## Conquering Alzheimer's: a look at the therapies of the future

**Researchers are looking to drug combinations, vaccines and gene therapy as they forge the next generation of treatments for the condition.**

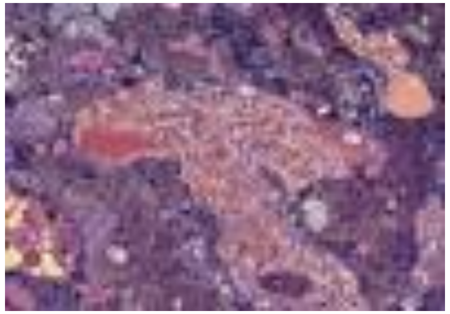
Alison Abbott, April 4, 2023. Nature.



## How one man’s rare Alzheimer’s mutation delayed the onset of disease

**Genetic resilience found in a person predisposed to early-onset dementia could potentially lead to new treatments.**

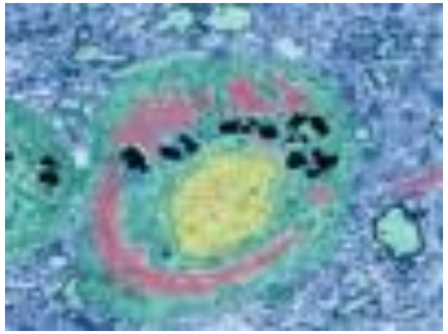
Sara Reardon, May 15, 2023. Nature



## This is how an Alzheimer’s gene ravages the brain

**Study in cells and mice suggests that the variant *APOE4* affects the all-important insulation around nerve cells.**

Elie Dolgin, November 16, 2022. Nature



## The brain cells linked to protection against dementia

**People with an abundance of specific neurons are more likely to escape cognitive decline despite having signs of Alzheimer’s in their brains.**

Sara Reardon, September 28, 2022. Nature

**Researchers have sifted through genomes from thousands of individuals in an effort to identify genes linked to Alzheimer's disease. But these scientists have faced a serious obstacle: it's hard to know for certain which of those people have Alzheimer's. There's no foolproof blood test for the disease, and dementia, a key symptom of Alzheimer's, is also caused by other disorders. Early-stage Alzheimer's might cause no symptoms at all.**

Max Kozlov, November 10, 2023. Nature

**Predictive algorithms have changed the world, and all the worlds to come, and there is no going back," they say in their articulate book on how current and future artificial-intelligence algorithms will change apprehension of risk and affect human behaviour. They accept that economics, business, finance and some aspects of medicine are not governed by scientific logic.**

The Age of Prediction
Igor Tulchinsky & Christopher E. Mason MIT Press (2023)

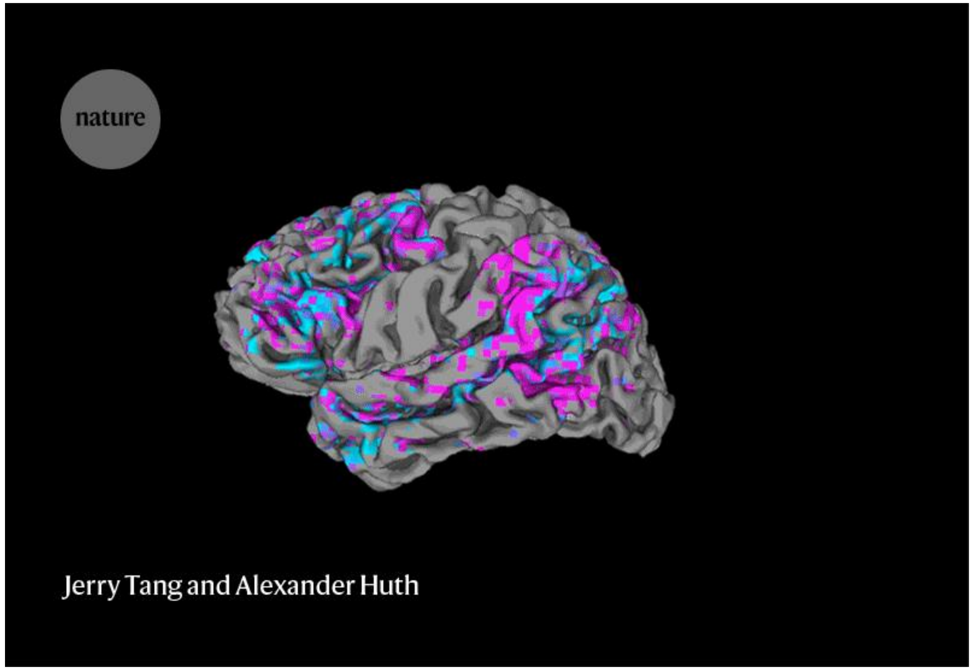
**Why genes mean less than you think, and other reads: Books in brief**

Andrew Robinson, November 17, 2023, Nature

# AI shows promise for finding Alzheimer's genes

Jack Shields

- One AI algorithm sorts through large numbers of brain images and identifies scans that include characteristics of Alzheimer's.
- Another machine-learning method identifies important structural features of the brain that could help scientists spot new signs of Alzheimer's in brain scans.
- The AI classifier reportedly detected Alzheimer's in brain scans with over 90% accuracy.
- Efforts to develop these algorithms are led by Paul Thompson, a neuroscientist at the University of Southern California in Los Angeles.
- The bottom line: Medical AI is still advancing rapidly. Advanced algorithms will soon unlock new treatment options for a plethora of diseases and health conditions.



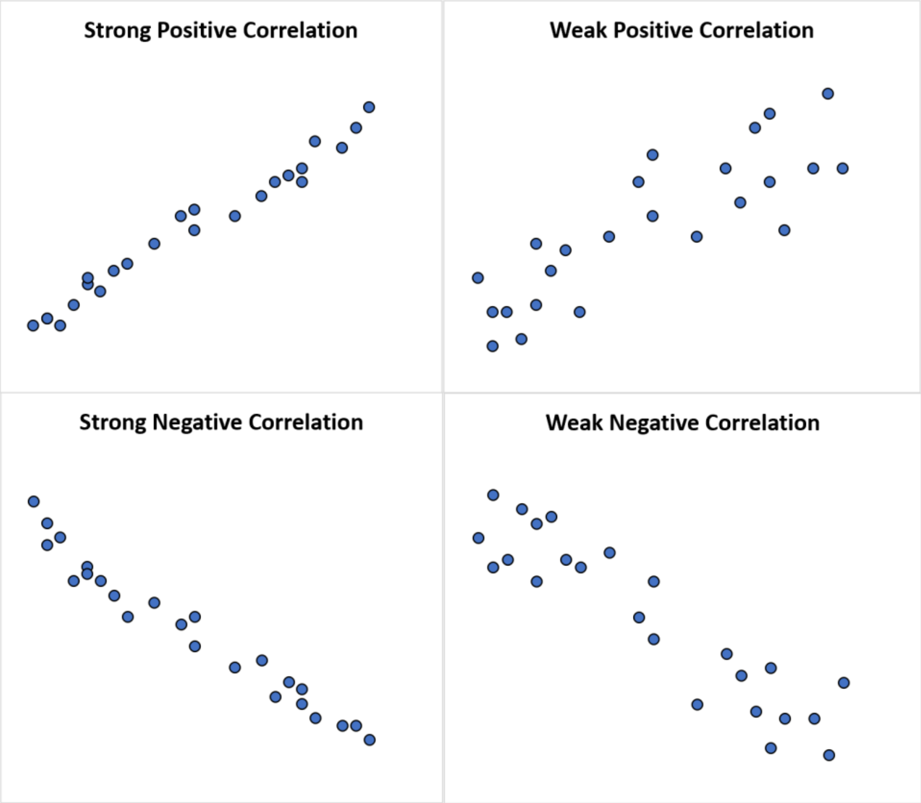
Jerry Tang and Alexander Huth

# Classical Approach to Genetic Studies of Complex Diseases

- **Association**

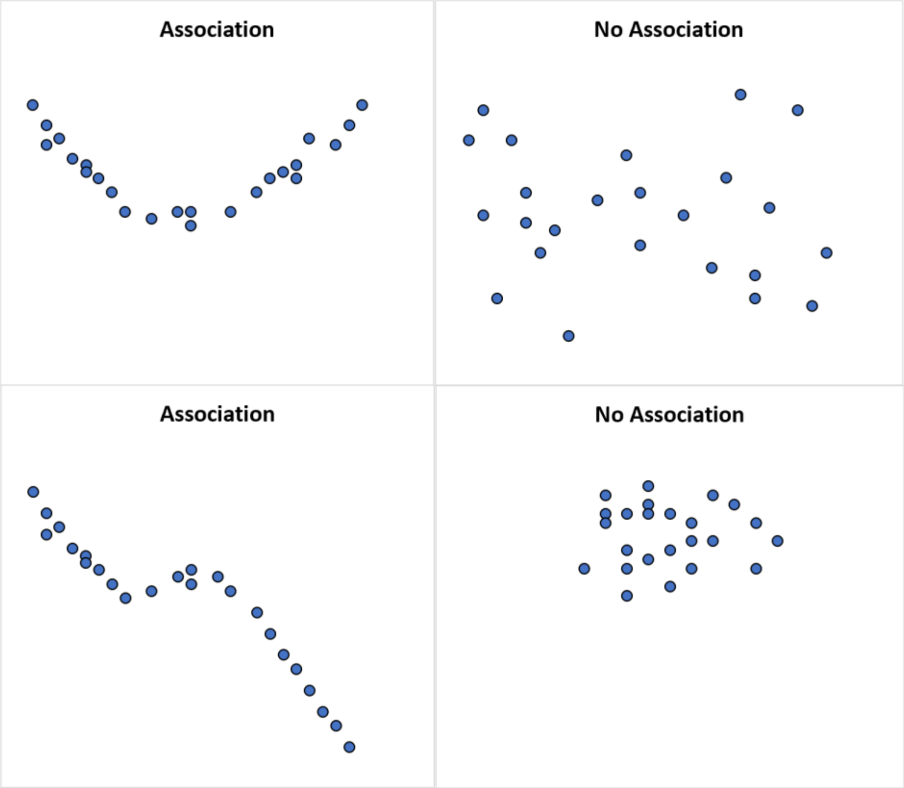
  **Pearson Correlation Coefficient**

  This is a measure of the linear association between two random variables

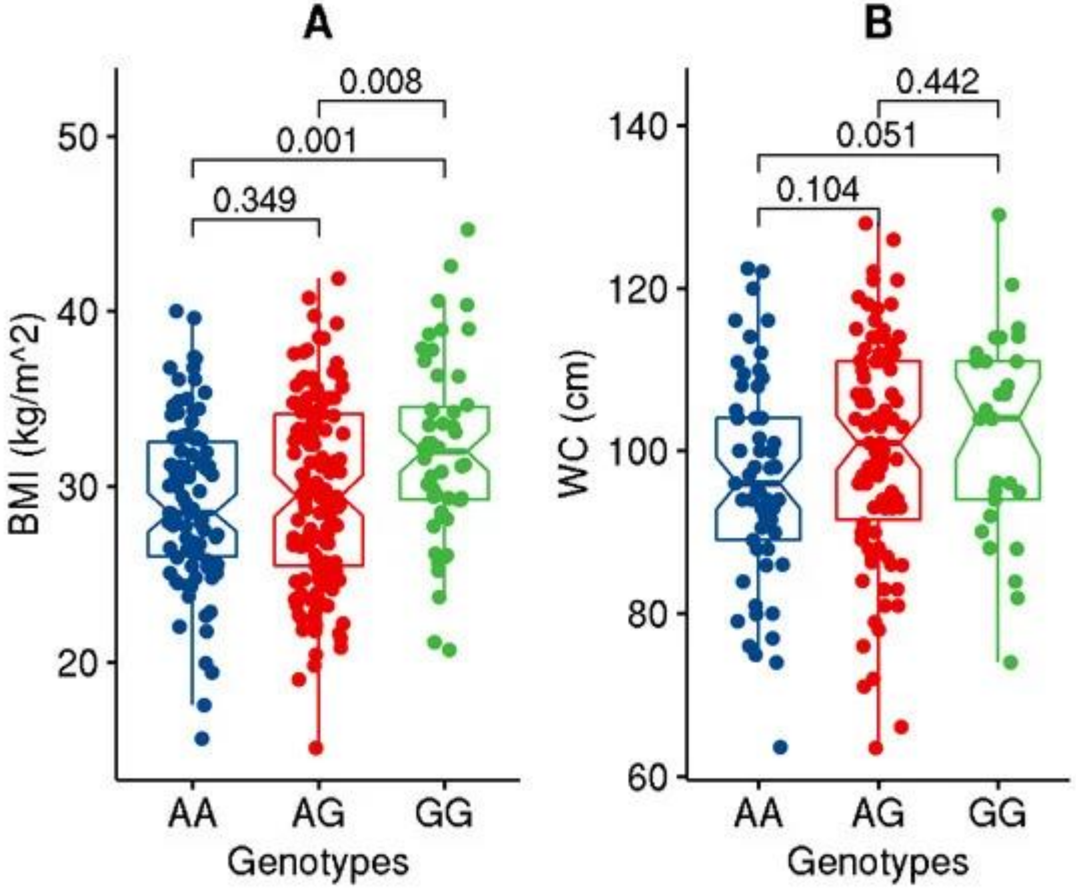


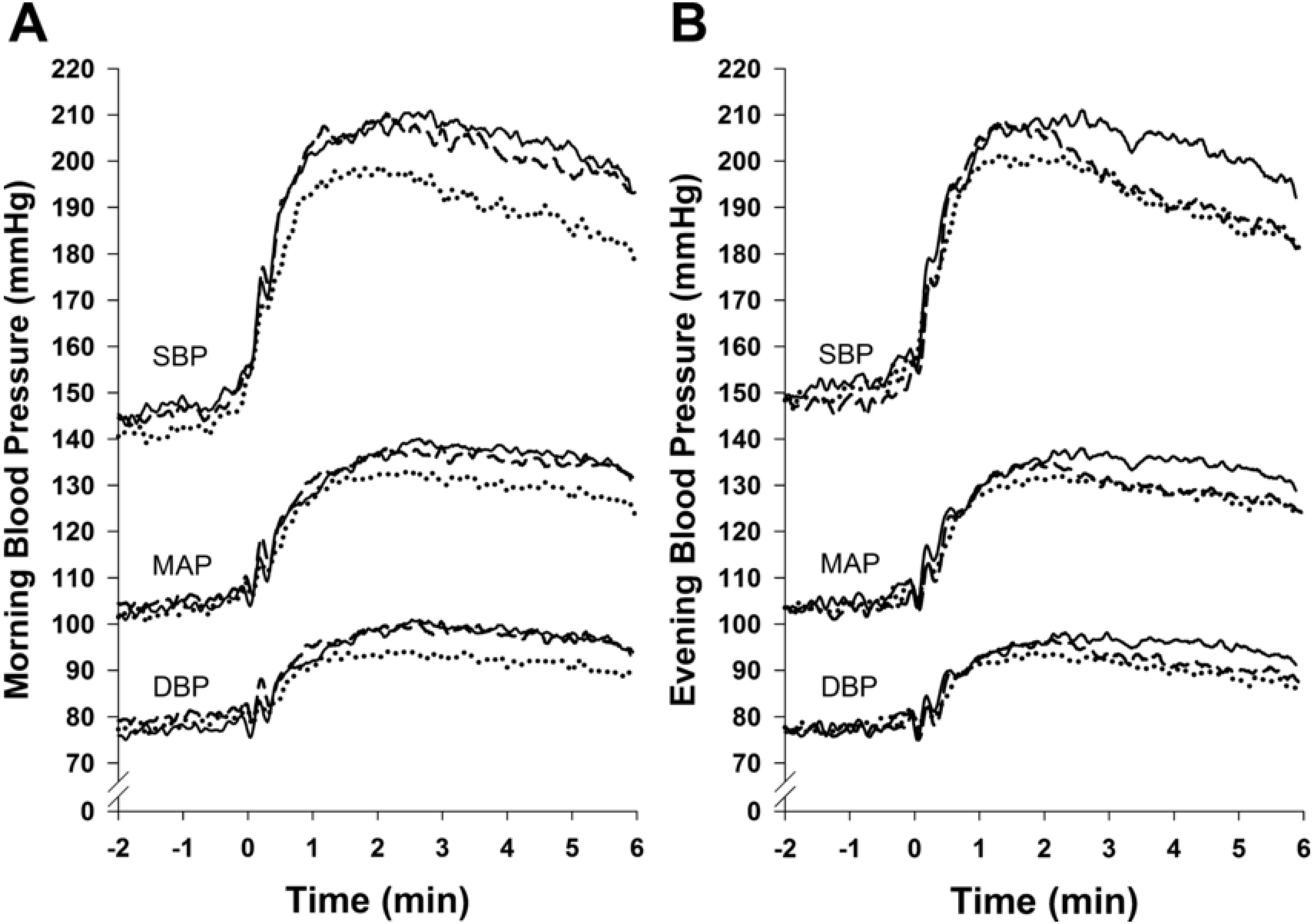
## Association

the word association can tell us whether or not there is any relationship between two random variables: linear or non-linear.

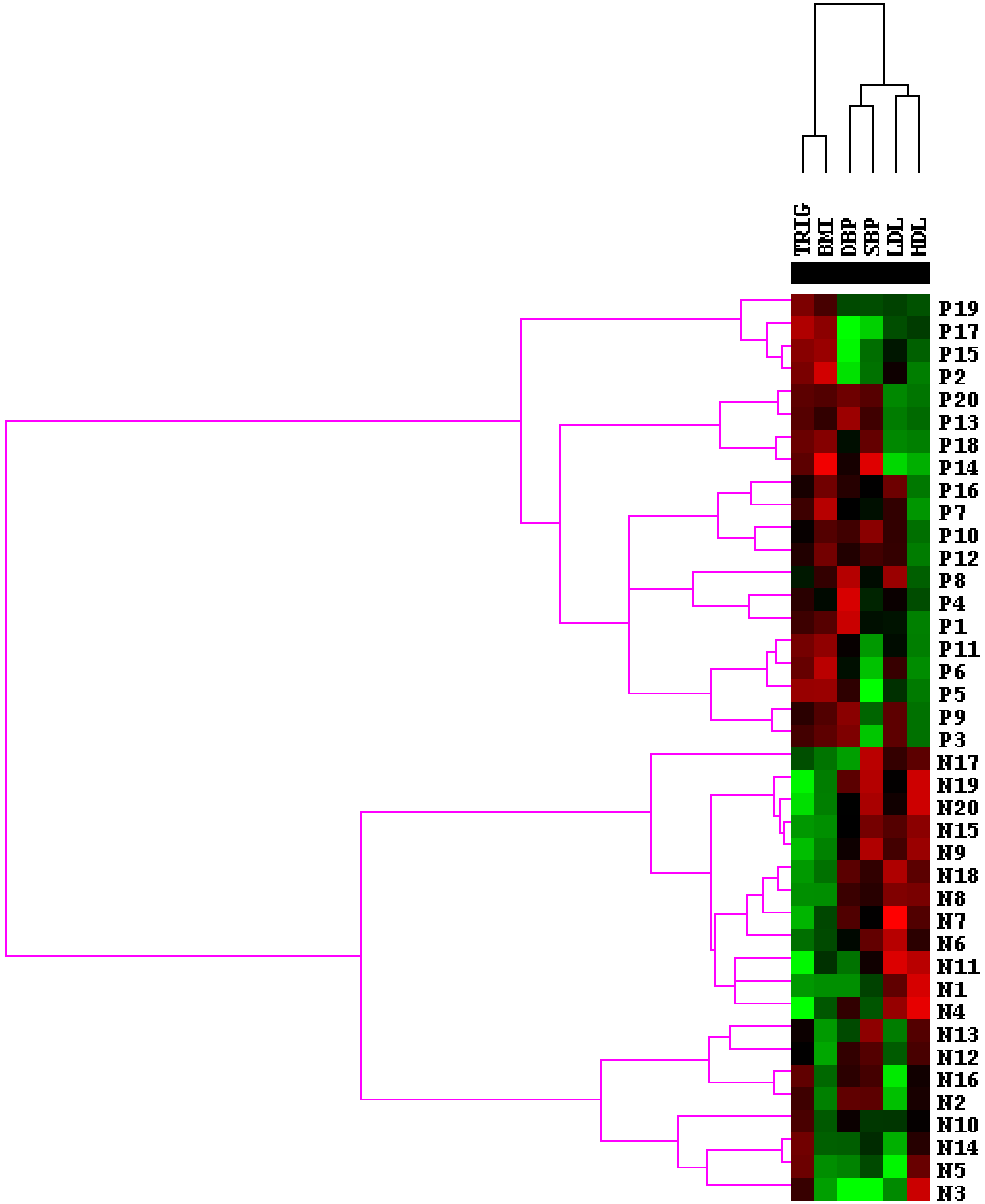


# Mapping Quantitative Trait Loci

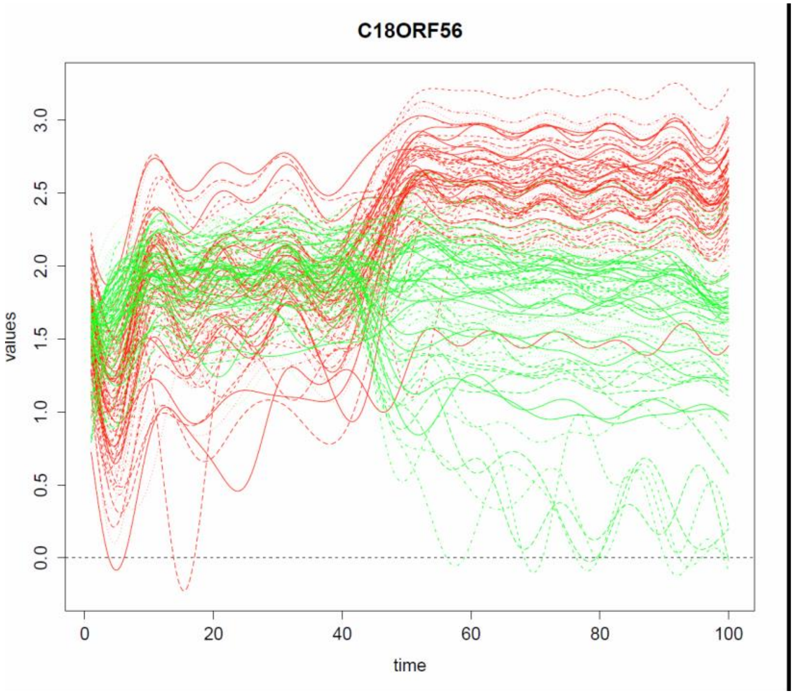


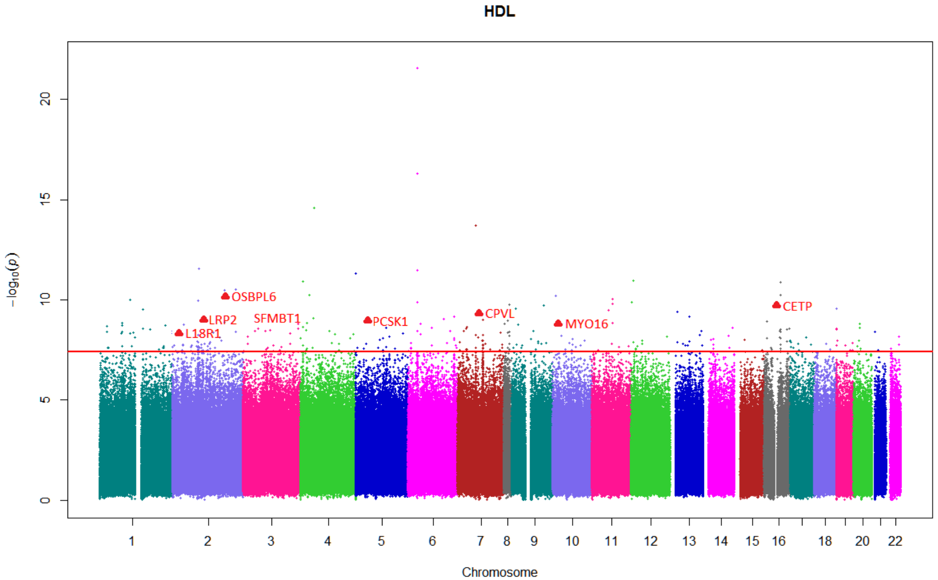

A
B
Morning Blood Pressure (mmHg)
Evening Blood Pressure (mmHg)
220
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
0
SBP
MAP
DBP
-2
-1
0
1
2
3
4
5
6
Time (min)

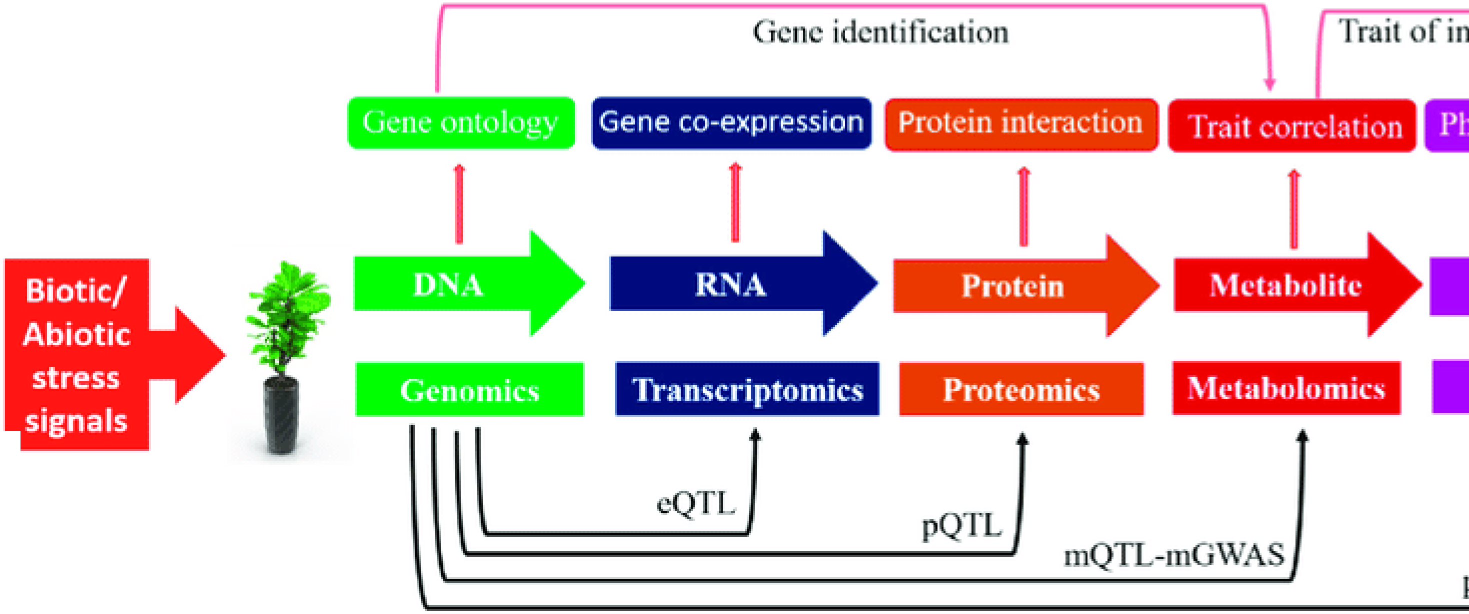
A quantitative trait may be a risk factor of diseases.



# RNA-Seq






Gene identification

Trait of in

Gene ontology

Gene co-expression

Protein interaction

Trait correlation

Ph

Biotic/
Abiotic
stress
signals

DNA

RNA

Protein

Metabolite

Genomics

Transcriptomics

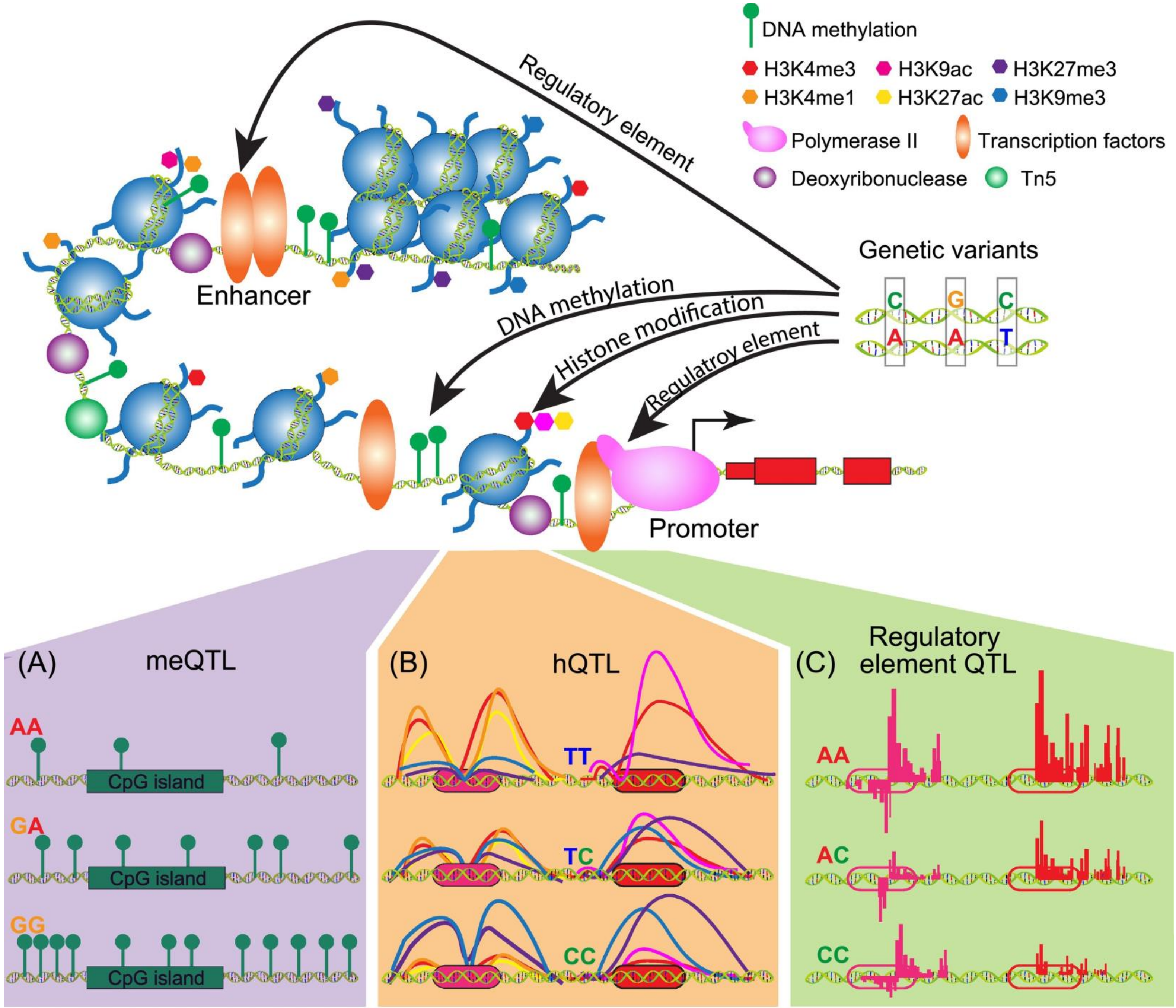
Proteomics

Metabolomics

eQTL

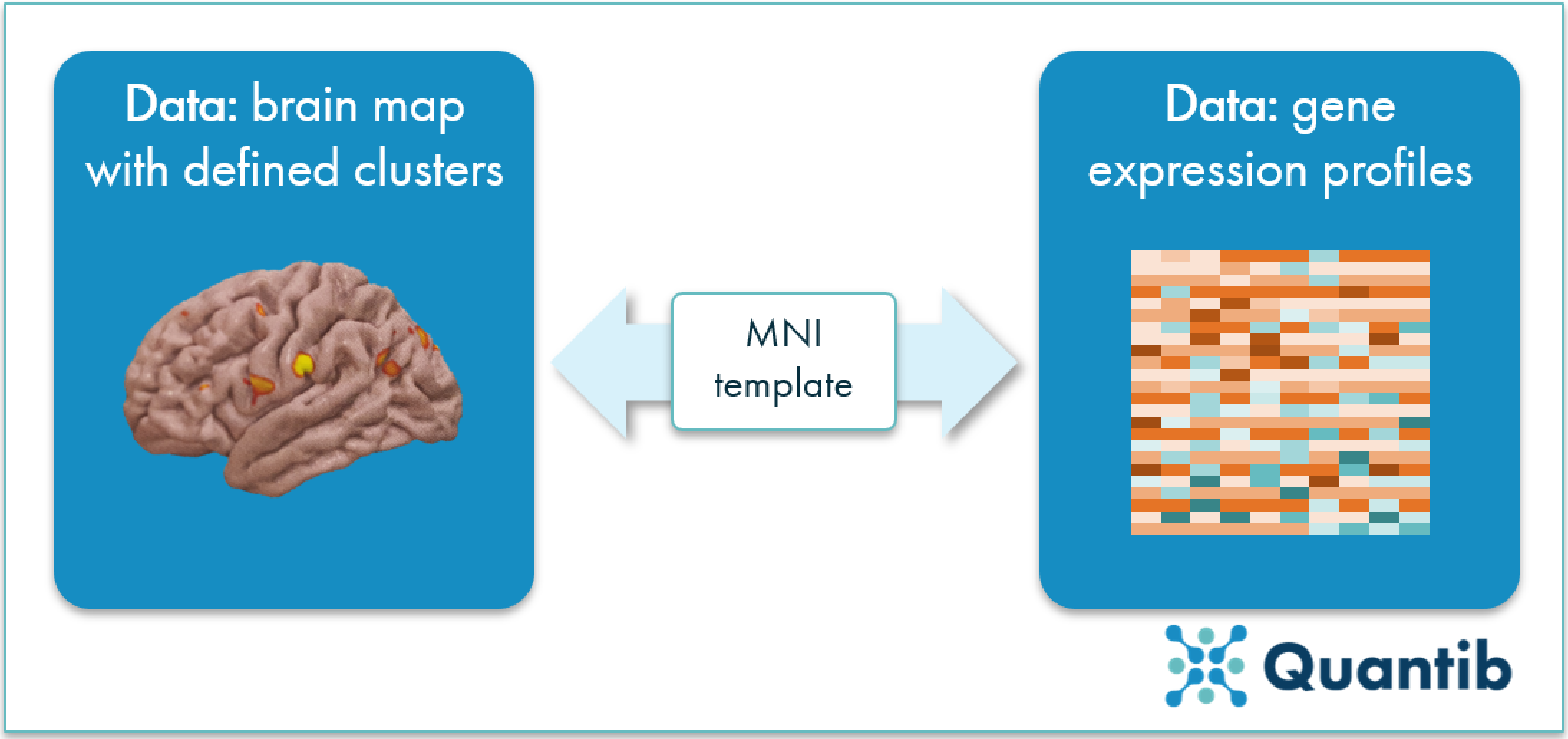
pQTL

mQTL-mGWAS



Trends in Genetics

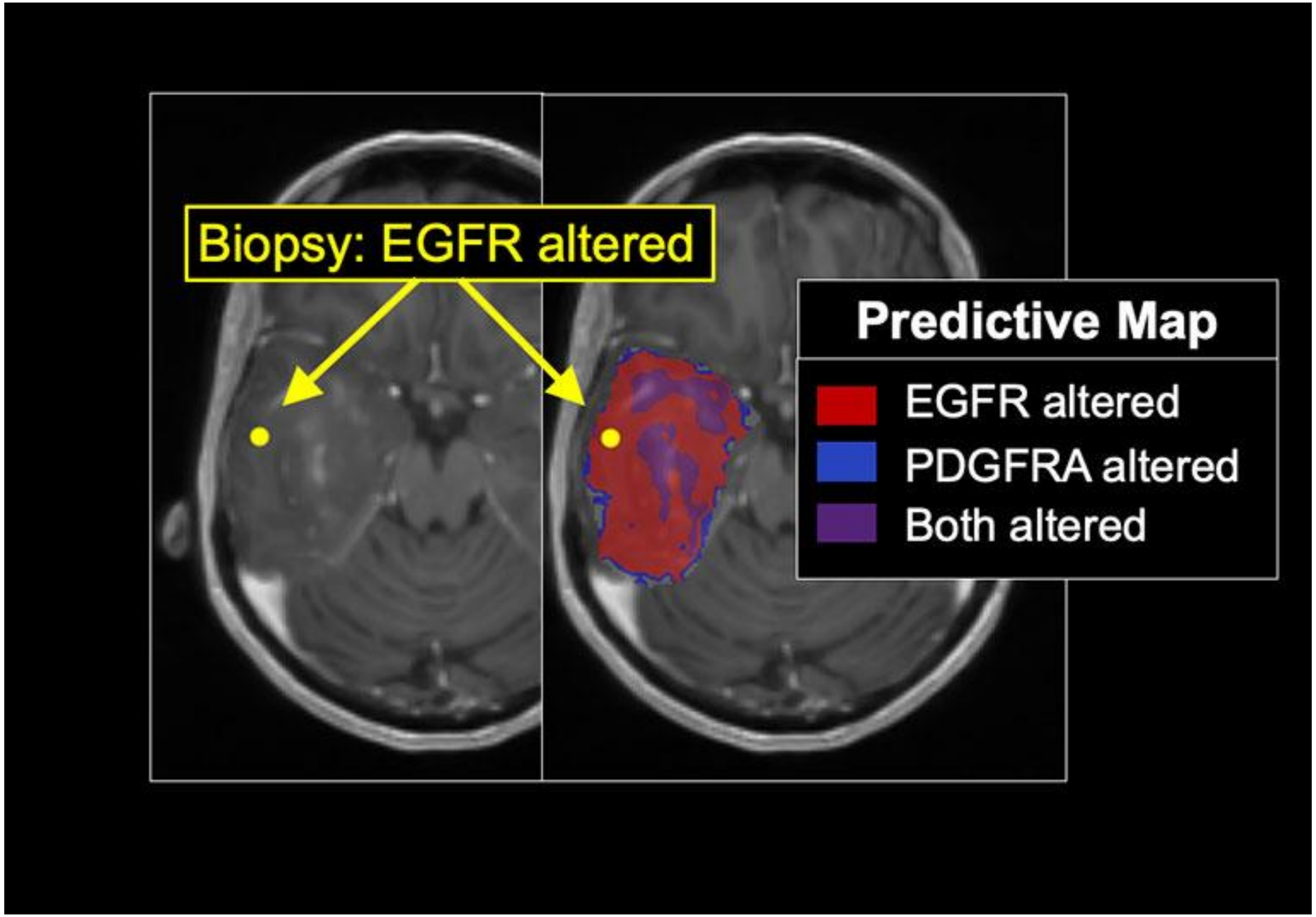
# Montreal Neurological Institute (MNI)



Data: brain map with defined clusters

MNI template

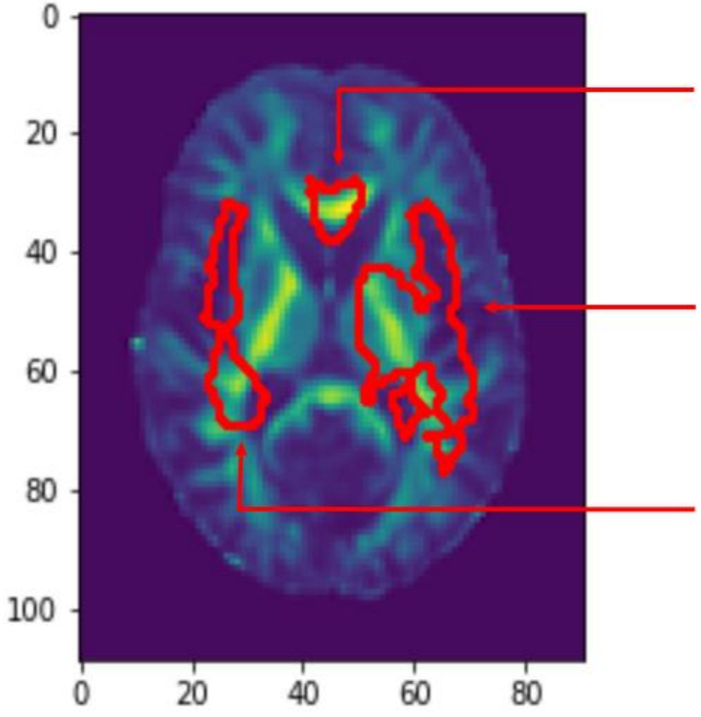
Data: gene expression profiles



**Radiogenomics: Exploring MRI's potential for predicting heterogeneous tumor biology**

Mayo Clinic, March 26, 2022

# Causal Regions of Interest (ROIs) for AD



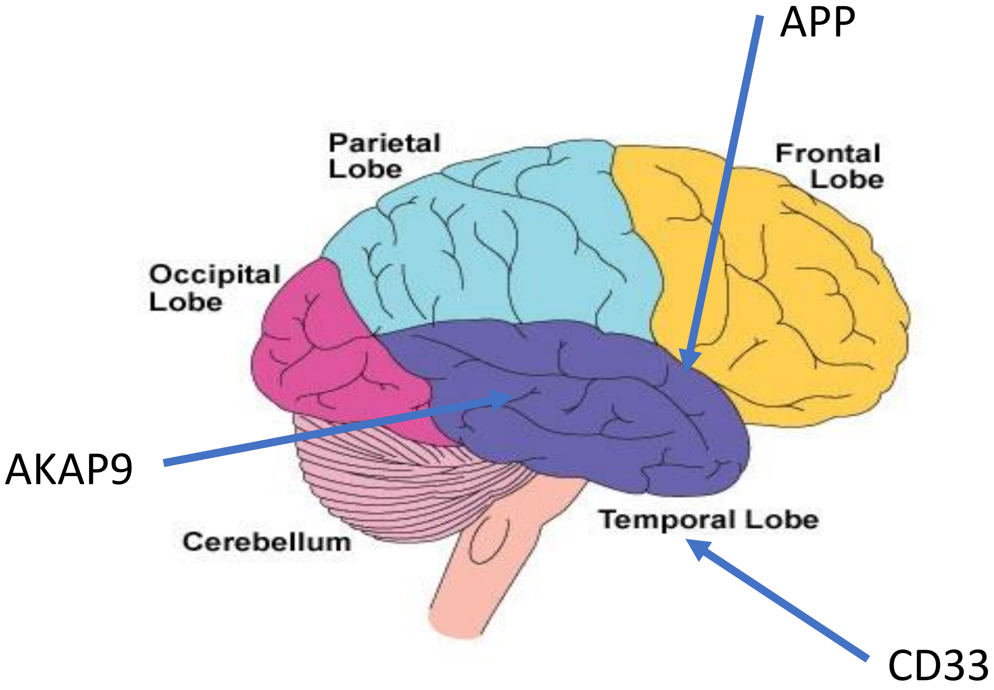
ROI 18 (the ventricles and enlarged ventricle) that was enclosed by red line

ROI 17 (the right temporal lobe) that was enclosed by red line

ROI 14 (the left temporal lobe) that was enclosed by red line

| Time Point | Number of significant ROIs | ROI Index | P-value |
|---|---|---|---|
| **Baseline** | 2 | 18 | 0.0005 |
| | | | |
| **6 months** | 4 | 14 | 0.0108 |
| | | 17 | 0.0155 |
| | | 18 | 0.0010 |
| | | | |
| **12 months** | 4 | 14 | 0.0018 |
| | | 17 | 0.0107 |
| | | 18 | <0.00005 |
| | | | |
| **24 months** | 16 | 14 | <0.00005 |
| | | 17 | 0.0239 |
| | | 18 | <0.00005 |

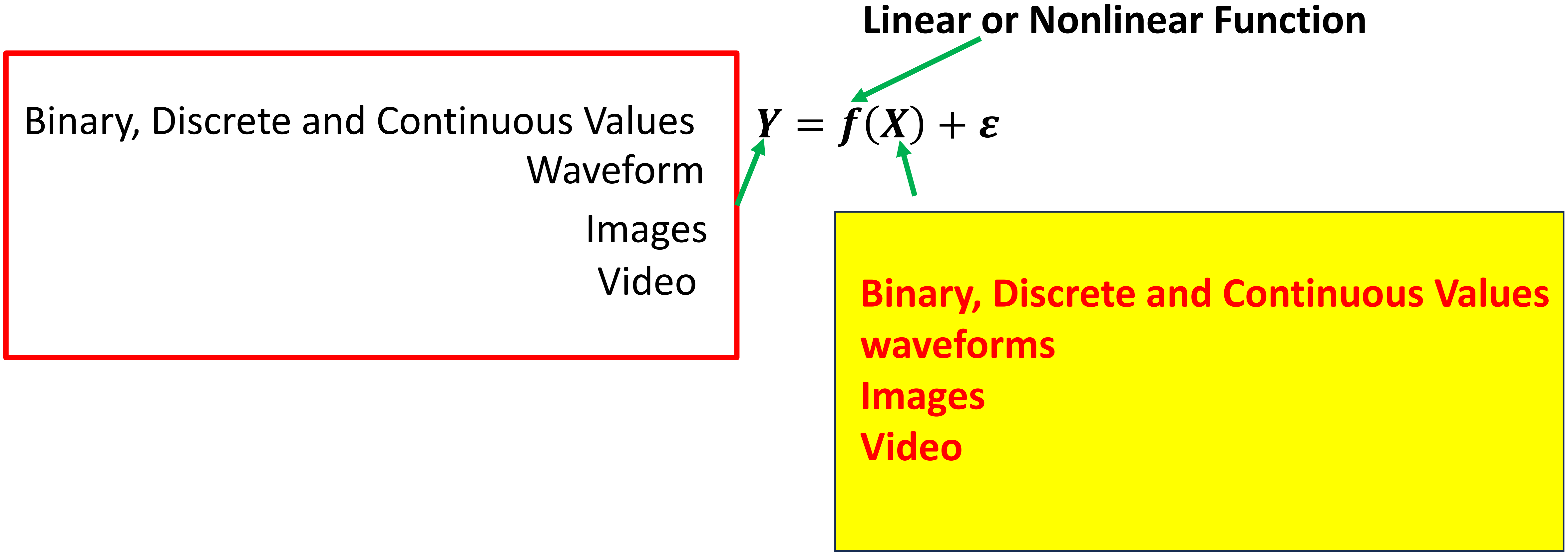
# Causal genes for imaging phenotypes



| | **APP** | **CD33** | **AKAP9** |
|---|---|---|---|
| Location | 21q21.3 | 19q13.3 | 7q21-q22 |
| Pathway | APP processing | Immune response | – |
| Function | Neuronal development, Synaptic formation and repair, β-Amyloid production | Clathrin-mediated endocytosis, Cell signaling | Signal transduction |
| References | Cruts et al.,2012;<br>Yoshikai et al.,1990;<br>Thinakaren and Koo,2008;<br>De Jonghe et al.,2001 | Biffi et al., 2012;<br>Bradshaw et al.,2013 | Logue et al,2014;<br>Venkatesh et al.,2016 |

# Classical Statistical Methods for Genetic Studies of Complex Disease

- **General Model for Diseases:**



**Linear or Nonlinear Function**

$$Y = f(X) + \varepsilon$$

Binary, Discrete and Continuous Values
Waveform
Images
Video

**Binary, Discrete and Continuous Values**
**waveforms**
**Images**
**Video**

**However, current analytical methods are designed to perform a single task, only providing a partial picture of the multi-modal data.**

# Key Question 1:

**How to unify different biological values in different space?**

## Solution

**Embedding: Map them to Euclidean Space**

# Key Question 2:

**How to implement nonlinear functions?**

## Solution

**Hierarchically organized neural networks**

As artificial intelligence technology continues to evolve,

- **various improved multimodality and multi-task masked autoencoder**,
- **X-Decoder which is a generalized decoding model that can predict pixel-level segmentation and language tokens seamlessly,**
- and **joint embedding models** are being developed.
- These models often incorporate novel network architectures and learning strategies to further improve performance and generalization capabilities.

# MC-JEPA: A JOINT-EMBEDDING PREDICTIVE ARCHITECTURE FOR SELF-SUPERVISED LEARNING OF MOTION AND CONTENT FEATURES

# A general framework for self-supervised learning in speech, vision and language

Models and code are available at
www.github.com/pytorch/fairseq/tree/master/examples/data2vec

# OMNIVORE: A Single Model for Many Visual Modalities

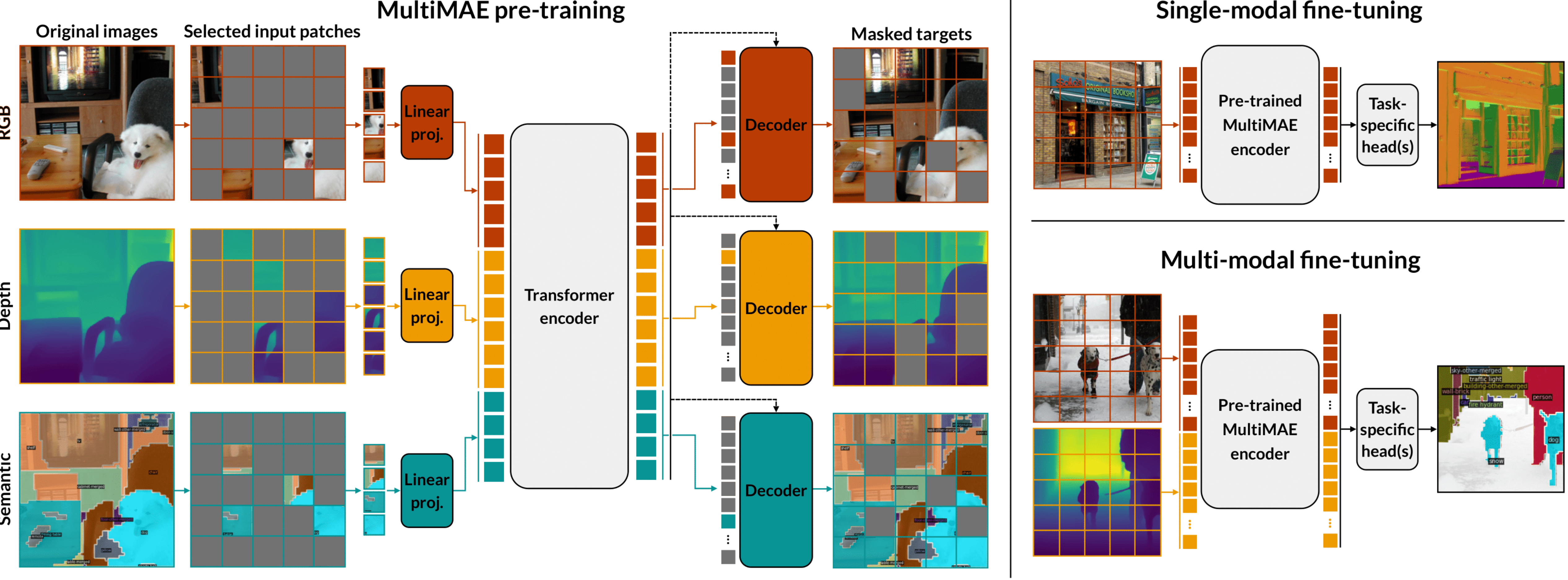
R. Girdhar, M. Singh, N. Ravi, L. van der Maaten, A. Joulin and I. Misra
Omnivore: A Single Model for Many Visual Modalities
IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022
[arXiv] [code/models] [BibTex]

**Efficient and accurate sequence generation with small-scale protein language models**

**Reprogramming Pretrained Language Models for Protein Sequence Representation Learning**

**ProtFIM: Fill-in-Middle Protein Sequence Design via Protein Language Models**

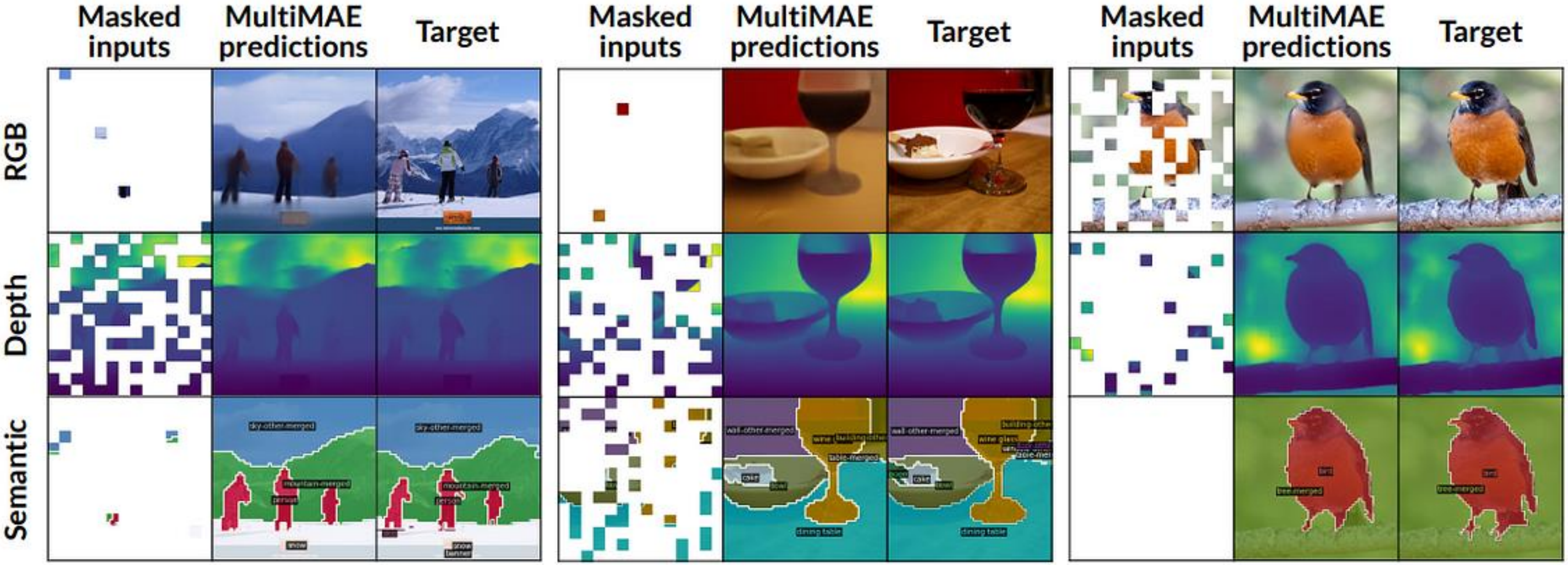
**Structure-informed Language Models Are Protein Designers**



## MultiMAE: Multi-modal Multi-task Masked Autoencoders

Roman Bachmann et al. 2022

https://multimae.epfl.ch



MultiMAE pre-training objective. We randomly select 1/6 of all 16×16 image patches from multiple modalities and learn to reconstruct the remaining 5/6 masked patches from them. The figure shows validation examples from ImageNet, where masked inputs (left), predictions (middle), and non-masked images (right) for RGB (top), depth (middle), and semantic segmentation (bottom) are provided. Since we do not compute a loss on non-masked patches, we overlay the input patches on the predictions.

# Multi-modal encoder

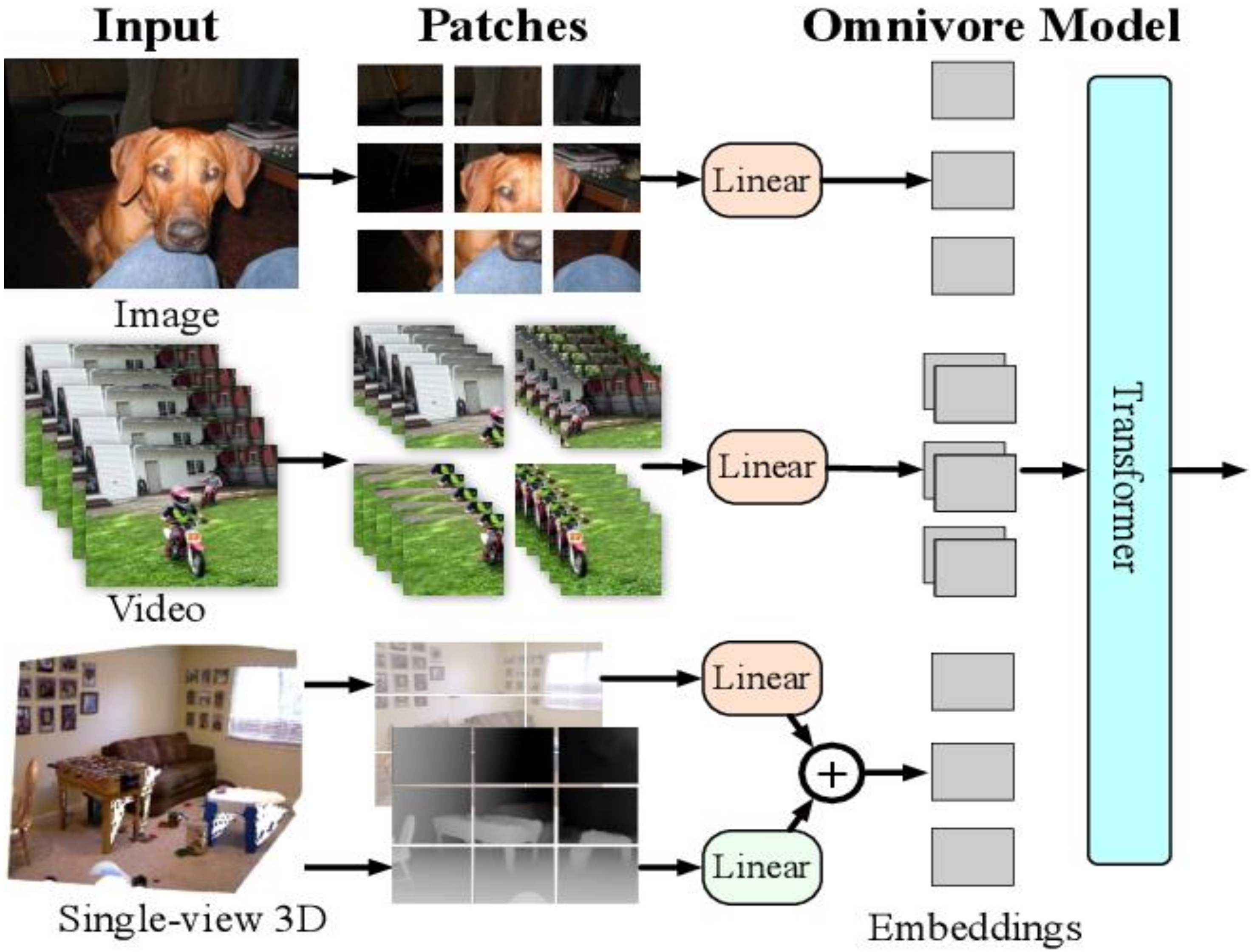
- Our multi-modal Transformer encoder is a ViT [26], but with patch projection layers for each additional input modality.
- 16×16 patches of each modality are projected to tokens with the correct Transformer dimension using a different linear projection for each modality
- Projected patches are concatenated into a sequence of tokens and given as input to the same Transformer encoder.
- We also add an additional global token with a learned embedding, similar to the class-token used in ViT

# Decoders

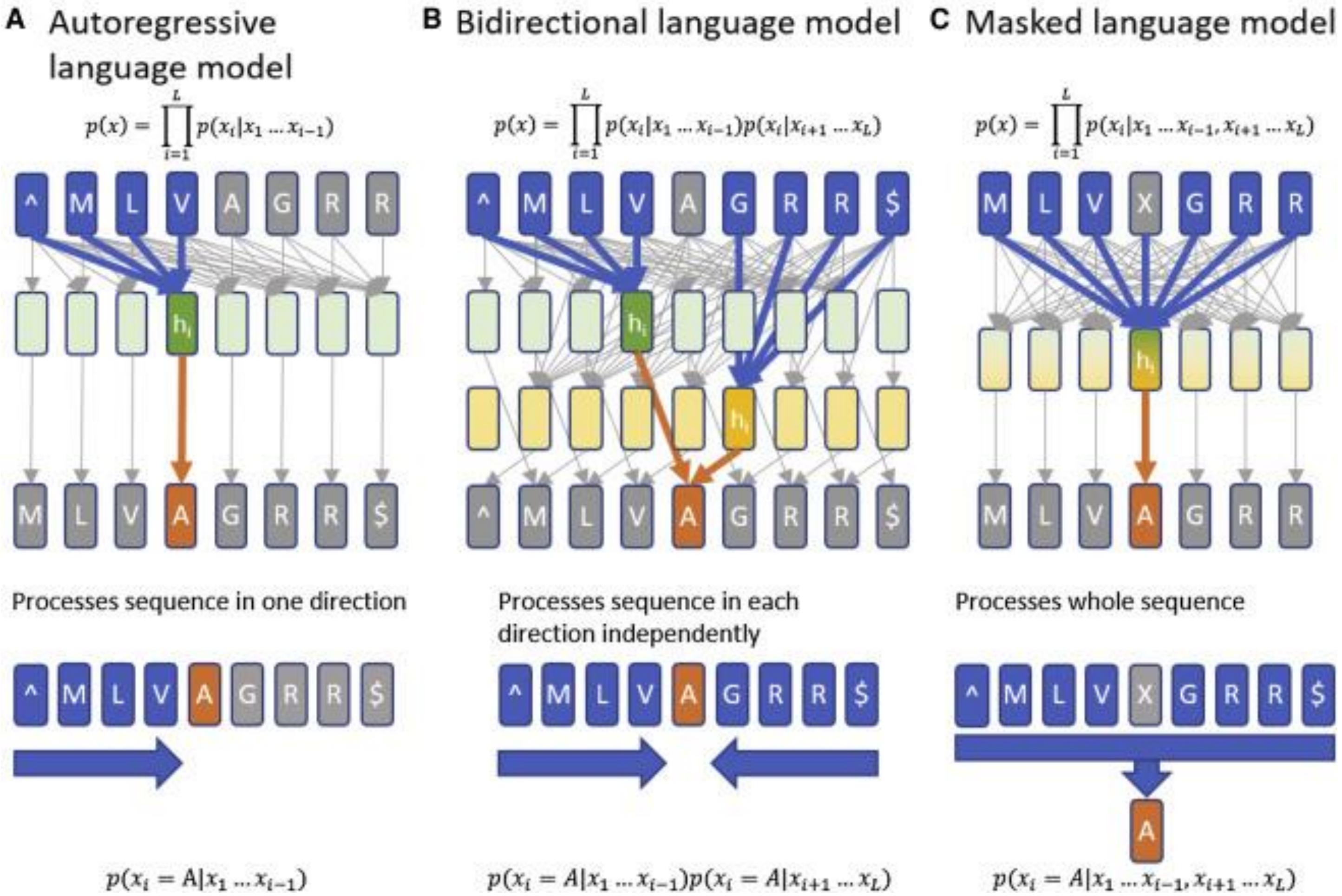
- To reconstruct the masked-out tokens from the visible tokens, we use a separate decoder for each task
- The input to each decoder is the full set of visible tokens from the respective task it is reconstructing. As in MAE [35], these visible tokens are decoded jointly with a set of mask tokens, which serve as placeholders for the decoders to write the reconstructed patches .
- To integrate information from the encoded tokens of other modalities, we add a single cross-attention layer in each decoder using these tokens as queries and all the encoded tokens as keys / values.
- Sine-cosine positional embeddings and learned modality embeddings are added to the tokens before this step. This is then followed by a small MLP and Transformer blocks. Following MAE, we compute the losses only on the masked tokens

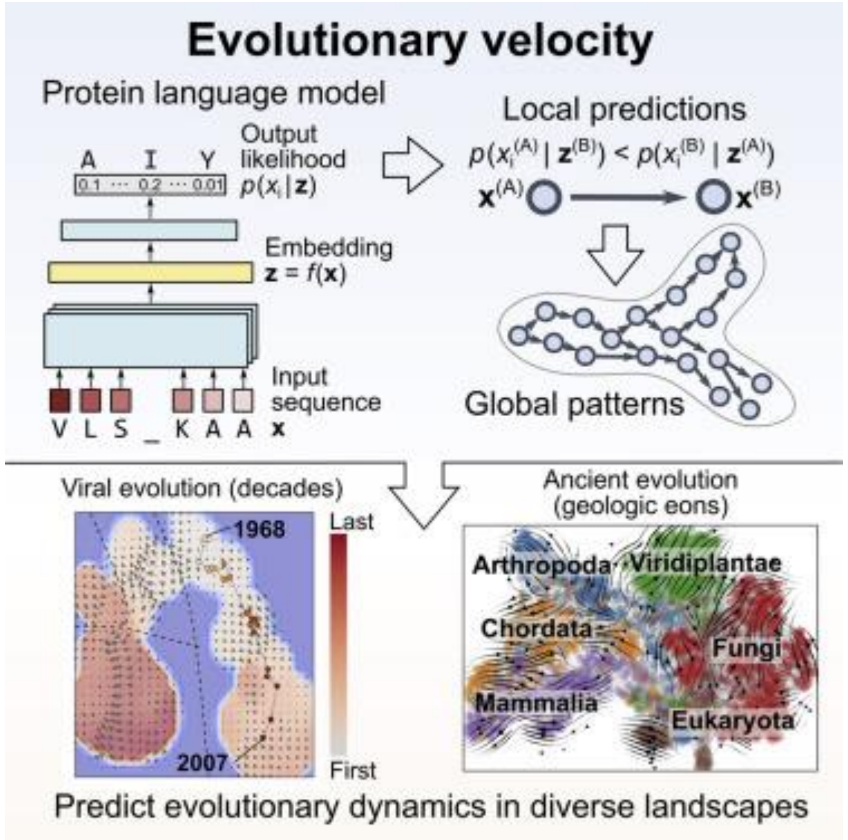
# Transfer tasks

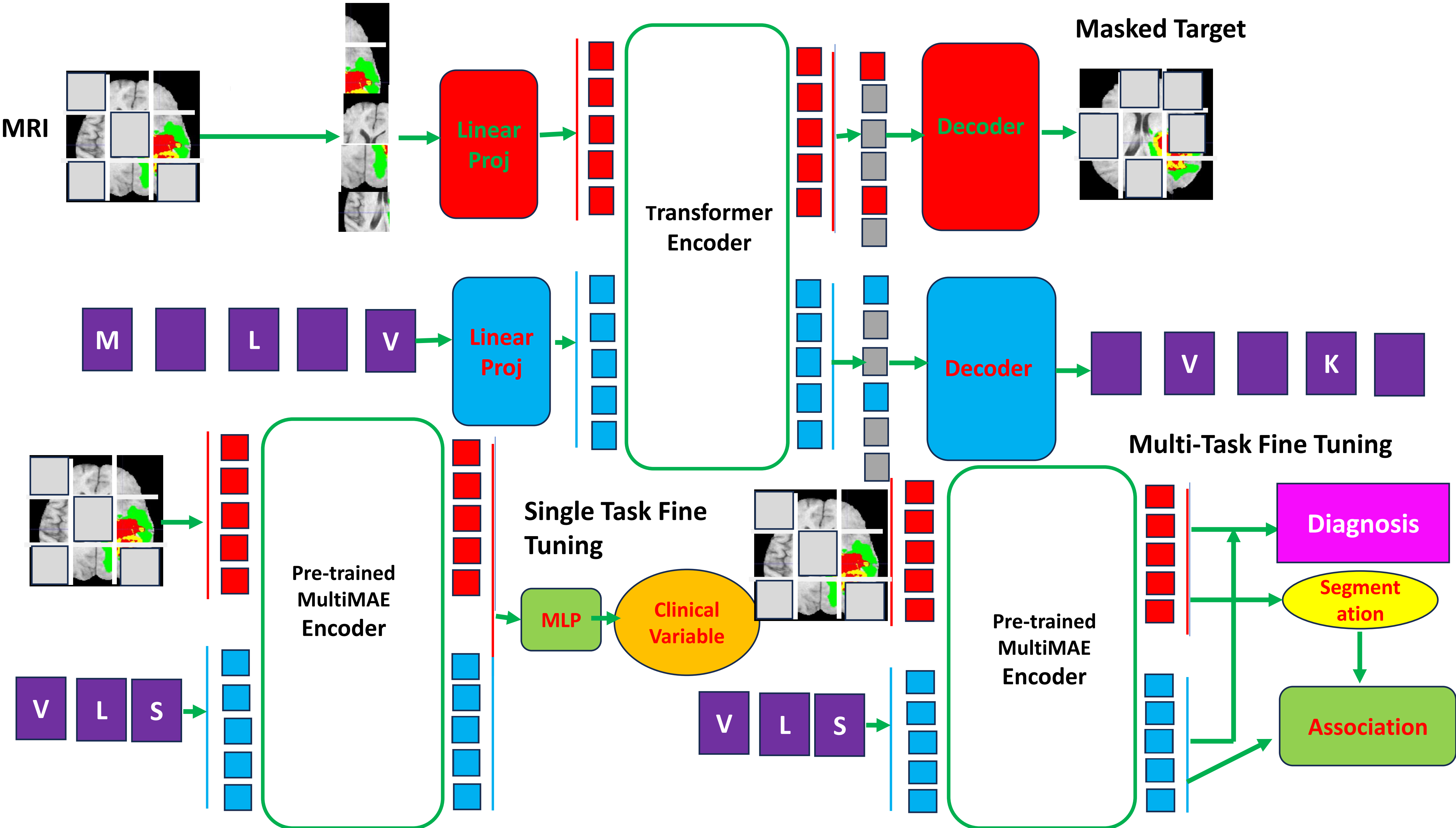
**For all transfers, we replace the pre-trained decoders by randomly initialized taskspecific heads, and train them along with the pre-trained encoder.**



OMNIVORE: A Single Model for Many Visual Modalities


A Autoregressive language model
$p(x) = \prod_{i=1}^{L} p(x_i|x_1 \dots x_{i-1})$
^ M L V A G R R
$h_i$
M L V A G R R $
Processes sequence in one direction
^ M L V A G R R $
$p(x_i = \mathrm{A}|x_1 \dots x_{i-1})$
B Bidirectional language model
$p(x) = \prod_{i=1}^{L} p(x_i|x_1 \dots x_{i-1}) p(x_i|x_{i+1} \dots x_L)$
^ M L V A G R R $
$h_i$
$h_i$
^ M L V A G R R $
Processes sequence in each direction independently
^ M L V A G R R $
$p(x_i = A|x_1 \dots x_{i-1}) p(x_i = A|x_{i+1} \dots x_L)$
C Masked language model
$p(x) = \prod_{i=1}^{L} p(x_i|x_1 \dots x_{i-1}, x_{i+1} \dots x_L)$
M L V X G R R
$h_i$
M L V A G R R
Processes whole sequence
^ M L V X G R R $
A
$p(x_i = A|x_1 \dots x_{i-1}, x_{i+1} \dots x_L)$


Evolutionary velocity
Protein language model
A I Y
0.1 ··· 0.2 ··· 0.01
Output likelihood $p(x_i\,|\,\mathbf{z})$
Embedding $\mathbf{z} = f(\mathbf{x})$
Input sequence
V L S _ K A A $\mathbf{x}$
Local predictions
$p(x_i^{(A)}\,|\,\mathbf{z}^{(B)}) < p(x_i^{(B)}\,|\,\mathbf{z}^{(A)})$
$\mathbf{x}^{(A)}$
$\mathbf{x}^{(B)}$
Global patterns
Viral evolution (decades)
1968
2007
Last
First
Ancient evolution (geologic eons)
Arthropoda
Viridiplantae
Chordata
Fungi
Mammalia
Eukaryota
Predict evolutionary dynamics in diverse landscapes


MRI
Linear Proj
Transformer Encoder
Decoder
Masked Target
M
L
V
Linear Proj
Decoder
V
K
Multi-Task Fine Tuning
Pre-trained MultiMAE Encoder
Single Task Fine Tuning
MLP
Clinical Variable
V
L
S
Pre-trained MultiMAE Encoder
V
L
S
Diagnosis
Segmentation
Association

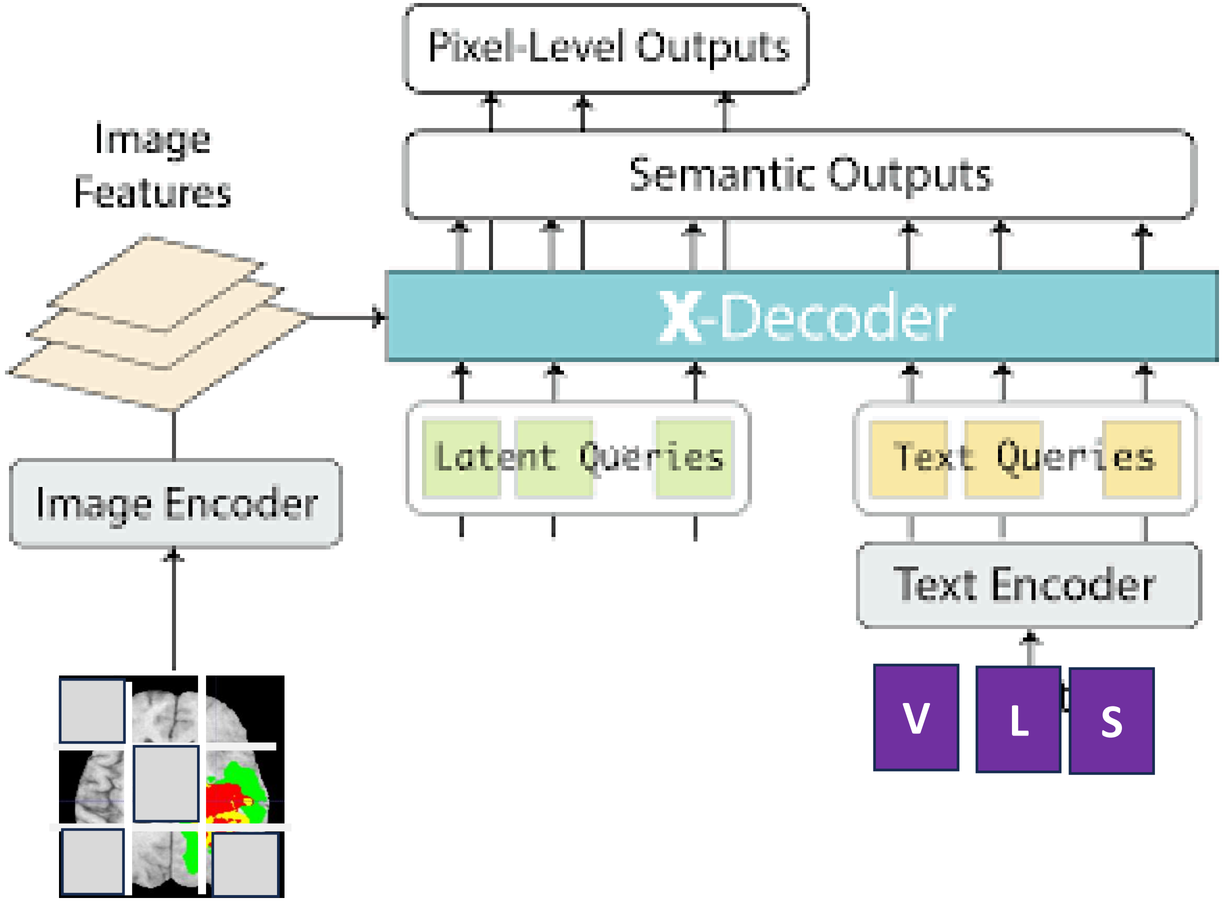
# Generalized decoding for pixel, image, and language.

Code, demo, video and visualization are available at: https://x-decodervl.github.io.

# Segment everything everywhere all at once

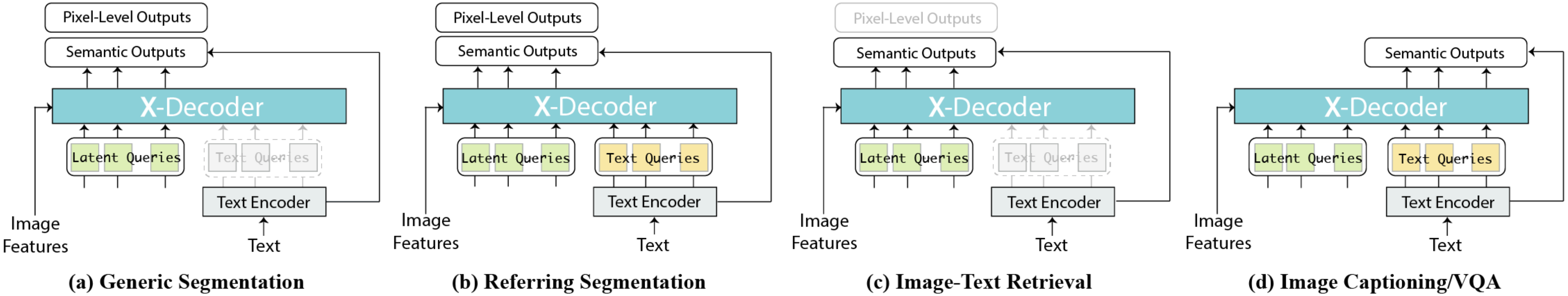
# Masked-attention Mask Transformer for Universal Image Segmentation

https://bowenc0221.github.io/mask2former

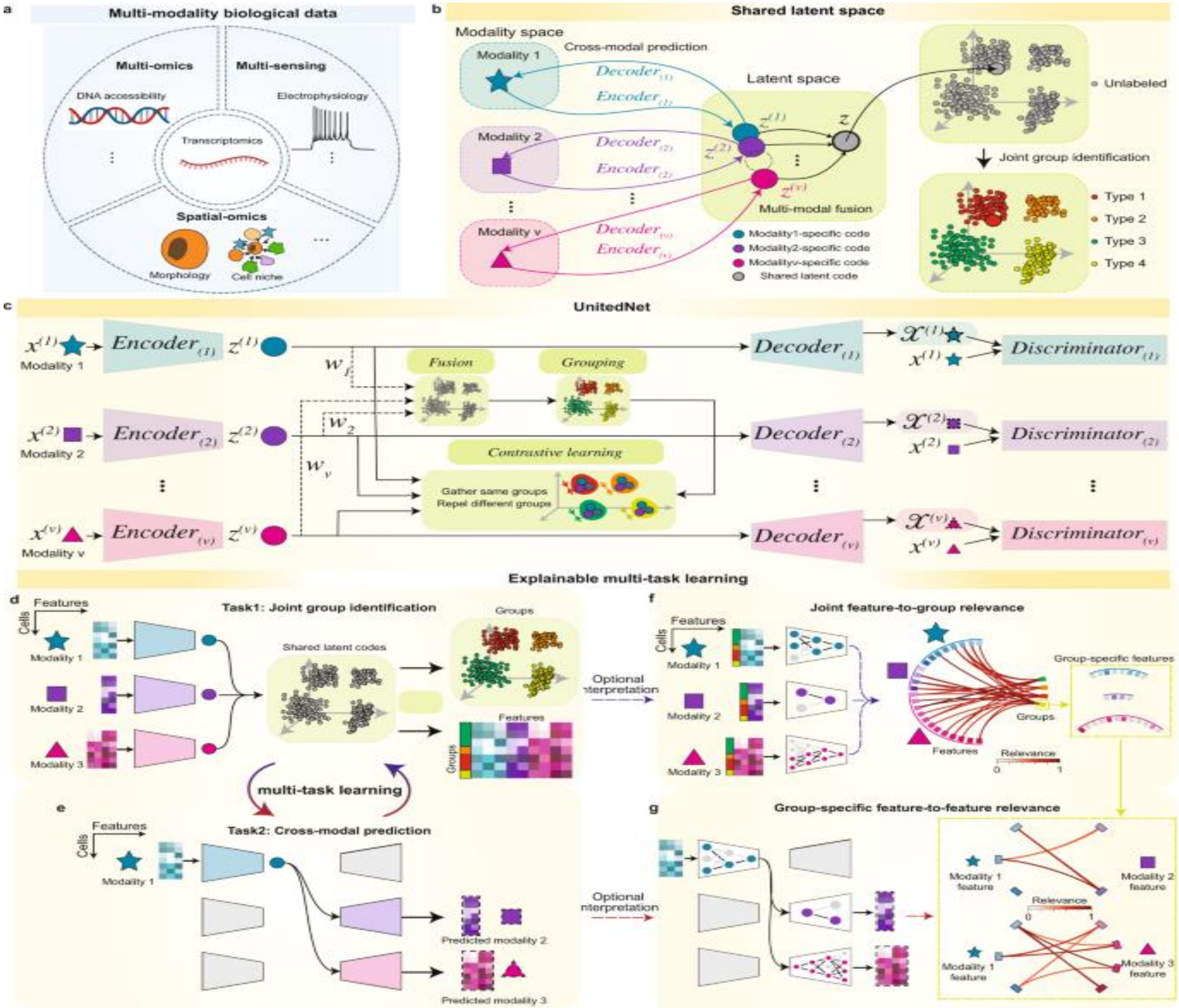


Generalized decoding for pixel, image, and language

Xueyan Zou el a. 2023

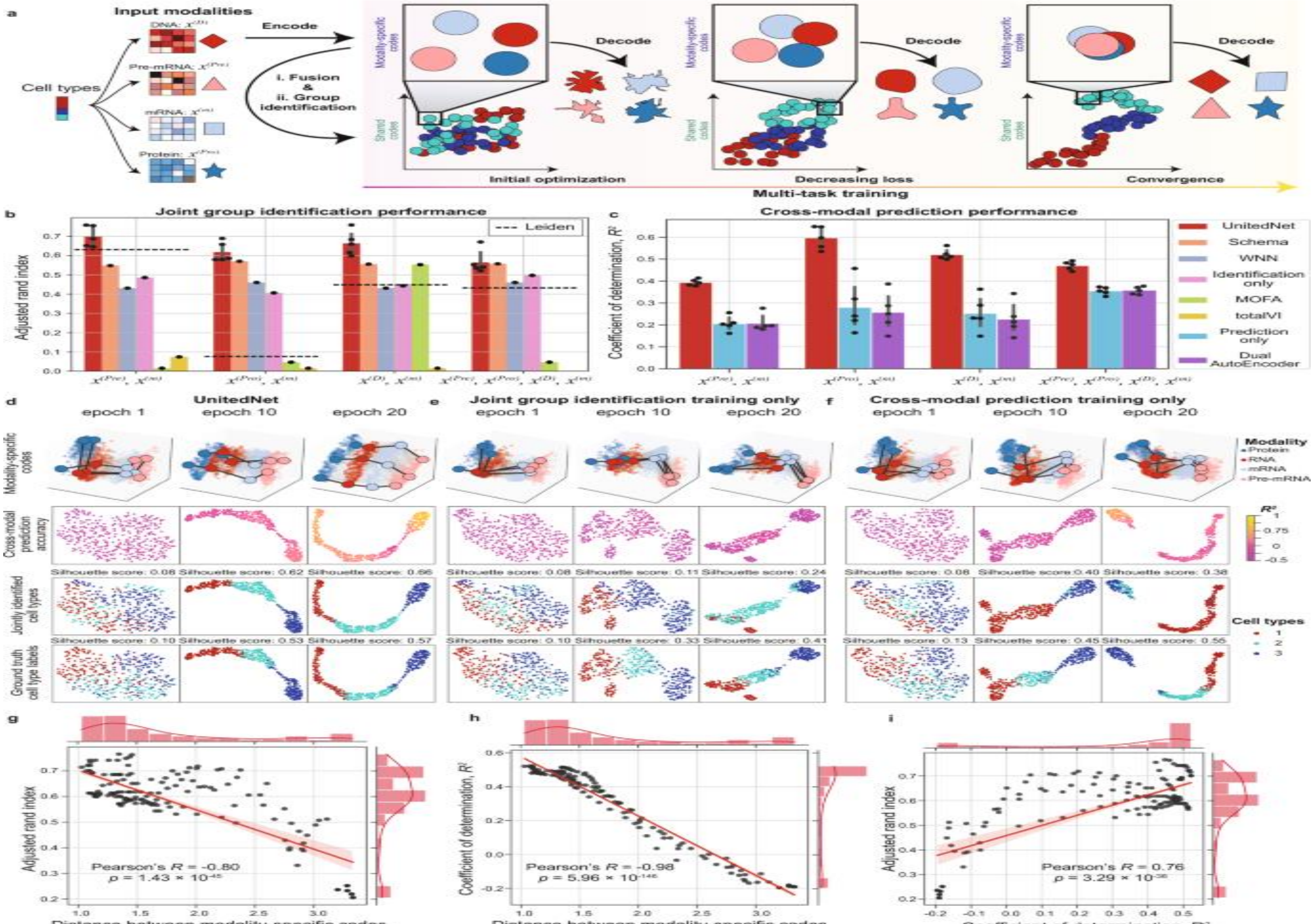


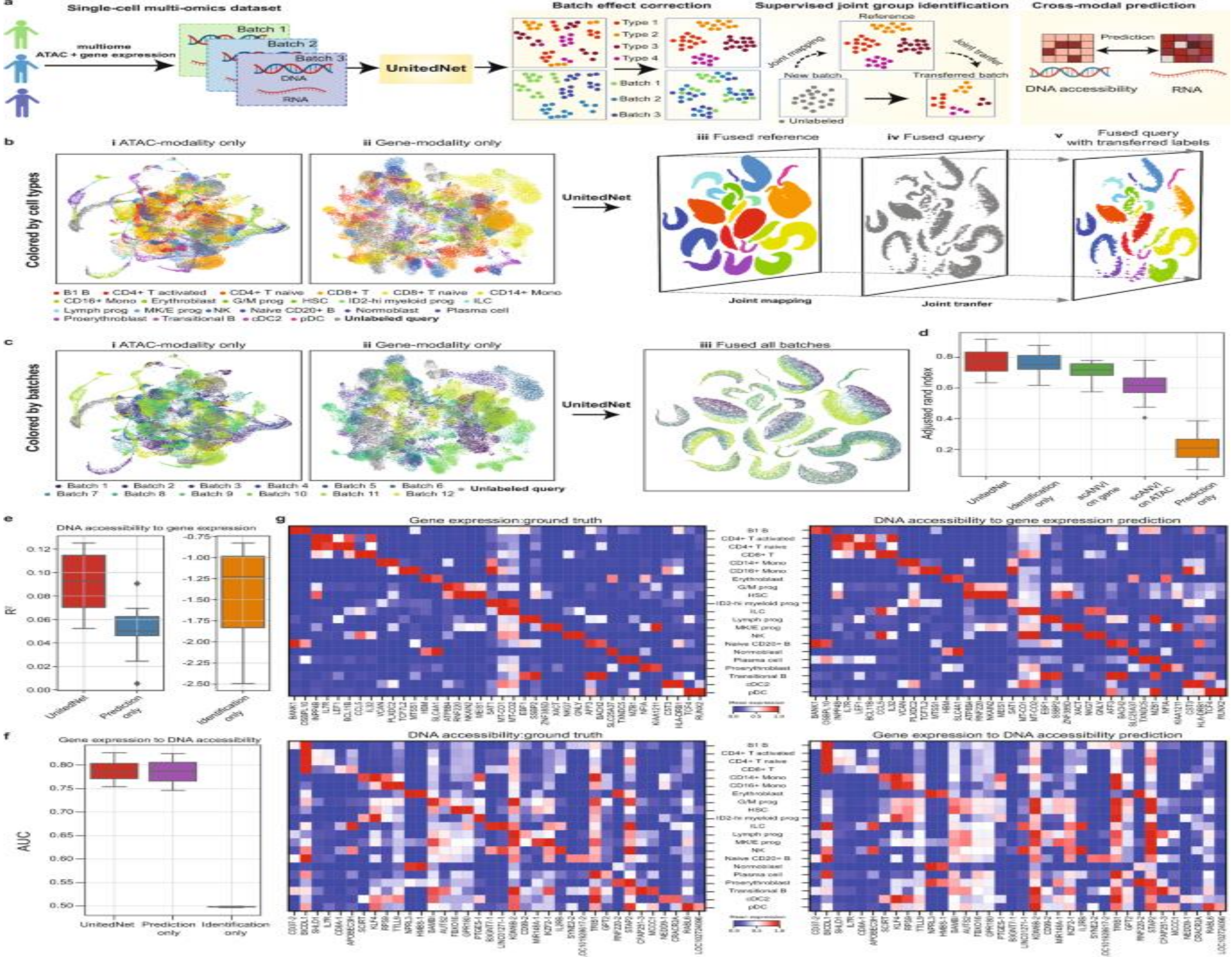
Unifying four different types of tasks with our proposed X-Decoder. From left to right, they are: (a) generic semantic/instance/panoptic segmentation; (b) referring segmentation; (c) image-text retrieval and (d) image captioning and VQA. The components with white text indicate not applied



# Explainable multi-task learning for multi-modality biological data analysis

Xin Tang et al. 2023


a
Input modalities
DNA: $x^{(D)}$
Pre-mRNA: $x^{(Pre)}$
mRNA: $x^{(m)}$
Protein: $x^{(Pro)}$
Cell types
Encode
i. Fusion & ii. Group identification
Modality-specific codes
Shared codes
Decode
Initial optimization
Decreasing loss
Convergence
Multi-task training
b
Joint group identification performance
Adjusted rand index
Leiden
$x^{(Pre)}, x^{(m)}$
$x^{(Pro)}, x^{(m)}$
$x^{(D)}, x^{(m)}$
$x^{(Pre)}, x^{(Pro)}, x^{(D)}, x^{(m)}$
c
Cross-modal prediction performance
Coefficient of determination, $R^2$
UnitedNet
Schema
WNN
Identification only
MOFA
totalVI
Prediction only
Dual AutoEncoder
d
UnitedNet
e
Joint group identification training only
f
Cross-modal prediction training only
epoch 1
epoch 10
epoch 20
Modality-specific codes
Cross-modal prediction accuracy
Jointly identified cell types
Ground truth cell type labels
Modality
Protein
RNA
mRNA
Pre-mRNA
$R^2$
Cell types
1
2
3
Silhouette score: 0.08
Silhouette score: 0.62
Silhouette score: 0.66
Silhouette score: 0.10
Silhouette score: 0.53
Silhouette score: 0.57
Silhouette score: 0.08
Silhouette score: 0.11
Silhouette score: 0.24
Silhouette score: 0.10
Silhouette score: 0.33
Silhouette score: 0.41
Silhouette score: 0.08
Silhouette score:0.40
Silhouette score: 0.38
Silhouette score: 0.13
Silhouette score: 0.45
Silhouette score: 0.55
g
Adjusted rand index
Pearson's $R$ = -0.80
$p = 1.43 \times 10^{-45}$
Distance between modality-specific codes
h
Coefficient of determination, $R^2$
Pearson's $R$ = -0.98
$p = 5.96 \times 10^{-146}$
Distance between modality-specific codes
i
Adjusted rand index
Pearson's $R$ = 0.76
$p = 3.29 \times 10^{-36}$
Coefficient of determination, $R^2$


a
Single-cell multi-omics dataset
multiome
ATAC + gene expression
Batch 1
Batch 2
Batch 3
DNA
RNA
UnitedNet
Batch effect correction
Type 1
Type 2
Type 3
Type 4
Batch 1
Batch 2
Batch 3
Supervised joint group identification
Reference
Joint mapping
Joint tranfer
New batch
Transferred batch
Unlabeled
Cross-modal prediction
Prediction
DNA accessibility
RNA
b
Colored by cell types
i ATAC-modality only
ii Gene-modality only
UnitedNet
iii Fused reference
iv Fused query
v Fused query with transferred labels
Joint mapping
Joint tranfer
B1 B
CD4+ T activated
CD4+ T naive
CD8+ T
CD8+ T naive
CD14+ Mono
CD16+ Mono
Erythroblast
G/M prog
HSC
ID2-hi myeloid prog
ILC
Lymph prog
MK/E prog
NK
Naive CD20+ B
Normoblast
Plasma cell
Proerythroblast
Transitional B
cDC2
pDC
Unlabeled query
c
Colored by batches
i ATAC-modality only
ii Gene-modality only
UnitedNet
iii Fused all batches
Batch 1
Batch 2
Batch 3
Batch 4
Batch 5
Batch 6
Unlabeled query
Batch 7
Batch 8
Batch 9
Batch 10
Batch 11
Batch 12
d
Adjusted rand index
0.8
0.6
0.4
0.2
UnitedNet
Identification only
scANVI on gene
scANVI on ATAC
Prediction only
e
DNA accessibility to gene expression
R²
0.12
0.10
0.08
0.06
0.04
0.02
0.00
-0.75
-1.00
-1.25
-1.50
-1.75
-2.00
-2.25
-2.50
UnitedNet
Prediction only
Identification only
f
Gene expression to DNA accessibility
AUC
0.80
0.75
0.70
0.65
0.60
0.55
0.50
UnitedNet
Prediction only
Identification only
g
Gene expression:ground truth
DNA accessibility to gene expression prediction
DNA accessibility:ground truth
Gene expression to DNA accessibility prediction
Mean expression
0.0
0.5
1.0
BANK1
OSBPL10
INPP4B
IL7R
LEF1
BCL11B
CCL5
IL32
VCAN
PLXDC2
TCF7L2
MTSS1
HBM
SLC4A1
ATP8B4
RNF220
NKAIN2
MEIS1
SAT1
MT-CO1
MT-CO3
EBF1
SSBP2
ZNF385D
XACT
NKG7
GNLY
AFF3
BACH2
SLC25A37
TXNDC5
MZB1
NFIA
KIAA1211
CST3
HLA-DRB1
TCF4
RUNX2
CD37-2
BICDL1
SHLD1
IL7R
CD84-1
APOBEC3H
SCIRT
KLF4
RPS6
TTLL3
NPRL3
HMBS-1
BANK1
AUTS2
TBXDC18
GPR183
PTGES-1
BIGKIT1-1
LINC02127-1
KDM4B-2
CD99-2
MIR146A-1
IKZF2-1
IL2RB
SYNE2-2
LOC101928817-2
TRIB1
GPT2
RNF220-2
STAP2
CFAP251-3
MCCC1
NEDD9-1
CRACR2A
RAB4A
LOC102724096

## Explainable multi-task learning for multimodality biological data analysis

Xin Tang et al. 2023

Jia Liu,

Jie Ding, dingj@umn.edu; jia_liu@seas.harvard.edu

## AdaMTL: Adaptive Input-dependent Inference for Efficient Multi-Task Learning

## Dynamic Weighted Multitask Learning and Contrastive Learning for Multimodal Sentiment Analysis

## Advances and Challenges of Multi-task Learning Method in Recommender System: A Survey

# Multitask Attention-Based Neural Network for Intraoperative Hypotension Prediction