## DIYgenomics AI-first Digital Biology Methods for Disease and Longevity - 12/5/23

Melanie Swan, Renato P. dos Santos, Takashi Kido, Eric Roland, Robert Dowd (arXiv: 2307.02502)

- Aim: AI methods to facilitate and extend experimentalist efforts
- Approach: interact with biology at the level of math not data
- Very big data: imaging, PPI (OhmNet), genomics, pathway (Kegg), evolutionary DBs

Constant first-principles thinking to identify unexpected approaches

- 1. Protein networks approach, disease graphs, molecular therapies
  - Protein-to-gene not gene-to-protein (AlphaMissense enabled by AlphaFold)
  - Protein-protein interaction networks: nodes (proteins) and edges (interactions)
- 2. Drug design not drug discovery (Isomorphic Labs, Stokes 2020 (Halicin), Bronstein 2020 (MaSIF))
- 3. Treat pathway not condition (Kellis, DIYgenomics) with 10-drug one-shot intervention
- 4. Mobilize short-cuts, pre-connected latent space, synthetic data gen, open DBs (scale.ai)
  - Evolutionary DB tip-offs: changing genes suggest spatial proximity (AlphaFold)
  - OCR math equation extraction from PDFs unnecessary, exists in Stack Exchange

Interacting with biosystems at the level of mathematics (plus data) as more expedient lever

- Cancer: spectral graph Laplacian of cancer genome (Braun, 2012, Hearing the Shape of Cancer)
- Genomics: 25% total mutations neutral effect per number theory (Tagaki funct) (Ord 2023)
- Genomics: 60% total mutations missense; 32% pathogenic per asymmetry (AlphaMissense)
  - Equivariance (allowable displacement-rotational symmetries define poss. space)
  - Laplacian matrices (meta-mathematical measure of pairwise distance relations)
  - Position graph representation learning, neural algorithmic reasoning (math agent)

Longevity same implied AI-solution of very large data + straightforward self-learning algorithms

- RL wellness agent, SIR longevity (cycle through sustaining-intervening-recovering pool)
- AI math agent disease self-play competitions with protein-protein organoid tissue DBs
- Mechanism-intervention (immunotherapies, epigenetic editing (CpG methylations), TEs)

	Supervised Learning	Unsupervised Learning	<b>Reinforcement Learning</b>
Input	Labeled data	Unlabeled data	State-action pairs
Goal	Learn a function	Learn underlying structure	Maximize future rewards
Search	Classification (Relational DB)	Semantic search (Vector DB)	Exploration-exploitation
Example	Image recognition, translation	GPTs, LLMs, AlphaFold	AlphaGo, AlphaStar
Apple Ex.	This thing is an apple	This thing looks like other apples	Eat this thing to stay alive
DIYgenomics		Biomarker-Pathway-Gene-SNP	Math agent disease self-play

Table 1. Longevity and Health Preservation: Apply Mix-Match of all Classes of AI Learning Problems.

## DIYgenomics.org Vector DB Semantic Search Project for Human Aging

Concept: apply biomarkers-pathways of aging to personal genome profiles for precision health Project Suitability: multimodal data, large public data stores, semantic relationships unmapped

	Data Tier	Sample Data Values	Data Source (pilot)	Data Source (at-scale)
1	Biomarker	Epigenetic clocks, glycans,	Moqri et al., 2023, Cell 186(18);	Larger scope of "human
		telomere, inflammation,	Aging Biomarker Consortium, 2023,	biomarkers of aging"
		mitochondrial function	Science China Life Sciences 66.	
2	Pathway	Metabolism, etc.	Single pathway	Kegg, BioCYC
3	Gene	APOE, APP, PSEN, TP-53	Nebula, DIY genomics, Gene Card	AlphaMissense, dbSNP
4	SNP	rs7412, rs5215	Citizen 1 & Citizen 2 WGS VCFs	OpenSNP (n=7,839)