Universal Multilayer Network Embedding Reveals a Causal Link Between GABA Neurotransmitter and Cancer

Léo Pio-Lopez[1] and Michael Levin[1,2*]

[1] *Allen Discovery Center, Tufts University, Medford, MA, USA*
[2] *Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115, USA*
[*] *corresponding author: michael.levin@tufts.edu*

Keywords— network embedding, multi-layer network, drug repositioning, GABA, cancer, AI, machine learning, drug discovery

Running title— Multilayer network embedding

Abstract

The volume and complexity of biological data have significantly increased in recent years, often represented as network models continue to increase at a rapid pace. However, drug discovery in the context of complex phenotypes are hampered by the difficulties inherent in producing machine learning algorithms that can integrate molecular-genetic, biochemical, physiological, and other diverse datasets. Recent developments have expanded network analysis techniques, such as network embedding, to effectively explore multilayer network structures. Multilayer networks, which incorporate various nodes and connections in formats such as multiplex, heterogeneous, and bipartite networks, provide an effective framework for merging diverse and multi-scale biological data sources. However, current network embedding methods face challenges and limitations in addressing the heterogeneity and diversity of these networks. Therefore, there is an essential need for the development of new network embedding methods to manage the complexity and diversity of multi-omics biological information effectively. Here, we report a universal multilayer network embedding method MultiXVERSE, which is to the best of our knowledge the first one capable of handling any kind of multilayer network. We applied it to a molecular-drug-disease multiplex-heterogeneous network. Our model made new predictions about a link between GABA and cancer that we verified experimentally in the Xenopus laevis model.

1 Introduction

Network graph models are highly effective for depicting real-world objects through their relationships and interactions [1]. They offer valuable insights into the connections between different entities and are utilized as tools to investigate complex systems across various fields [2, 3, 4]. A significant challenge in machine learning involves converting high-dimensional graph-based data into a feature vector. Indeed, these methods were originally designed for vector data and cannot be directly applied to biological datasets such as biological networks [5, 6]. Network embedding, also known as graph representation learning, addresses this issue by transforming network data into formats compatible with conventional machine learning tools, thereby broadening the scope of machine learning applications in network analysis.

Network embedding techniques have proven highly effective across numerous applications, including community detection, node classification, and link prediction. Capable of handling vast networks with millions of nodes, these techniques are particularly valuable in the era of big data. Consequently, network embeddings are increasingly used to analyze various large-scale networks, such as social [3], neuronal [4] or molecular networks [2, 7, 8].

The volume and complexity of biological data have significantly increased in recent years, often represented as multilayer network models [1, 9]. Multilayer networks, which incorporate various nodes and connections in formats such as multiplex, heterogeneous, and bipartite networks, provide an especially effective framework for merging diverse and multi-scale biological data sources [1, 9]. However, current network embedding methods face challenges and limitations in addressing the heterogeneity and diversity of these networks [8]. Therefore, there is an essential need for the development of new network embedding methods to manage the complexity and diversity of multilayer networks effectively.

In this work, we extended MultiVERSE [8], a multilayer network embedding algorithm tailored for the application of machine learning techniques to these multilayer networks but limited to a maximum of two multiplex networks. MultiVERSE is based on the VERSE framework [10], and coupled with Random Walks with Restart (RWR) on multilayer networks [9]. Recently, RWR has been extended to a universal random walk with restart using a method called MultiXrank [11] allowing the exploration of any kind of multilayer networks. We extended MultiVERSE with MultiXrank and it is now a universal multilayer network embedding method that we named MultiXVERSE. Our method can handle any multilayer network defined as a composition of various multiplex and monoplex networks interconnected through bipartite interaction networks (see Figure 3 for an example). Within this multilayer structure, each network may also be weighted and/or directed. And we can add as many multiplex and bipartite networks as we want with this extension without limitations, except of course computational power.

Consequently, MultiXVERSE provides a means to network embedding on these multilayer networks, which are characterized by their rich and complex interactions. This approach is particularly effective in representing the multi-scale interactions typically observed in biological systems. For biology, this approach allows to aggregate network data from drugs, diseases, genes, patients etc... in the same network representation and machine learning can be applied on the resulting embeddings for a wide variety of application including drug repositioning, new predicted gene-disease or drug-target links, the discovery of specific biological functional modules for diseases integrating genes and drugs etc... To the best of our knowledge, it is the first time that network embedding can be applied to any kind of multilayer networks without any limitations on the number of multiplex networks or the type of neworks (weighted, directed, undirected) .

In this article, we applied MultiXVERSE to a biological multilayer network containing data on gene, drug, and diseases interactions and evaluated the quality of the embedding using link prediction (a standard approach in multilayer network embedding [8, 12, 13]). Second, we clustered the embeddings to find functional biological modules, which revealed new predictions of

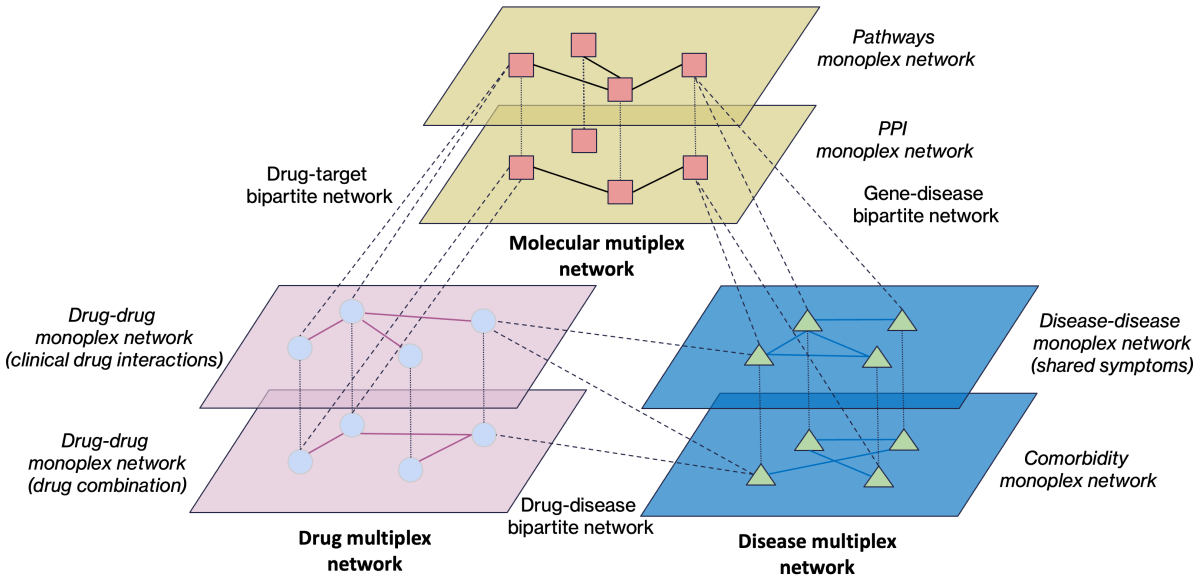

Figure 1: A universal multilayer network consisting of 3 distinct multiplex networks (gene, drig and disease), each represented by different colors (yellow, purple, and blue). Each of these multiplex networks consists of various types of nodes, squares are genes, circles are drugs and triangles are diseases. They are interconnected through three bipartite networks (gene-disease, drug-target and drug-disease), which are visualized here as bipartite interactions for clarity. The number of multiplex networks and layers inside the multiplex networks is arbitrary and could be more.

a link between GABA and cancer. Third, we applied link prediction on the embeddings of GABA agonists drugs and found new links between GABA receptors and cancer. Neurotransmitters are emerging targets in cancer [14, 15] and as such is a case study particularly suited for evaluating our system. Finally, we experimentally tested the new prediction of GABA as a potential cause of cancer in a tadpole melanocyte model, and validated the prediction linking GABA-modulating drugs to a cancer-like phenotype in the absence of classic carcinogens, oncogenes, or DNA damage as the initiator of cellular conversion.

2 Materials and methods

2.1 MultiXVERSE: a universal multilayer network embedding

We extended MultiVERSE [8] to MultiXVERSE for universal multilayer network embedding . This method computes the similarities between nodes using random walks with restart on any kind of multilayer network [11] and optimizes the embeddings using Kullback-Leibler minimization. We present the general method of MultiXVERSE in this section.

Within the MultiXVERSE framework, it is necessary to formulate a similarity metric for the multiplex-heterogeneous network, designated as G_{MH}. This metric, denoted $sim_G : V_{MH} \times V_{MH} \to \mathbb{R}$, maps pairs of nodes within V_{MH} to a real number, reflecting their level of similarity. It is defined as follows:

$$\forall v \in V_{MH}, \sum_{u \in V} sim_G(v, u) = 1\,. \tag{1}$$

Hence, $sim_G(v,.)$, which signifies the similarity metric for any given node v within the multiplex-heterogeneous network G_{MH}, is conceptualized as a probability distribution. Given that, one can obtain the normalized similarity distribution within the embedding space by applying the softmax function. Formally, let w_i represent the embedding of node i within this space. Consequently, the similarity between the embeddings of two nodes w_u and w_v is characterized by the dot product $w_u \cdot w_v^T$, yielding the following expression:

$$sim_{Emb}(v,.) = \frac{exp(w_v \cdot w^T)}{\sum_{i=1}^{n} exp(w_v \cdot w_i)}\,. \tag{2}$$

The purpose of MultiXVERSE is to closely estimate the similarity distribution within the embedding space, represented as $sim_{Emb} : V_{MH} \times V_{MH} \to \mathbb{R}$, such that for all v in V_{MH}, the relationship $sim_G(v,.)$ is approximated by $sim_{Emb}(v,.)$. The optimization during the learning phase is executed through the minimization of the Kullback-Leibler divergence between the two similarity measures:

$$\sum_{v \in V_M} KL(sim_G(v,.) \parallel sim_{Emb}(v,.)) \tag{3}$$

By keeping only the terms related to sim_{Emb} as sim_G is constant, we derive the objective function as follows:

$$\mathcal{L} = -\sum_{v \in V_M} sim_G(v,.)\ log(sim_{Emb}(v,.)) \tag{4}$$

At each iteration, since sim_{Emb} is constructed as a softmax function, it necessitates normalization across the entire network's nodes, a process that is computationally intensive. Analogous to the methodologies employed in the original MultiVERSE and VERSE algorithms, Noise Contrastive Estimation (NCE) is utilized to approximate these computations [16].

To sum up, in this framework, the similarity in the multiplex-heterogenous network is computed using MultiXrank [11]. And MultiXVERSE applies Kullback-Leibler minimization to optimize the embeddings. The parameters for random walks with restart on the multilayer networks are the following: $r = 0.7$, $\eta = \left[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right]$, $\lambda = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$, $\delta_1 = 0.5$, $\delta_2 = 0.5$, $\delta_3 = 0.5$. and $\tau = \left\{\begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}, \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix}, \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}\right\}$.

The reader can refer to the original article for more details on the method and particular implementation of the algorithm [8] and [11]. The code for MultiXVERSE can be found at https://github.com/LPioL/MultiXVERSE.

2.1.1 Datasets for the gene-drug-disease multilayer network

We used several different datasets to construct the multiplex-heterogeneous network, which we constructed to be composed of one human molecular multiplex network (3 layers), one drug multiplex network (4 layers) and one disease monoplex network. The multiplex networks are linked by 3 bipartite networks: drug-disease, gene-disease, and drug-target networks.

The multiplex networks are the following:

- **Human molecular multiplex network:** This network is a molecular network, extracted from [8], composed of 3 layers:

    1. A protein-protein interaction (PPI) layer which integrates 4 datasets: Hi-Union, APID (apid.dep.usal.es) (Level 2, human only), Lit-BM (http://www.interactome-atlas.org/download).

2. The second layer is a pathways layer constructed from the human Reactome data [17] extracted from NDEx [18].
3. This last layer is a molecular complexes layer corresponding to the fusion of Hu.map [19] and Corum [20].

- **Drug multiplex network:** The multiplex drug network integrates several sources and interaction types and has been extracted from [11]. Data is derived from Cheng et al. [21] and the pharmacological drugs interaction network available at snap.stanford.edu were utilized. In this network, drugs are named according to DrugBank conventions, encompassing both the multiplex network and its associated bipartite networks:

  1. The first layer corresponds to clinical drug interactions. It includes 14,822 clinically reported adverse drug-drug interactions among 667 drugs.
  2. The second is the experimental drug combinations layer. It contains 737 experimentally validated drug combinations involving 376 drugs.
  3. The third represents the predicted drug combinations and includes 2,080 network-predicted combinations for hypertensive drugs, covering 65 different drugs.
  4. The last layer of the drug multiplex network includes the pharmacologic drug-drug interactions and consists of 48,514 interactions determined by the pharmacological effects of one drug on another, involving 1,514 drugs.

- **Disease network:** The Disease multiplex network (DIS) has been structured into two layers, each representing different aspects of disease relationships:

  1. Disease-Disease Network Based on Shared Symptoms: Originating from a bipartite disease-symptoms network, this layer forms connections based on the cosine distance between diseases, retaining all interactions where this distance is above 0.5, indicating significant symptom overlap.
  2. Comorbidity Network: This layer integrates epidemiological data to illustrate the comorbidity relationships among diseases, highlighting epidemiological correlations.

  Each layer provides a unique perspective on disease interactions, encompassing treatment similarities, symptom relationships, and epidemiological data. .

The bipartite networks are:

- **Gene-Disease Network:** We extracted the curated gene-disease bipartite network from the DisGeNET database in order to connect the two molecular and disease multiplex networks.

- **Drug-target Network:** This network combines drug-target associations from multiple sources including DrugBank Release Version 5.1.8 (accessible at https://go.drugbank.com/releases/latest), DrugCentral release v10.12 (available at https://drugcentral.org/download), and associations described by Cheng et al [21].

- **Disease-Drug Network:** The associations between diseases and drugs are obtained from the Comparative Toxicogenomics Database.

2.1.2 Evaluation of the approach

Link prediction Like previous studies [8, 12, 13, 22], we employed link prediction to assess the efficacy of our embeddings and validate our universal multilayer network embedding method for network biology and medicine. Our link prediction methodology entails initially removing 30% of bipartite edges randomly in each bipartite networks to form a training multilayer network. Subsequently, we employ a Random Forest classifier to this training network, as described in [8], and perform evaluations on a withheld subset consisting of 30% of the edges. The binary classifier's training method includes the utilization of various operators on the node embeddings. These operators comprise Hadamard, Weighted-L1, Weighted-L2, Average, and cosine.

The objective of this validation method is to ascertain the quality of the embeddings in the discovery of novel drug-gene-disease associations. At present, conducting direct comparative analyses with alternative methodologies is not practicable due to the unique nature of the embedding process for multilayer networks with three distinct node types from 3 differents multiplex networks, a feature not yet paralleled in the existing literature.

Case study on cancer and neurotransmitters The second approach we used for validation is to test the method on a case study, here the to assess the link between neurotransmitter and cancer [23, 24, 25]. Serotonin has already been linked to cancer [23, 24, 25] and we want to know if other neurotransmitters may be predicted by our model which could lead to the discovery of new targets and drug repositioning for cancer.

In order to test our system on these new results, we focused on biological modules. Once MultiXVERSE has been applied to the drug-disease-gene multilayer network, we used a clustering method on the embeddings and analyzed the clusters. The clustering method we applied is spherical k-means [26] with $k = 500$ applied on the embedding.

We then applied link prediction to GABA agonist drugs using a Random Forest classifier with the operator 'Average'.

2.2 Experimental testing:materials and method

2.2.1 Animal husbandry

Animal care was done in compliance with, and approval from, the Institutional Animal Care and Use Committee (IACUC) under protocol number M2023-18 of Tufts University. Xenopus embryos were collected according to standard protocols [27] in 0.1X MMR ((Marc's Modified Ringers) pH 7.8 + 0.1% Gentamicin. Xenopus embryos were staged according to [28].

2.2.2 Drug exposure

Stocks of muscimol (Tocris 0289) were kept at 10mM concentration in DMSO. Embryos were exposed in 0.1X MMR during stages 12-43 in muscimol at a final concentration of 50 μM.

2.2.3 Histology

Embryos at stage 43-45 were embedded in JB4 according to the manufacturer's directions (Polysciences), and sectioned on a Leica microtome at 20μ. They were then photographed on a Nikon SMZ-1500 microscope.

	ROC-AUC			
Operators	Gene-disease	Drug-target	Drug-disease	Average
Hadamard	0.88 ± 0.002	$\mathbf{0.92 \pm 0.001}$	0.88 ± 0.003	$\mathbf{0.90 \pm 0.005}$
Weighted_L1	0.88 ± 0.002	0.62 ± 0.01	0.77 ± 0.002	0.76 ± 0.005
Weighted_L2	0.88 ± 0.003	0.63 ± 0.01	0.76 ± 0.001	0.76 ± 0.004
Average	$\mathbf{0.94 \pm 0.002}$	$\mathbf{0.93 \pm 0.003}$	$\mathbf{0.91 \pm 0.002}$	$\mathbf{0.93 \pm 0.002}$
Cosine	0.55 ± 0.005	0.83 ± 0.004	0.70 ± 0.002	0.70 ± 0.004

Table 1: ROC-AUC scores for link prediction using MultiXVERSE. Link predictions are computed for the bipartite interactions of the multiplex-heterogeneous networks. The scores higher than 0.9 are highlighted in bold.

3 Results

3.1 Computational results

3.1.1 Evaluation results using link prediction

The ROC-AUC is superior to 0.9 with Average operators for all bipartite networks (see Table 3.1.1), meaning that the method can predict with high precision the removed 30% of gene-disease, drug-disease and drug-target links from the corresponding multiplex-heterogeneous networks.

The variance across all operators is minimal, indicating that the network embedding method demonstrates high robustness and consistency in each iteration of the link prediction evaluation test.

3.1.2 Consistency of MultiXVERSE with MultiVERSE results on the progeria cluster

To assess the quality of the clustering of our embeddings, we analyzed the progeria cluster similarly to [8] (see Figure 3A). Hutchinson-Gilford Progeria Syndrome (HGPS) is a rare genetic disorder that causes premature aging. It is characterized by symptoms such as slowed postnatal growth, facial structural abnormalities, premature cardiovascular diseases, lipodystrophy, hair loss, and widespread osteodysplasia. HGPS arises from mutations in the LMNA genes, leading to the production of a deleterious version of the Lamin A protein, known as Progerin.

The results are similar to those obtained with MultiVERSE: LMNA and HGPS were both found to be associated. We found several genes in both cluster including ZMPSTE24 in the cluster that is associated to accelerated aging in the literature and LMNA [29, 30], but also LEMD2, RGS18, MARVELD1, KCNK13, IZUMO2, PERM1, LINC01857, and KIF12n. The cluster share diseases to including muscular dystrophy, the Werner syndrome (the adult premature aging syndrome), deformities of the hand and foot and cardiac disease associated to progeria [31].

KCNK13 is an especially interesting gene and encodes a K^+ potassium ion channel (Potassium Two Pore Domain Channel Subfamily K Member 13). It is related to the Birk-Barel syndrome (BIBARS) - a rare genetic disorder characterized by motor and speech delay, impaired intellectual development, early feeding difficulties, muscular hypotonia, hyperactivity, aggression, and facial dysmorphism. This syndrome shares part of its phenotype with HGPS. HGPS has also been related to bioelectricity [32] which can fall in the context of aging (or premature aging) as a channelopathy [33].

Therefore, we conclude that we have similar results to the previous version of MultiVERSE [8], even if we have more multiplex and bipartite networks (MultiVERSE has been applied on a gene-disease mutilayer network) and a different set of networks, showing a good robustness to

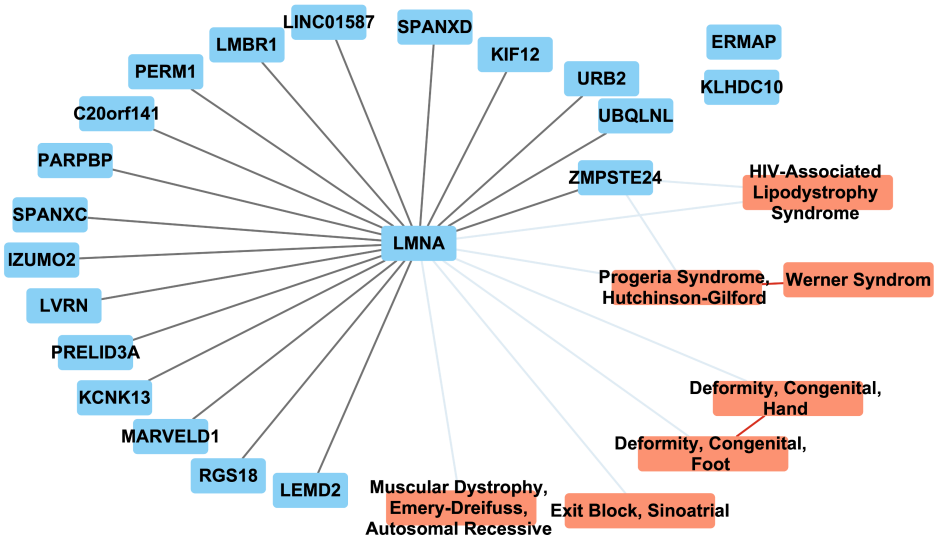

Figure 2: Network representation of the progeria cluster. Blue, orange are respectively genes, diseases. Black, light blue, and red links are respectively molecular multiplex, gene-disease, and disease multiplex links.

the integration of new data.

3.1.3 The clustering of the embeddings shows serotonin and GABA pathways linked to cancer and developmental disorder

In order to learn more about the link between cancer, developmental disorders and neurotransmitters [23, 24, 25], we analyzed the different clusters integrating those three components. We found that several clusters show a link between neurotransmitters including GABA and serotonin with cancer of malformations. One cluster (see Figure 3B) includes EPO and Darbepoetin alpha. Recombinant human erythropoietin is commonly used in clinical settings to treat anemia associated with cancer and chemotherapy. However, recent clinical trials indicate that rhEPO might also negatively affect disease progression and patient survival [34]. Interestingly, EPO is known to increase GABA currents [35] suggesting an implication of GABA neurotransmitter in the adverse effect of EPO in cancer development.

A second cluster is linking developmental disorders, neurotransmitter drugs (gabapentin with lamotrigine) with serotonin syndrome and large cell carcinomas (see Figure 3C). Gabapentin is a structural analogue of the inhibitory neurotransmitter gamma-aminobutyric acid (GABA) [36]. Lamotrigine is an anti-glutamate agent and may enhance GABAergic transmission [37]. Lamotrigine can also augment serotonin re-uptake inhibitors [38]. This cluster suggests a link between GABA, serotonin and cancer and that has been recently studied [23, 24, 25].

Lastly, we found a cluster (see Figure 3D) including developmental disorders Trimetazidine that is an anti-ischemic drug that can inhibit platelet aggregation and regulate the expression of serotonin in a rodent model [39].

These results indicate GABA pathways as potentially implicated in cancer development in addition of serotonin pathways [23, 24, 25].

Our model predicted links between cancer and serotonin and GABA via clustering, and the

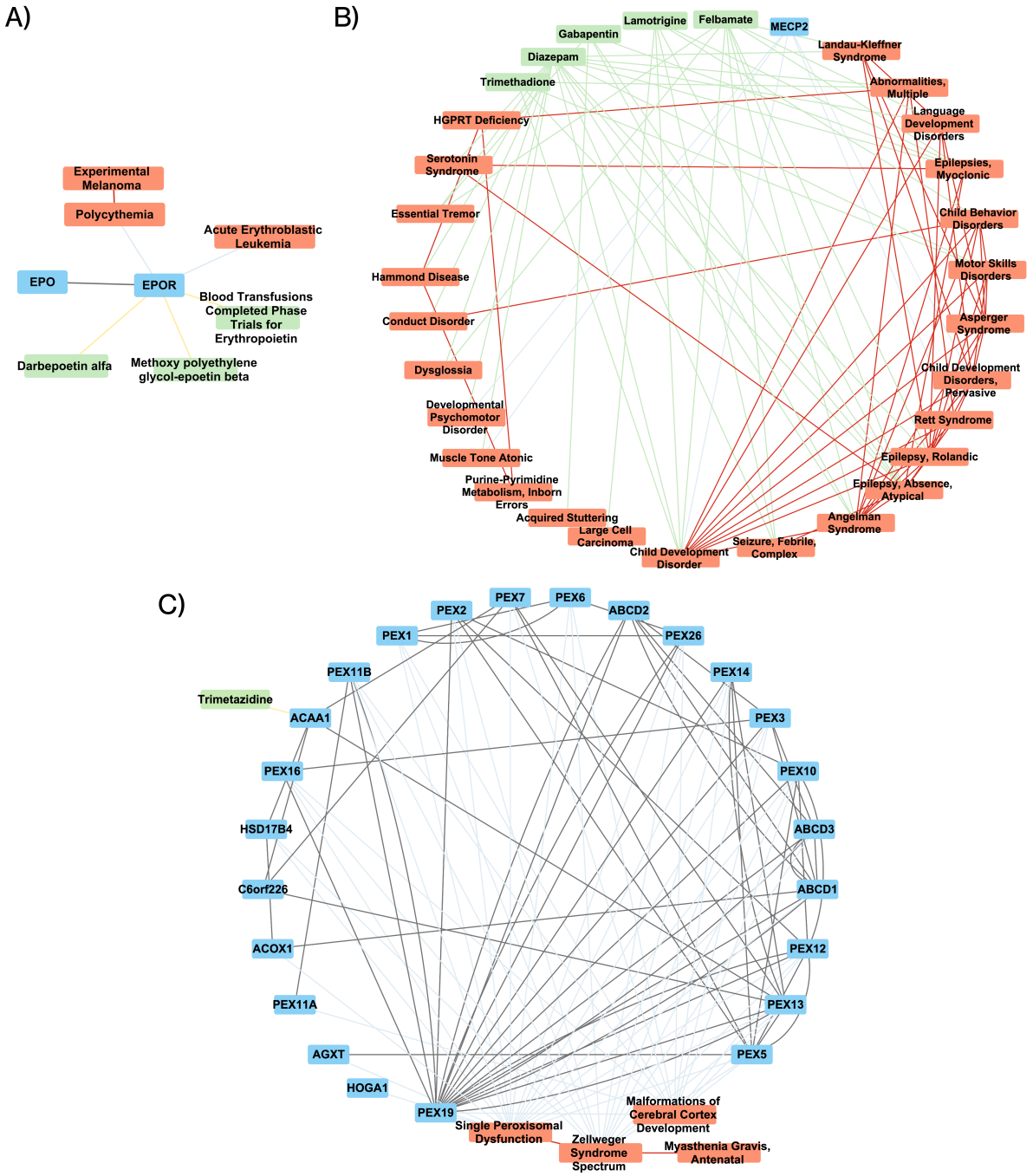

Figure 3: Network representations of the clusters integrating cancer or developmental disorders and neurotransmitters. A) Network cluster linking drugs regulating GABA and melanoma B) Network cluster linking GABA drugs like gabapentin and large cell carcinoma. C) Network cluster linking trimetazidine and malformations of cortex development. Blue, orange and green boxes are respectively genes, diseases and drugs. Black, light blue, yellow, light green and red links are respectively molecular multiplex, gene-disease, drug-target, drug-disease, disease multiplex links.

serotonin link has been validated by published data [23, 24, 25]. We decided to test the link between GABA and cancer using link prediction.

3.1.4 GABA drugs show different types of cancer in the first 10 predictions

In order to validate the link between GABA and cancer, we applied link prediction using the embeddings on Baclofen, Zaleplon, Clobazam, Progabide, Zolpidem and Gabapentin. These drugs are GABA agonists [40, 41, 42, 43, 44].

All of the tested drugs showed a link with cancer in the 10 first predictions, except with Zolpidem that showed a link with rectum neoplasm at the 18th prediction. Baclofen had a prediction for adrenal cortical carcinoma, Zaplelon for cancer of the esophagus and adrenal cortical carcinoma, Clobazam for soft tissue sarcoma and experimented neoplasm, Progabide for soft tissue neoplasm, and Gabapentin for soft tissue neoplasm.

The new prediction of GABA agonist as a potential trigger of cancer had not yet been validated; thus, we sought to test it . experimentally.

3.2 Experimental results

In order to test the prediction that GABA pathway modulation should induce a cancer-like phenotype, we used larval Xenopus laevis - a powerful model system commonly used to understand carcinogenic dysregulation of cell function [45, 46, 47]. One type of cancer that is especially readily investigated in frog embryos is the conversion of melanocytes, from normal pigment cells to a hyper-proliferative, invasive melanoma-like phenotype [23, 47, 25, 48]. In order to perturb GABA signaling, we used muscimol – a well-known GABA agonist [49, 50, 51, 52, 53]. Muscimol itself is not in the original data we used in our model but it is a GABA(A) agonist like Progabide [54, 55] that was found in the link prediction (see above), enabling us to test the utility of the model's categorical predictions for novel drugs that it did not have direct experience with. Fifty Xenopus embryos were exposed to 50 mM muscimol between the stage 12 to st. 45 (after completion of gastrulation through swimming tadpole stages), and then sectioned. The results are shown in Figure 4. In contrast to controls, all of the exposed animals exhibited a drastic hyperpigmentation due to the melanocytes' changing shape (from their normal round form to a much more elongated morphology), and migrating into inappropriate regions that are normally clear. This phenotype has previously been characterized quantitatively with respect to melanocyte shape and number, as well as the expression of cancer-related markers and melanoma-like migration into gut, brain, and vascular tissues [23, 47, 25, 24, 48]. These results confirm the prediction of the model and link GABA-modulating drugs to a cancer-like phenotype in the absence of classic carcinogens, oncogenes, or DNA damage as the initiator of cellular conversion.

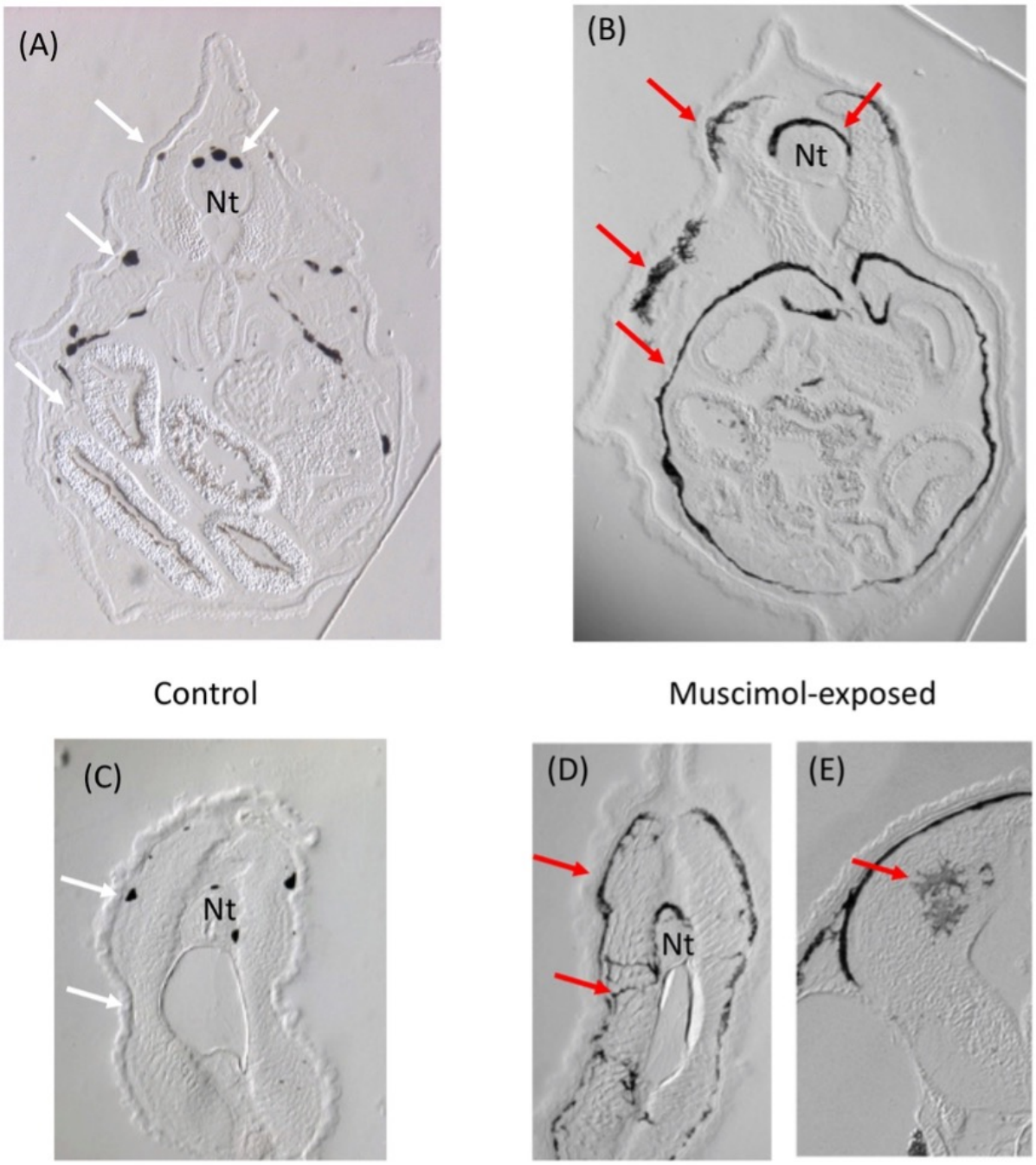

Figure 4: Melanocyte conversion phenotype induced by muscimol exposure. All panels are transverse sections; A,B are taken through the gut, while C,D are taken through the mid-tail. (A) Control embryos show small numbers of discrete, round melanocytes (white arrows). (B) In contrast, muscimol exposure induces melanocytes to over-proliferate and form long, stretched out projections that cover the gut cavity and other locations (red arrows). Compared to round melanocytes in the trunk and tail of controls (C), melanocytes in muscimol-treated animals (D) can be clearly seen to have an abnormal invasive shape and distribution (D). (E) Closeup of melanocytes invading the neural tube after muscimol exposure. Nt = neural tube.

4 Discussion

In this work, we presented what is, to the best of our knowledge the first universal multiplayer network embedding method with no limitations on number of multiplex and bipartite networks thanks to recent developments in RWR [11]. We applied MultiXVERSE to a multilayer network containing gene, drug, and diseases interactions and evaluated the quality of the embedding using both clustering and link prediction to demonstrate the quality of the embeddings. Our model predicted links between cancer, serotonin, and GABA via clustering. The serotonin link has been validated by published data [23, 24, 25]. Next, we tested the new prediction from clustering between GABA and cancer using link prediction and found again predictions between GABA agonists and cancer. Finally, we tested this new prediction experimentally and confirmed the ability of GABA-modulating drugs to induce a cancer-like phenotype in a vertebrate model system in vivo, in the absence of classic carcinogens, oncogenes, or DNA damage as the initiator of cellular conversion. These kinds of data have important implications for understanding of cancer etiology and possible normalization [56, 57, 58] efforts using neurotransmitter modulators, and need to be tested in preclinical mammalian models next.

Our main domain of application here was drug discovery for cancer but the embeddings of a gene-drug-disease multilayer networks may be used for many other applications. Different questions in network biology could be addressed, by finding new genes related to specific diseases, such as the interesting target – the ion channel KCNK13 - revealed by our case study on the progeria cluster. This may be especially relevant due to the recent hypotheses about the role of bioelectric signaling in aging [33, 59, 60]. We also applied link prediction for different drug-disease associations. Our method for drug repurposing has currently one significant limitation: in the case of new drug-target predictions, we don't know if the drug will activate or inactivate the target. This may be resolved by using directed networks in the multiplex networks and training link prediction models on the embedding integrating the directional information of the links.

Our system revealed predicted a causal link between GABA and cancer development that we validated experimentally in a vertebrate model system in vivo. This link has been studied before and it has been found that GABA has a driver role in controlling stem and proliferative cell state through GHB production in glioma [61]. It has been reported too that membrane potential differences and GABA(A) receptor expression in hepatic tumor versus non-tumor stem cells [62]. However, contradictory evidences has been reported showing that GABA could have an inhibitory effect on tumor progression or cell proliferation [63, 64]. Consequently, more research is necessary to understand the impact of GABA on cancer.

To improve the capabilities of Large Language Models (LLMs) in processing text-enriched images, researchers have developed embeddings specifically designed to capture image contexts. These embeddings are integrated as soft prompt inputs in LLMs, enhancing the models' ability to effectively handle visual information [65]. Similarly, we could use the node embeddings as input to use the power of LLMs for generative drug discovery including the richness of multiplex-heterogeneous network biological data. One relevant effort that can be integrated into this framework in the future is the bioinformatics of shape, which seeks to formalize and make amenable to machine learning data on large-scale anatomical outcomes in embryogenesis, regeneration, cancer, and bioengineering [66, 67]. We expect that future systems that combine pattern inference with diverse multi-modal datasets, comprising physiological, anatomical, and molecular-biological data will be a critical aid to human scientists and clinicians seeking to develop interventions for a wide range of biomedical applications.

Acknowledgements

We thank Julia Poirier for assistance in the manuscript, Axel De Baat for helpful comments, Dayong Qiu for histology, and Maria Lobikin for early work on muscimol. We gratefully acknowledge support via a sponsored research agreement with Astonishing Labs.

Conflict of Interest statement

M.L.'s lab receives funding in the form of sponsored research agreements from Astonishing Labs and Morphoceuticals, companies that operates in the space of drug discovery and biomedicine.

References

[1] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, "Multilayer networks," *Journal of Complex Networks*, vol. 2, no. 3, pp. 203–271, 2014.

[2] M. Zitnik, M. M. Li, A. Wells, K. Glass, D. M. Gysi, A. Krishnan, T. Murali, P. Radivojac, S. Roy, A. Baudot, *et al.*, "Current and future directions in network biology," *arXiv preprint arXiv:2309.08478*, 2023.

[3] L. Liao, X. He, H. Zhang, and T.-S. Chua, "Attributed social network embedding," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 12, pp. 2257–2270, 2018.

[4] G. Ma, C.-T. Lu, L. He, S. Y. Philip, and A. B. Ragin, "Multi-view graph embedding with hub detection for brain network analysis," in *2017 IEEE International Conference on Data Mining (ICDM)*, pp. 967–972, IEEE, 2017.

[5] P. Csermely, N. Kunsic, P. Mendik, M. Kerestély, T. Faragó, D. V. Veres, and P. Tompa, "Learning of signaling networks: molecular mechanisms," *Trends in biochemical sciences*, vol. 45, no. 4, pp. 284–294, 2020.

[6] I. A. Kovács, R. Mizsei, and P. Csermely, "A unified data representation theory for network visualization, ordering and coarse-graining," *Scientific reports*, vol. 5, no. 1, p. 13786, 2015.

[7] W. Nelson, M. Zitnik, B. Wang, J. Leskovec, A. Goldenberg, and R. Sharan, "To embed or not: network embedding as a paradigm in computational biology," *Frontiers in genetics*, vol. 10, 2019.

[8] L. Pio-Lopez, A. Valdeolivas, L. Tichit, É. Remy, and A. Baudot, "Multiverse: a multiplex and multiplex-heterogeneous network embedding approach," *Scientific Reports*, vol. 11, no. 1, pp. 1–20, 2021.

[9] A. Valdeolivas, L. Tichit, C. Navarro, S. Perrin, G. Odelin, N. Levy, P. Cau, E. Remy, and A. Baudot, "Random walk with restart on multiplex and heterogeneous biological networks," *Bioinformatics*, vol. 35, no. 3, pp. 497–505, 2018.

[10] A. Tsitsulin, D. Mottin, P. Karras, and E. Müller, "Verse: Versatile graph embeddings from similarity measures," in *Proceedings of the 2018 World Wide Web Conference*, pp. 539–548, International World Wide Web Conferences Steering Committee, 2018.

[11] A. Baptista, A. Gonzalez, and A. Baudot, "Universal multilayer network exploration by random walk with restart," *Communications Physics*, vol. 5, no. 1, p. 170, 2022.

[12] H. Zhang, L. Qiu, L. Yi, and Y. Song, "Scalable multiplex network embedding.," in *IJCAI*, vol. 18, pp. 3082–3088, 2018.

[13] A. Bagavathi and S. Krishnan, "Multi-net: A scalable multiplex network embedding framework," in *International Conference on Complex Networks and their Applications*, pp. 119–131, Springer, 2018.

[14] S.-H. Jiang, L.-P. Hu, X. Wang, J. Li, and Z.-G. Zhang, "Neurotransmitters: emerging targets in cancer," *Oncogene*, vol. 39, no. 3, pp. 503–515, 2020.

[15] R. Mancusi and M. Monje, "The neuroscience of cancer," *Nature*, vol. 618, no. 7965, pp. 467–479, 2023.

[16] M. Gutmann and A. Hyvärinen, "Noise-contrastive estimation: A new estimation principle for unnormalized statistical models," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pp. 297–304, 2010.

[17] D. Croft, A. F. Mundo, R. Haw, M. Milacic, J. Weiser, G. Wu, M. Caudy, P. Garapati, M. Gillespie, M. R. Kamdar, *et al.*, "The reactome pathway knowledgebase," *Nucleic acids research*, vol. 42, no. D1, pp. D472–D477, 2014.

[18] D. Pratt, J. Chen, D. Welker, R. Rivas, R. Pillich, V. Rynkov, K. Ono, C. Miello, L. Hicks, S. Szalma, *et al.*, "Ndex, the network data exchange," *Cell systems*, vol. 1, no. 4, pp. 302–305, 2015.

[19] K. Drew, C. Lee, R. L. Huizar, F. Tu, B. Borgeson, C. D. McWhite, Y. Ma, J. B. Wallingford, and E. M. Marcotte, "Integration of over 9,000 mass spectrometry experiments builds a global map of human protein complexes," *Molecular systems biology*, vol. 13, no. 6, 2017.

[20] M. Giurgiu, J. Reinhard, B. Brauner, I. Dunger-Kaltenbach, G. Fobo, G. Frishman, C. Montrone, and A. Ruepp, "Corum: the comprehensive resource of mammalian protein complexes—2019," *Nucleic acids research*, vol. 47, no. D1, pp. D559–D563, 2019.

[21] F. Cheng, I. A. Kovács, and A.-L. Barabási, "Network-based prediction of drug combinations," *Nature communications*, vol. 10, no. 1, pp. 1–11, 2019.

[22] L. Pio-Lopez, "Drug Repositioning Using Multiplex-Heterogeneous Network Embedding: A Case Study on SARS-CoV2," in *Complex Networks & Their Applications X* (R. M. Benito, C. Cherifi, H. Cherifi, E. Moro, L. M. Rocha, and M. Sales-Pardo, eds.), (Cham), pp. 731–741, Springer International Publishing, 2022.

[23] D. Blackiston, D. S. Adams, J. M. Lemire, M. Lobikin, and M. Levin, "Transmembrane potential of glycl-expressing instructor cells induces a neoplastic-like conversion of melanocytes via a serotonergic pathway," *Disease models & mechanisms*, vol. 4, no. 1, pp. 67–85, 2011.

[24] D. Lobo, M. Lobikin, and M. Levin, "Discovering novel phenotypes with automatically inferred dynamic models: a partial melanocyte conversion in xenopus," *Scientific Reports*, vol. 7, no. 1, p. 41339, 2017.

[25] M. Lobikin, D. Lobo, D. J. Blackiston, C. J. Martyniuk, E. Tkachenko, and M. Levin, "Serotonergic regulation of melanocyte conversion: A bioelectrically regulated network for stochastic all-or-none hyperpigmentation," *Science signaling*, vol. 8, no. 397, pp. ra99–ra99, 2015.

[26] C. Buchta, M. Kober, I. Feinerer, and K. Hornik, "Spherical k-means clustering," *Journal of Statistical Software*, vol. 50, no. 10, pp. 1–22, 2012.

[27] G. R. Sive, Hazel L and R. Harland, "Early development of xenopus laevis: a laboratory manual," p. 249, 2000.

[28] P. D. Nieuwkoop and J. Faber, *Normal table of Xenopus laevis (Daudin). A systematical and chronological survey of the development from the fertilized egg till the end of metamorphosis.* Amsterdam: North-Holland Publishing Company, 1967.

[29] I. Varela, J. Cadinanos, A. M. Pendás, A. Gutiérrez-Fernández, A. R. Folgueras, L. M. Sánchez, Z. Zhou, F. J. Rodríguez, C. L. Stewart, J. A. Vega, *et al.*, "Accelerated ageing in mice deficient in zmpste24 protease is linked to p53 signalling activation," *Nature*, vol. 437, no. 7058, pp. 564–568, 2005.

[30] H. J. Worman and S. Michaelis, "Prelamin a and zmpste24 in premature and physiological aging," *Nucleus*, vol. 14, no. 1, p. 2270345, 2023.

[31] J. P. van Tintelen, R. M. Hofstra, H. Katerberg, T. Rossenbacker, A. C. Wiesfeld, G. J. du Marchie Sarvaas, A. A. Wilde, I. M. van Langen, E. A. Nannenberg, A. J. van der Kooi, *et al.*, "High yield of lmna mutations in patients with dilated cardiomyopathy and/or conduction disease referred to cardiogenetics outpatient clinics," *American heart journal*, vol. 154, no. 6, pp. 1130–1139, 2007.

[32] C.-Y. Lo, Y.-W. Tjong, J. C.-Y. Ho, C.-W. Siu, S.-Y. Cheung, N. L. Tang, S. Yu, H.-F. Tse, and X. Yao, "An upregulation in the expression of vanilloid transient potential channels 2 enhances hypotonicity-induced cytosolic ca2+ rise in human induced pluripotent stem cell model of hutchinson gillford progeria," *PLoS One*, vol. 9, no. 1, p. e87273, 2014.

[33] L. Pio-Lopez and M. Levin, "Aging as a loss of morphostatic information: a developmental bioelectricity perspective," *Ageing Research Reviews*, p. 102310, 2024.

[34] J. Szenajch, G. Wcislo, J.-Y. Jeong, C. Szczylik, and L. Feldman, "The role of erythropoietin and its receptor in growth, survival and therapeutic response of human tumor cells: from clinic to bench—a critical review," *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, vol. 1806, no. 1, pp. 82–95, 2010.

[35] C. Roseti, P. Cifelli, G. Ruffolo, E. Barbieri, M. Guescini, V. Esposito, G. Di Gennaro, C. Limatola, A. Giovannelli, E. Aronica, *et al.*, "Erythropoietin increases gabaa currents in human cortex from tle patients," *Neuroscience*, vol. 439, pp. 153–162, 2020.

[36] M. Rose and P. Kam, "Gabapentin: pharmacology and its use in pain management," *Anaesthesia*, vol. 57, no. 5, pp. 451–462, 2002.

[37] B. Costa and N. Vale, "Understanding lamotrigine's role in the cns and possible future evolution," *International Journal of Molecular Sciences*, vol. 24, no. 7, p. 6050, 2023.

[38] J. G. Reid, M. J. Gitlin, and L. L. Altshuler, "Lamotrigine in psychiatric disorders," *J Clin Psychiatry*, vol. 74, no. 7, pp. 675–84, 2013.

[39] M. Liu, W. Wei, C. R. Stone, L. Zhang, G. Tian, and J. N. Ding, "Beneficial effects of trimetazidine on expression of serotonin and serotonin transporter in rats with myocardial infarction and depression," *Neuropsychiatric Disease and Treatment*, pp. 787–797, 2018.

[40] A. Finnimore, M. Roebuck, D. Sajkov, and R. McEvoy, "The effects of the gaba agonist, baclofen, on sleep and breathing," *European Respiratory Journal*, vol. 8, no. 2, pp. 230–234, 1995.

[41] H. Noguchi, K. Kitazumi, M. Mori, and T. Shiba, "Binding and neuropharmacological profile of zaleplon, a novel nonbenzodiazepine sedative/hypnotic," *European journal of pharmacology*, vol. 434, no. 1-2, pp. 21–28, 2002.

[42] R. Huddart, J. S. Leeder, R. B. Altman, and T. E. Klein, "Pharmgkb summary: clobazam pathway, pharmacokinetics," *Pharmacogenetics and genomics*, vol. 28, no. 4, pp. 110–115, 2018.

[43] K. J. Bergmann, "Progabide: a new gaba-mimetic agent in clinical use," *Clinical neuropharmacology*, vol. 8, no. 1, pp. 13–26, 1985.

[44] E. Sanna, F. Busonero, G. Talani, M. Carta, F. Massa, M. Peis, E. Maciocco, and G. Biggio, "Comparison of the effects of zaleplon, zolpidem, and triazolam at various gabaa receptor subtypes," *European journal of pharmacology*, vol. 451, no. 2, pp. 103–110, 2002.

[45] L. J. Hardwick and A. Philpott, "An oncologist's friend: How xenopus contributes to cancer research," *Developmental biology*, vol. 408, no. 2, pp. 180–187, 2015.

[46] L. J. Hardwick and A. Philpott, "Xenopus models of cancer: expanding the oncologist's toolbox," *Frontiers in Physiology*, vol. 9, p. 424568, 2018.

[47] M. Lobikin, B. Chernet, D. Lobo, and M. Levin, "Resting potential, oncogene-induced tumorigenesis, and metastasis: the bioelectric basis of cancer in vivo," *Physical biology*, vol. 9, no. 6, p. 065002, 2012.

[48] J. Morokuma, D. Blackiston, D. S. Adams, G. Seebohm, B. Trimmer, and M. Levin, "Modulation of potassium channel function confers a hyperproliferative invasive phenotype on embryonic stem cells," *Proceedings of the National Academy of Sciences*, vol. 105, no. 43, pp. 16608–16613, 2008.

[49] A. Y. Benkherouf, K.-R. Taina, P. Meera, A. J. Aalto, X.-G. Li, S. L. Soini, M. Wallner, and M. Uusi-Oukari, "Extrasynaptic δ-gabaa receptors are high-affinity muscimol receptors," *Journal of neurochemistry*, vol. 149, no. 1, pp. 41–53, 2019.

[50] G. A. Johnston, "Muscimol as an ionotropic gaba receptor agonist," *Neurochemical research*, vol. 39, pp. 1942–1947, 2014.

[51] M. J. Leach and J. A. Wilson, "Gaba receptor binding with 3h-muscimol in calf cerebellum.," *European journal of pharmacology*, vol. 48, no. 3, pp. 329–330, 1978.

[52] J. Scheel-Krüger, A. R. Cools, and P. M. Van Wel, "Muscimol a gaba-agonist injected into the nucleus accumbens increases apomorphine stereotypy and decreases the motility," *Life Sciences*, vol. 21, no. 11, pp. 1697–1702, 1977.

[53] I. Shoulson, D. Goldblatt, M. Charlton, and R. J. Joynt, "Huntington's disease: treatment with muscimol, a gaba-mimetic drug," *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, vol. 4, no. 3, pp. 279–284, 1978.

[54] G. Bartholini, "Pharmacology of the gabaergic system: effects of progabide, a gaba receptor agonist," *Psychoneuroendocrinology*, vol. 9, no. 2, pp. 135–140, 1984.

[55] A. Wahab, U. Heinemann, and K. Albus, "Effects of γ-aminobutyric acid (gaba) agonists and a gaba uptake inhibitor on pharmacoresistant seizure like events in organotypic hippocampal slice cultures," *Epilepsy research*, vol. 86, no. 2-3, pp. 113–123, 2009.

[56] M. V. Maffini, J. M. Calabro, A. M. Soto, and C. Sonnenschein, "Stromal regulation of neoplastic development: age-dependent normalization of neoplastic mammary cells by mammary stroma," *The American journal of pathology*, vol. 167, no. 5, pp. 1405–1410, 2005.

[57] J. C. Kasemeier-Kulesa, J. M. Teddy, L.-M. Postovit, E. A. Seftor, R. E. Seftor, M. J. Hendrix, and P. M. Kulesa, "Reprogramming multipotent tumor cells with the embryonic neural crest microenvironment," *Developmental Dynamics*, vol. 237, no. 10, pp. 2657–2666, 2008.

[58] B. Mintz and K. Illmensee, "Normal genetically mosaic mice produced from malignant teratocarcinoma cells.," *Proceedings of the National Academy of Sciences*, vol. 72, no. 9, pp. 3585–3589, 1975.

[59] B. Anderson, "Bioelectricity: A top-down control model to promote more effective aging interventions," *Bioelectricity*, vol. 6, no. 1, pp. 2–12, 2024.

[60] B. B. Silver and C. M. Nelson, "The bioelectric code: reprogramming cancer and aging from the interface of mechanical and chemical microenvironments," *Frontiers in Cell and Developmental Biology*, vol. 6, p. 21, 2018.

[61] E. A. El-Habr, L. G. Dubois, F. Burel-Vandenbos, A. Bogeas, J. Lipecka, L. Turchi, F.-X. Lejeune, P. L. C. Coehlo, T. Yamaki, B. M. Wittmann, *et al.*, "A driver role for gaba metabolism in controlling stem and proliferative cell state through ghb production in glioma," *Acta neuropathologica*, vol. 133, pp. 645–660, 2017.

[62] W. Bautista, V. Perez-Alvarez, F. Burczynski, A. Raouf, T. Klonisch, and G. Minuk, "Membrane potential differences and gabaa receptor expression in hepatic tumor and non-tumor stem cells," *Canadian Journal of Physiology and Pharmacology*, vol. 92, no. 1, pp. 85–91, 2014.

[63] H. M. Schuller, H. A. Al-Wadei, and M. Majidi, "Gamma-aminobutyric acid, a potential tumor suppressor for small airway-derived lung adenocarcinoma," *Carcinogenesis*, vol. 29, no. 10, pp. 1979–1985, 2008.

[64] A. Ortega, "A new role for gaba: inhibition of tumor cell migration," *Trends in pharmacological sciences*, vol. 24, no. 4, pp. 151–154, 2003.

[65] W. Hu, Y. Xu, Y. Li, W. Li, Z. Chen, and Z. Tu, "Bliva: A simple multimodal llm for better handling of text-rich visual questions," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, pp. 2256–2264, 2024.

[66] D. Lobo, T. J. Malone, and M. Levin, "Towards a bioinformatics of patterning: a computational approach to understanding regulative morphogenesis," *Biology Open*, vol. 2, no. 2, pp. 156–169, 2012.

[67] D. Lobo, E. B. Feldman, M. Shah, T. J. Malone, and M. Levin, "A bioinformatics expert system linking functional data to anatomical outcomes in limb regeneration," *Regeneration*, vol. 1, no. 2, pp. 37–56, 2014.