A COMPLEX SYSTEMS ANALYSIS OF CANCER NETWORKS

Uthamacumaran, A.¹

a_utham@live.concordia.ca

Concordia University, Montreal

ABSTRACT

Cancers are complex, adaptive ecosystems. They remain the lead cause of disease-related, pediatric death in North America. The emerging field of complexity science has redefined cancer as a computational system with intractable, algorithmic complexity. Herein, a tumor and its heterogeneous phenotypes are discussed as dynamical systems having multiple, chaotic attractors. Machine learning, Network science and information theory are discussed as current tools for cancer network reconstruction. The fluid dynamics of cancer-cell fate transitions and chemical pattern formation are briefly reviewed. Deep Learning architectures, delay-embedding algorithms and computational fluid models are proposed for better forecasting gene expression patterns in cancer ecosystems. Cancer cell decision-making is investigated within the framework of complexity theory.
INTRODUCTION

Cancer is the leading cause of ‘disease’-related death among children in North America. The heterogeneous clonal expansion, replicative immortality, patterns of longevity, evasion of death signals, hijacked immune system, self-sufficient growth signals, metastatic invasion and other emergent properties of cancer indicate cancer is a complex adaptive system (Hanahan and Weinberg, 2011). A complex adaptive system is a nonlinear dynamical system of many interacting agents which adaptively respond to the perturbations of their environment. The signatures of a complex adaptive system include emergence, self-organized patterns, interconnected multi-level structures, sudden critical transitions (edge of chaos), computational irreducibility, unpredictability and multi-scaled, feedback loops. In simple terms, the collective whole cannot be defined by the sum of its interacting parts (Wolfram, 1988; 2002).

Waddington (1942) first described cellular development as an energy landscape where cell fate bifurcations are visualized as balls rolling up hills (energy barriers) and down valleys (attractors) (Bhattacharya et al., 2011; Wang et al., 2011). The cell fate decision-making is governed by the gene expression dynamics of underlying gene regulatory networks (GRNs) and their environment. Due to the stochastic nature of molecules, the changes in molecular concentrations are generally defined by Langevin equations (Davila-Vederrain, 2015). The transitions between the attractors are considered as random walks on a network, following Boltzmann-Gibbs statistics (Perkins et al., 2014). The time-evolution of a cell state’s probability density on the attractor space is given by the Fokker-Planck equations:

$$\frac{\partial \rho}{\partial t} = - \sum_i \frac{\partial}{\partial x_i} [A_i(x)\rho] + \frac{1}{2} \sum_{i,j} D_{ij}(x) \frac{\partial^2}{\partial x_i x_j} \rho$$

where the mean path is found from the Path integral formulation. Here, D is the diffusion matrix, A is the drift vector characterizing the drag and random forces encountered in Brownian motion, x the spatial coordinate in one dimension and \(\rho\) the probability (information) density. The mean path is determined from a path integral assuming a Gaussian kernel for the noise, as: \(\rho(x, t) = \int Dx \exp\{-S(x)\} = \int Dx \exp\{-\int L(x(t))dt\}\), where the sum over all possible paths contributes to the cell fate trajectory. Herein, S is the action and L is the Lagrangian or weight of each path such that the probability flux \(\frac{\partial \rho}{\partial t} + \nabla \cdot J = 0\) is conserved (Chunhe and Wang, 2013; Wang, 2014). Assuming steady-state solutions, the energy potential at a given point in the landscape is \(V(x) = -\ln \rho(x)\) and corresponds to a cell state’s energy density. However, certain gene expressions can switch the cell fate from a basin of attraction into another. Examples include the EMT/MET (epithelial-mesenchymal transition) switches observed in cancer pattern formation and induced pluripotent stem cells reprogramming (iPSC) by the Yamanaka transcription factors. Sudden bursts of transcription factors or morphogens’ expression result in state transitions.

Assume a cell state to be defined as a configuration of steady states \(x(t)\), represented by a gene expression matrix. At the critical transition state, the effects of a bifurcation (control) parameter \(\mu\) is given by, \(\dot{x}(t) = F(x(t), \mu)\). In complex, high-dimensional attractor spaces such as the Waddington landscape, the bifurcation parameters \(\mu\), are typically not known and used to denote qualitatively the
edge of criticality in state bifurcations (Mojtahedi et al., 2016). In fluid dynamics, the bifurcation parameter is often considered to be the Reynolds number (Ruelle, 1995). The attractors (energy minima) are generally assumed to be fixed points or periodic orbits. However, cancer cell fates are observed as unstable attractors fluctuating between valleys of the landscape, with differential gene expression patterns and spatiotemporal heterogeneity (Huang and Kauffman, 2013; Zhou et al., 2014; Li et al., 2016). Gene expression is often termed a stochastic process (Elowitz, 2002). However, in argument, cancer cells are strange attractors of the Waddington landscape, where the cell fate can fluctuate back and forth between valleys in an ‘apparent random’ motion (FIGURE 1). That is, by definition, they exhibit sensitive dependence to initial conditions (i.e. chaos) and their bounded phase space has a fractal-dimension. Unlike a random process, a strange attractor exhibits specific patterns that can be mapped in the peak analysis of its corresponding frequency spectra with time-delay embedding algorithms.

Spin glasses are statistical tools to model emergent behaviours in initially disordered, many-body systems (Mezard and Montanari, 2009). The NK model, a spin-glass based, rugged energy landscape is used in the computational modelling of many complex systems, including the combinatorial optimization of Boolean GRNs (Kauffman and Levin, 1987; Weinberger, 1991). Criticality in spin-glass phase transitions are reminiscent of the chaotic bifurcations observed in chemical pattern formation systems (Wolf et al., 2018). For e.g., EMT (epithelial-mesenchymal transition) states are reminiscent of glassy phase transitions (Font-Clos, 2018). Spin glass statistics are also the foundations of neural networks, powerful algorithms within complexity science emerging in Waddington landscape reconstruction.

Complex systems theory, or simply, complexity theory, emerged from dynamical systems theory in the 1960s. It surged in the 1980s when Wolfram demonstrated cellular automata, simple computer programs following simple, discrete rules, produced highly complex behaviours observed in Nature (Wolfram, 1984). While current approaches of cancer modelling assume relatively well-defined, reducible models with steady-state solutions, complexity science indicates reconstructing the cancer attractor space as an intractable, computer-science problem. Such problems currently do not have analytical solutions and hence, require approximations by computational heuristics and algorithms (Wolfram, 2002). The emergent paradigm in complexity science is two-fold: (a) the information dynamics of biological networks are representable as a set of strings or, rows and columns of an adjacency matrix (graph), and (b) algorithmic complexity theory provides approximate solutions and long-term predictions of otherwise, analytically intractable problems in biological networks reconstruction.

Thus, Complexity is related to NP-completeness. To illustrate NP-completeness, consider the following problem. Is searching necessary to find a needle in a haystack? This depends on our tools. If a magnet was available, the vast space of possibilities need not be explored to find the needle. This is an analogy of optimization. The P vs. NP problem asks, can we solve searching problems without searching? NP problems are quickly checkable but are either intractable or solved by brute-force searching. As such, there are three layers of complexity to consider when approaching a complexity problem: time, space and algorithmic complexity. The problem was first informally addressed in a letter by Gödel to von Neumann (1956).
Finding the clique or master GRNs (gene regulatory networks) controlling cancer stem cell (CSC) fate transitions is a combinatorially complex, optimization problem. Such complex optimization/decision problems are studied in a branch of Algorithmic Information Theory (AIT), known as Kolmogorov-Chaitin complexity, wherein causal trajectories in complex interactomes are predicted on the grounds of information dynamics and computations. AIT treats cancer as a computer program with unresolved, algorithmic complexity (Zenil et al., 2016).

Hence, machine learning algorithms and data science process the information flow from single-cell multi-omics to infer causal trajectories between the identified signaling pathways (Saelens et al., 2019; Yang et al., 2019). Information measures such as correlation scores and entropy are used to infer network dynamics from single-cell datasets (Chan et al., 2017). However, current approaches lack the time component in their analyses. In proposition, time-series scRNA-Seq and multi-omics will allow us to map cancer (stem) cells as dynamical systems and thereby, test for the emergence of strange attractors in cell fate decision making.

STEM CELL REPROGRAMMING

Reprogramming cell fates was an intractable problem until the pioneering works of Nobel Laureates Shinya Yamanaka and John Gurdon. Today, the Yamanaka factors can be replaced by a cocktail of small molecules with higher reprogramming efficiency (Hou et al., 2013). The chromatin remodelling by the chemical factors allows the facilitated binding of Yamanaka transcription factors (TFs) and overcome the epigenetic barriers to dedifferentiate cells to iPSC states. Micro-RNA-based reprogramming generated iPSC clones claimed with up to 90% efficiency in human fibroblasts (Kogut et al., 2018). There was > 200-fold increase in reprogramming efficiency when the culture media contained antagonists of TGF-β, MEK/ERK inhibitors (mitogen activated kinases) and thiazovinin (ROCK inhibitor) (Xiong et al., 2019; Saito et al., 2019). Many alternate algorithms have been discovered in systems biology for chemically altering the epigenetic landscape of differentiated cell states and thereby, minimizing their trajectories towards stem cell attractors (Rais et al., 2013; Ranquist et al., 2017; Hernandez et al. 2018). This illustrates the P vs. NP problem in complexity theory. That is, when one NP-problem is solved, a class of similar problems are solved. However, in the context of cancer stem cells, signal effects vary on the patient, tissue type and microenvironment, making reprogramming apparently difficult (O’Brien-Ball and Biddle, 2017).

The mere notion of ‘reprogramming’ indicates cells are computational systems (computers), whose cell fates correspond to specific chemical programs (Zenil et al., 2016). More specifically, cancer is a computer program in an infinite loop with no halting condition (Zenil et al., 2016). Algorithmic Information Theory was demonstrated as a tool for reverse engineering complex, biological networks. Moreover, unhealthy states of the dynamical immune system were defined as strange attractors, while healthy states were considered as fixed attractors in state space.

What is a cancer stem cell? This is an unresolved complexity problem. In highly fluid cancers such as leukemia, distinct cancer stem cells are identified. For instance, the 17-gene stemness score (LSC17 score) is a prognostic biomarker used for assessing acute myeloid leukemia (AML) relapse in clinical care (Shlush et al., 2014; Ng et al., 2016). However, in solid tumors, the question remains debated whether all cancer cells are (potentially) stem cells (Battle and Clevers, 2017). Current findings suggest the phenotypic plasticity of cancer stem cells (CSCs) are less constrained than believed, dynamic and microenvironment dependent. Identifying the master GRN coordinating cancer stemness remains an
intractable problem and thus, a roadblock for cancer stem cell reprogramming; the primary reason being the epigenetic complexity of tumors.

Epigenetic dysregulation is a hallmark of cancer. The epigenetic burden is greater in pediatric cancers than in adults (Det et al., 2018; Filbin and Monje, 2019). The epigenome remains a black box in systems biology, governing stem cell reprogramming both in vivo and in vitro. Moreover, there are multiple, complex programs such as the transfer of exosomes and microRNAs involved in the cancer stemness problem. For instance, by regulating the PI3K/AKT/mTOR signaling pathway, miR-126 confers leukemic stem cells’ self-renewal, quiescence (dormancy) and therapy resistance (Lechman et al., 2016). Despite the complexity, several patterns in cancer stem cell networks have been recognized in the past decade indicating possible routes towards their cell fate reprogramming. To illustrate, a few examples of epigenetic patterns identified in highly morbid, pediatric brain cancers are given.

Stem cell fate transitions encounter many environmental fluctuations and molecular heterogeneities. However, transcriptomics revealed glial-tumor interactions regulate brain metastases via upregulated EMT/MET pathways. The EMT pathways coordinate chemical pattern formation (i.e. morphogenesis) and confer dynamic switching in cancer stem cell fates (Wingrove et al., 2019).

Transcription factors can exhibit chaotic dynamics (Heltberg et al., 2019). For instance, Nanog heterogeneity arises from fluctuations in gene networks and sudden burst-like, intermittent transcription in the coexisting states (Smith et al., 2017). The Myc network, is a global amplifier of gene expression. Molecular studies revealed Myc increases transcription duration and heterogeneity (Patange et al., 2019). Sox2, another Yamanaka factor, is also an essential driver of cancer stem cell sub-populations in glioblastoma (Suva et al., 2013; 2014). PI3K/mTOR (mammalian target of rapamycin) and MEK/ERK pathways were shown to be critical to the self-renewal in glioma stem cells (GSC) and mediate cancer stemness in brain tumors (Sunayama et al. 2010). Regardless of the divergent, clonal evolutions of tumors, key driver mutations in histone H3 post-translational modifications (onco-histones) and IDH1/2 (isocitrate dehydrogenase) were tractable in all clusters indicating patterns of epigenetic reprogramming in cancer stem cells (Suva et al., 2014; Salloum et al., 2017).

Progression to higher grade glioma displayed an overall decrease in methylation, and hypermethylation of a small subset of CpG islands associated with developmental regulators, including FOX, SOX and TBX family genes. The epigenetic reprogramming may arrest cells into a permanently self-renewing state (Bai et al., 2016). TERT promoter/telomerase mutations often occurred later for rapid growth and relapsed tumors, indicating a critical attractor in cancer development (Korber et al., 2019; Stead and Verhaak, 2019). TERT promoter mutations is needed years pre-diagnosis of chromosomal alterations for detectable tumors. Such causal developments indicate ‘sensitive dependence on initial conditions’ (i.e. cascading effects).

Single-cell transcriptomics-based cell lineage reconstruction revealed four distinct targetable clusters in pediatric medulloblastoma (Northcott et al., 2017; Vladoiu et al., 2019). Apart from differential gene expression signatures, CpG island methylation profiles were also distinct between the clusters as shown by t-SNE (t-distributed stochastic neighbor embedding) plots. CRISPR screens identified stemness and chemotherapy resistance regulators in patient-derived glioblastoma stem cells including members of Sox (Sox9, Socs3, USP8, DOT1L and protein ufmylation) (Macleod et al., 2019). The epigenetic landscape of GBMs (Glioblastoma Multiforme) shows tremendous spatiotemporal heterogeneity. However, a core set of neurodevelopmental transcription factors (POU3F2, SOX2, SALL2, OLIG2) were identified to be essential for GBM propagation in brain cancer stem cells (Suva et al., 2014). As seen, single cell multi-
omics are indicating a link between the Yamanaka factors and the epigenetic reprogramming of cancer stem cells. Machine learning algorithms and information-graph theoretic approaches are currently employed to infer causal patterns from these multi-omics datasets and reverse-engineer the cancer interactome.

**MACHINE LEARNING**

Machine Learning is paving the future of precision oncology. The algorithms are trained to identify characteristic features in complex datasets and classify/predict outcomes based on learned pattern recognition. Deep learning healthcare will allow AI (artificial intelligence)-assisted decision-making in precision oncology (Esteva et al., 2019; Topol, 2019). IBM Watson’s individualized cancer diagnosis is a good example. Deep convolutional networks can assess complex drug interactions in patients. They are also used in quality assessment of protein folding and protein structure prediction from sequence (Tong and Altman, 2009). Digital pathology, cell-lineage reconstruction from biopsies, evidence-based drug screening (personalized pharmacogenomics), and data-science driven identification of therapeutic targets/biomarkers are merely few examples of Deep learning healthcare (Silberbush et al., 2019).

Deep convolutional networks use multi-layered information processing and back-propagation to classify-predict complex signals such as images, speech and video data (LeCun et al., 2015). Deep learning nets are revolutionizing clinical pathology and medical imaging reconstruction (Shan et al., 2019; Zhang et al., 2019). Unsupervised methods of machine learning can perform signal processing, feature extraction and pattern recognition on newly presented datasets, while supervised learning methods classify cancer from non-cancerous data sets based on trained databases.

For e.g., Deep learning nets cluster-classified the diverse subgroups in PDAC (pancreatic cancer), a highly heterogeneous and morbid disease. Six molecular and clinically distinct subtypes of PDAC were identified with 160 subtype-specific markers (Zhao et al., 2018). Multi-omics data and machine learning are used to predict metabolic pathways in cancer resistance (Castello and Martin, 2018). Deep learning can predict microsatellite instabilities in gastric cancers and thus, improve immunotherapy decisions in patients (Kather et al., 2019). Bayesian hierarchical clustering of heterogeneous stem cells is feasible with microfluidics and droplet-Seq technology (Shun et al., 2019). Such cutting-edge Drop-Seq technologies will also pave point-of-diagnosis and personalized interventions. These are merely examples to illustrate oncology is transitioning towards Deep learning healthcare.

Machine learning is at the frontier of tackling the cancer stemness problem. A recent finding shows cells expressing CSC-associated cell membrane markers in Glioblastoma (GBM) do not represent a clonal entity defined by distinct functional properties and transcriptomic profiles, but rather is a ‘plastic state’ that most cancer cells can adopt (Dirkse et al., 2019). In other words, Dirkse et al. (2019) claim phenotypic plasticity is a non-hierarchical, emergent behaviour in cancer stem cells. The t-SNE projections of GBM scRNA-Seq found cellular subpopulations resembling different expression subtypes co-occurring in the same tumor which adapt to heterogeneous phenotypes in disease progression (Patel et al., 2014; Yuan et al., 2018). Hence, all GBM cancer cells are claimed to be cancer stem cells as opposed to hierarchy-based, clonal subections.

On the other hand, another group’s findings suggest conserved neural trilineage, a hierarchy with glial progenitor-like cells at the apex in GBM (Couturier et al., 2018). A droplet-based scRNA-Seq of patient-derived GBM cells, generated two or three cancer clusters by t-SNE and the Louvain community...
structure detection algorithm. Following PCA (principal component analyses) to assess intra-tumoral heterogeneity within the enriched GSCs (glioma-stem cells) of each grouping, the data suggested that GSCs are organized into progenitor, neuronal, and astrocytic gene expression programs, resembling a developing brain. However, the paper does not reject the micro-environment dependent phenotypic plasticity and interconvertibility of cell fates. These data lack time-series gene expression, a feat that is methodologically difficult. Hence, whether there is a fixed hierarchical clustering in stemness or any cancer cell can acquire (or is in) a stem cell fate depending on its environmental cues, is a highly debated, complexity problem.

Furthermore, cancers are complex adaptive ecosystems with multi-scaled, interaction networks. Liquid biopsies are enriched with circulating tumor cells (CTCs), circulating tumor DNA (ctDNA), tumor-derived extracellular vesicles, free nucleosomes and many other oncogenic biomarkers representing a vast repertoire of information on the patient’s tumor heterogeneity and therapy resistance outcomes. Exosomes are nanoscopic (~30-100 nm), heterogeneous packets of information released by cells forming complex, intercellular communication networks. Exosomes are emergent, reprogramming machineries used by cancer ecosystems to regulate cell fate dynamics (Guo et al., 2019). That is, exosomes can reprogram distant tissue microenvironments into pre-metastatic niches and horizontally transfer malignant traits such as therapy resistance to promote aggressive cancer phenotypes (Wei-xian et al., 2014; Hu et al., 2015; Zhang et al., 2018; Choi et al., 2018; Keklikogolu et al., 2019). Exosomes derived from hESC (human embryonic stem cell) microenvironment has been shown to suppress cancer phenotypes and reprogram a subset of malignant phenotypes to healthy-like, benign plastic states (Camussi et al., 2011; Zhou et al., 2017). An elevated expression of the Yamanaka factors and plausibly microRNAs in the nano-vesicles were identified as the mechanism for the cell fate reprogramming. Machine learning is reshaping our understanding of these many-layers of cancer complexity.

Machine learning algorithms applied on Intraoperative Raman spectroscopy can distinguish brain cancer stem cells from normal brain tissue (including both invasive and dense cancers) with an accuracy, sensitivity, and specificity > 90% (Jeremy et al., 2015). Background subtraction algorithms (e.g. rolling ball algorithm) and autofluorescence removal algorithms were employed to distinguish the cancer stem cells’ Raman bands from those of healthy tissues (Brusatori et al., 2017; Zhao et al., 2007).

Furthermore, machine learning is paving minimally invasive, cancer screening with liquid-biopsy derived exosomes characterization. Simple machine learning algorithms such as binomial classifiers (GLMnet) with the caret R package, can be trained to classify tumor-associated methylome patterns in cell-free plasma DNA (Shen et al., 2018). Hence, machine learning of aberrant epigenetic signatures in ctDNA are diagnostic indicators for cancer patients and provide precision oncology for relapse patients. Patient-derived exosomes were analyzed using Raman spectroscopy with a 785 nm laser at 5mW irradiance onto samples plated on calcium fluoride substrates. The cluster analysis and interpolation resolved Raman spectroscopy found that cancer and healthy exosomes clustered differently when subjected to simple dimensionality reduction techniques like PCA (Gualerzi et al., 2017; 2019).

Similarly, surface enhanced Raman scattering (SERS) signals of exosomes from normal and NSCLC (Non-small-cell lung carcinoma) cells on Gold-nanoparticle substrates were performed. The Raman spectra of cancerous exosomes showed unique peaks in the vibrational bands distinguishing NSCLC exosomes from healthy clusters when subjected to spectral decomposition by Principal Component Analysis (PCA) (Shin et al., 2018; Rojalin et al., 2019). Correlation analysis further revealed potential exosome surface protein markers, where cancerous exosomes had similar spectra (Park et al., 2017; Shin et al., 2018).
Similar findings with peak fitting algorithms and MCR-ALS algorithm (multivariate curve resolution-alternating least squares) on Raman spectra were used to cluster-classify pancreatic cancer exosomes on gold-nanoparticle plated SERS substrates (Banaei et al., 2017). Similarly, a recent study shows neural networks optimized with machine learning algorithms such as principal component analysis-linear discriminant analysis (PCA-LDA) and support vector machines (SVM) can distinguish oral cancer patients from healthy individuals by characteristic signatures in the Fourier transform infrared (FT-IR) spectroscopy of their salivary exosomes (Zlotogorski-Hurvitz et al., 2019). These findings indicate simple machine learning algorithms coupled to nanofluidics/nanoplasmonic chips with built-in SERS or other spectroscopic techniques, may pave point of care diagnostics for early detection and prognostic screening in cancer relapse patients (Rojalin et al., 2019).

NETWORKS RECONSTRUCTION

Machine learning algorithms are pioneering cell lineage reconstruction from single-cell RNA sequencing datasets. The gene expression patterns of patient-derived tumor samples can be reconstructed into graph-theoretic, flow networks. For example, dynamic Bayesian network analysis is an algorithm, where given protein concentration changes in time, the factors most connected in the regulatory networks are determined by Bayesian inference to reconstruct the GRN circuitry. Dynamic Bayesian networks model dynamical systems as steady states through probabilistic Boolean networks. Moreover, simple machine algorithms can display the growth patterns of cancer. Branching processes (also known as, random trees) simulated by the Gillespie algorithm are used to model clonal hematopoiesis in leukemia models. The birth-death process can rapidly undergo exponential growth and bifurcate into distinct lineages, whereby specific mutational subclones can be traced (Altrock et al., 2015).

The general approach to scRNA-Seq based cell lineage clustering and pseudotime inference is based on data filtering (pre-processing) followed by machine learning algorithms. Hence, the raw data reads are filtered and selected for most differentially expressed genes in the cells. False counts remain a current problem. The data is then fed into the machine learning algorithms as a gene expression matrix of cells (columns) by genes (rows), followed by dimensionality reduction techniques like PCA (principal component analysis) or t-SNE, etc. Then, network-graph theoretic approaches reconstruct the spatial neighborhood of cells with statistical inference (e.g. k means clustering (unsupervised) or k-nearest neighbor kNN graph (supervised)). Following, optimization algorithms are used to find trajectories and regulatory modules. Correlation analysis assigns a score to potential gene-gene interactions, after unsupervised community detection algorithms or partial-information decomposition algorithms are employed. Dispersion cell -cell variability measures generally use covariance, correlation measures, Bayesian-statistical inference and Shannon or metric entropy. The mean entropy generally increases in cell fate transitions (Chan et al., 2017).

Topslam estimates state transition pseudotime by mapping Individual cells to the surface of a Waddington-like landscape with Bayesian Gaussian process latent variable model/GPLVM, a nonlinear probabilistic dimensionality reduction (Zbiessele and Lawrence, 2017). These methods project high dimensional data into 2 or 3 components, whereby distances are interpreted as cell-cell variability for cluster analysis and constructing neighborhood graphs. However, such statistical inference methods are sensitive to fluctuations in gene expression data (environmental, intrinsic, etc.) and identify key regulators without understanding the system dynamics nor complexity driven by multi-scale molecular interactions. This is generally the downfall for stochastic modelling of complex biosystems. As mentioned, chaotic systems exhibit patterns (strange attractors) with apparent random behaviour.
Cell Router is a graph-theoretic, flow network-based trajectory detection algorithm (Lummertz da Rocha et al. 2018). First, dimensionality reduction (PCA, t-SNE, Diffusion map, etc.), is performed on the single-cell data set. Following, a kNN Graph (k-Nearest neighborhood) is assessed. Then, the Jaccard index finds the similarity between two cells where if they belong to the same cluster, a high correlation is observed. The Louvain community structure detection algorithm configures the populations by assessing weights of the graph by similarity of cell-cell interactions. As such, a source-to-sink, directed graph is projected mapping the GRN flow network. In a directed graph, the nodes are connected by flow arrows with the weighting indicating their capacity. The trajectories are found with Bellman-Ford algorithm, a type of cost flow optimization algorithm, and ranked by total flow, cost and length between vertices. The Bellman-Ford algorithm computes the shortest paths from a single source vertex to all vertices in a weighted graph. It is slower than the Dijkstra’s algorithm but can handle graphs with edge weights that are negative. Lastly, the optimized trajectories are ranked by GRN scores given by Pearson-Spearman correlation. Corresponding heat maps and dynamic curves of regulators are projected to obtain trajectories of the identified clusters, displayed as a GRN flow network.

Seurat is another algorithm that can be used as a pre-processing tool for cancer data sets. Seurat is a scRNA-Seq correlation and clustering, computational tool (Butler et al., 2018). The identified clusters can be further sorted using algorithms like CellRouter into flow networks. As discussed earlier, there is growing evidence challenging the hierarchical model of cancer stem cells. In argument, GBM shows that any cancer cell can be a stem cell depending on its microenvironment. The networks reconstruction algorithms discussed herein can validate these claims.

Single cell Energy path (scEPath) is a method for mapping the energy landscape of single-cell gene expression datasets. It reconstructs cell lineages and pseudo time inference (cell fate trajectory) with information-theoretic measures (Jin et al., 2018). The scRNA-seq data is initially pre-processed/filtered by removing low gene expressions. By calculating the Spearman correlation of the adjacency matrix, scEPath builds a GRN, calculates the normalized energy between expressions, and performs a linear dimensionality reduction with PCA. Data then undergoes cell clustering using an unsupervised framework called single-cell interpretation via multikernel learning (SIMLR). From the cell clusters, Boltzmann–Gibbs distributions are used to find the transition probabilities. To infer cell lineages, scEPath first constructs a probabilistic directed graph with maximum probability flow, equivalent to finding a minimum directed spanning tree (MDST) from Edmonds’ algorithm. Finally, scEPath uses the R “princurve” package to fit a principal curve of the clusters to compute the pseudo times.

The pre-processing in scEPath is similar to Slingshot. Slingshot is a trajectory-cell lineage classification algorithm using dimensionality reduction techniques and cluster analysis, as for any of the above (Street et al., 2018). Diffusion maps, PCA, etc. are used as dimensionality reduction followed by model-based clustering. The expectation-maximization algorithm, a class of maximum likelihood measures and Bayesian inference are applied to the dataset. As an alternate, k-means clustering can be assessed to reconstruct relative cell fate commitments. Lastly, minimum spanning trees (MST) are performed on the clusters using Prim’s algorithm to infer cell fate branching. Some of the best machine learning approaches in mapping cell fate choices consist of t-SNE followed by MST on single-cell datasets. However, neural networks are preferred for larger datasets.

J.J. Hopfield (1982) introduced neural networks with spin glass formalism. The Hopfield network is a recurrent neural network displaying associative memory. The Hopfield network is used as computational devices to approximate solutions of NP-hard problems such as the attractor reconstruction of cancer networks (Bossomaier and Green, 2000; Szedlak et al., 2014). Lang et al. (2014), using single-cell data,
demonstrated the Hopfield neural network can reconstruct the corresponding cell states of the Waddington landscape as spin glasses. Taking the orthogonal projection of key transcription factors’ gene expression created a subspace representative of attractors. Correlation measures and z-scores then defined the distinct clusters corresponding to the gene regulatory networks. Partially reprogrammed cell fates emerged as hybrids co-expressing signals from multiple cell fates in the landscape (i.e. spurious attractors) (Lang et al., 2014).

Hopland is a continuous, Hopfield neural network-based algorithm which interprets single-cell RNA-Seq or qPCR data for Waddington landscape reconstruction (Guo and Zheng, 2017). The Hopfield algorithm feeds the gene expression data matrix into the Hopfield network to reverse-engineer the Waddington landscape’s topography. The cell fate attractors are reconstructed based on gene to gene expression correlation. First, Isomap dimensionality reduction is performed. Isomap generally consists of a kNN nearest neighborhood graph followed by geodesic calculations between the nodes (genes) from the Dijkstra’s algorithm and multidimensional scaling. The fast-marching algorithm devised for solving the Eikonal equation, was used to calculate the geodesic distances on the landscape as the weights of edges connecting the cells. The gene networks were modeled on a triangulated-mesh, grid scheme to compute the geodesics based on cell-cell variability in gene expression clusters. Each gene is modelled as a neuron in the network, the cumulative energy values of which add up to each cell fate on the landscape. A Gaussian-mixture optimization algorithm was used to infer the parameters, and the mean values of the outputs (gene expression values) form the cell lineages.

Then, the Gradient descent algorithm optimizes the Hopfield network whereby the activation values (weights of the edges) undergo a relaxation process by minimizing the Lyapunov energy function. The energy minimization derives the local attractors. The Lyapunov energy function gives the energy values corresponding to the cell states, where a high energy corresponds to less differentiated states (hills), while low energy indicates differentiated states (valleys). In classical theory, cell differentiation paths follow the lowest potential energy in landscape. The Lyapunov energy function is given by:

\[ E = -\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} W_{ij} U_i U_j + \sum_{i=1}^{N} I_i U_i + \sum_{i=1}^{N} \int_{0}^{u_i} g^{-1}(u) du \]

Here g is the activation function, \( U_i = g_f(V_j) \), where V is the neuron outputs for N genes, W is the weights assessed from gene expression data as described above, \( \delta_i \) is the gene signal degradation rate, and \( I_i \) is the combination of propagation delays and regulations from environment, classified as ‘noise’. To simulate dynamic trajectories, the 1st order Runge-Kutta finite difference is used to solve the updating of neurons (genes) as an ordinary differential equation, while the Euler-Maruyama method is used for solving stochastic differential equations, such as the Langevin equations underlying the network dynamics. These two techniques are applied for high molecular concentrations inferred by the gene-gene correlations, whereas, the Gillespie algorithm is used by the neural network for low copy numbers (i.e. categorized as stochastic fluctuations) (Guo and Zheng, 2017). To visualize the landscape topography, the GPLVM dimensionality reduction technique is applied and lastly, the MST algorithm identifies the minimum trees connecting the clusters.

The application of a neural network-based classifier on whole-genome RNA-Seq data of cancer patients showed a near 86% success rate in diagnosing complex, metastatic cancers (n= 201) and 99% success rate on primary cancers (Grewal et al., 2019). However, esophageal cancers and adenocarcinoma were often misclassified. Neural networks were trained and optimized for the highest probabilistic inference
of pathological assessment. A variety of machine learning methods including ANN (artificial neural networks), Bayesian networks, multi-class support vector machines and decision trees were used as predictive models for the cross-validation and training in assessing cancer aggressiveness and pathological staging. Then, confidence scores were assigned in their performance evolution. The weight analysis of the neural network was used to identify the genes most important in class prediction of tumor subtype (Kourou et al., 2015; Grewal et al., 2019).

Deep learning neural networks are amidst the most effective heuristics currently used in biological networks reconstruction from scRNA-Seq profiles (Tian et al., 2019). scDeepCluster uses a model-based cluster analysis through multi-layered neural networks. Multi-kernel spectral clustering methods and community detection methods are amidst the most commonly used unsupervised learning methods in cell lineage reconstruction. However, such methods do not account for the false zero (low RNA capture) counts and scalability (dimensionality reduction) issues in scRNA-Seq cluster analysis. Deep learning embedded clustering was proposed as a solution. The autoencoder, a type of deep neural networks was used to replace the mean square error (MSE) loss function with a zero-inflated negative binomial (ZINB) model-based loss function. The autoencoder performs a nonlinear function mapping of the read count matrix of scRNA-Seq data to low dimensional latent space. Following, clustering analysis is performed by the Kullback–Leibler (KL) divergence which characterizes the relative Shannon information entropy. The KL divergence was assessed by the ‘deep embedded clustering’ (DEC) algorithm along with noise-reduction techniques to classify the distinct cell fate clusters (Tian et al., 2019).

Furthermore, time-delay embedding algorithms are barely in practice in biological networks reconstruction. As an example, convergent cross-mapping of cytokine networks within the hematopoietic system was recently used to identify causal trajectories in regulatory transcriptional networks (Krieger et al., 2018). Cytokine immune networks were inferred as directed graphs, from time-delay embedding mapping, where Pearson coefficients assessed the network connection strengths. However, such algorithms are limited to the size of the data set.

The efficiency of the discussed algorithms is model-dependent. For instance, Topslam outperformed Hopland in dissecting the scRNA-Seq of mouse embryo development, without the Hopfield element using only GPLVM and MST. Moreover, the data sets in these algorithms were highly filtered. As noted, the initial step was pre-processing the data reads, where low-gene expressions are filtered-out. However, Deep Learning is emerging as a solution to process large datasets as demonstrated by scDeepCluster. With time-series analysis, a filtered-out, low-gene expression signal can emerge as a strange attractor (i.e. sensitive dependence to small perturbations). Moreover, the attractors are assumed to be steady-states, whereas cancer’s Waddington landscape is a dynamical system governed by non-equilibrium statistical mechanics. The possibility of nonlocal cross-talks between these networks remains a query as well. Lastly, environmental perturbations are considered as ‘noise’, whereas the epigenome remains the governing principle for cell fate decision making and reprogramming. Distinguishing noise from chaotic processes remains an intractable problem in systems biology, mainly due to the lack of time-series datasets.

**PATTERN FORMATION**

Thus, many critical attractors of GRN circuitries are being revealed with multi-omics mapping of highly morbid cancers. Machine learning algorithms, and more recently Deep Learning, are cutting-edge technologies in interpreting these complex datasets, finding patterns within and mapping single cell fate trajectories. However, these methods can still be misleading as a lot of data filtering is done prior to...
analysis, wherein valuable information may be lost. Cell to cell variability is assumed on the reduced latent components from high dimensional data whereas, low-gene expression levels can result in chaotic fluctuations in cell activity. Moreover, what distinguishes a cancer stem cell remains an unsolved problem.

Herein, the reason to these problems is indicated as the ignorance of the fluid dynamics underlying cell fate transitions. The complexities emerging in cancer pattern formation are not accounted for in these algorithms.

Cancer (stem) cell fate bifurcations are governed by pattern formation systems, namely reaction-diffusion equations and simplified Navier-Stokes equations. The network reconstruction algorithms discussed above do not analyze the emergent behaviours in cell fate fluid dynamics. While the lack of time-series datasets is a fundamental limiting factor in testing the presence of strange attractors within the attractor landscape, the lack of consideration for fluid dynamics within cell fate transitions is proposed as the missing paradigm. As seen, the neural networks-based Waddington landscape reconstruction assume certain numerical approximation methods to model the gene expression dynamics as stochastic processes. The chemical kinetics are generally assumed to be governed by Hill-Langmuir equations or Arrhenius rate laws.

Strange attractors emerge as solutions of highly complex fluid flows, such as the turbulences of the Navier-Stokes equations. While cells are considered highly viscous micro-scale systems, chemical turbulence can occur in sudden, short puffs at the onset of pattern formation (i.e. intermittency). As paradoxical as these strange ideas may seem, recent findings challenge the dogma and indicate the emergence of low-Reynolds turbulences within cellular biosystems.

Turing (1952) first coined the term morphogen to describe the chemical oscillations that result in coat patterns of animals. The stripes, spots and spirals self-organized from Turing’s reaction-diffusion systems are given by nonlinear partial differential equations, the general form of which is given by Fick’s second law of diffusion with a reaction term, as:

\[
 \frac{\partial u}{\partial t} = D \nabla^2 u + R(u, t)
\]

where \(u\) is the velocity or concentration vector, \(\nabla\) is the gradient and \(R\) is the local reactions (coupling) term. The simplest reaction-diffusion equation pertaining to pattern formation is the Fisher-Kolmogorov–Petrovsky–Piskunov equation (FKPP) (Fisher, 1937; Kolmogorov et al., 1937). A simplified tumor growth-invasion system is given by the FKPP equations:

\[
 \frac{\partial N_1}{\partial t} = r_1 N_1 \left(1 - \frac{N_1}{K_1} - \frac{N_2}{K_2} \alpha_{12}\right) - d_1 L N_1 + \nabla \cdot (D N_1 (N_1, N_2) \nabla N_1)
\]

\[
 \frac{\partial N_2}{\partial t} = r_2 N_2 \left(1 - \frac{N_2}{K_2} - \frac{N_1}{K_1} \alpha_{21}\right) - d_2 L N_2 + \nabla \cdot (D N_2 (N_1, N_2) \nabla N_2)
\]

\(r_1\) and \(r_2\) are the growth rates of normal and tumor tissue respectively, \(d\) is the cellular susceptibility to excess acidic conditions, \(N_1 = \text{normal tissue density}, N_2(r, t) = \text{density of the neoplastic and } L(r, t) = \text{the excess proton } [H^+] \text{concentration (assumed normal tissue to be of logistic growth rate } r_1 \text{ and carrying...}

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 September 2019
doi:10.20944/preprints201909.0304.v1
capacity K), $\alpha$ is the Lotka-Volterra (competition) strength, D is the respective diffusion coefficients and death rate is proportional to L (Gatenby and Gawlinski, 1996). For the neoplastic growth, it is the same equation denoted by a subscript 2 and a lack of death amidst excess acid (adaptive). Changes in oxygen and nutrient uptakes, especially in hypoxia (a signature of most glycolytic cancers), are modelled using such equations. Reaction-diffusion equations are used to model the ECM (extracellular matrix) dynamics and cell-ECM adhesion remodelling (Ramis-Conde et al., 2008). Turing equations were shown to well-describe cancer-immune cell invasion dynamics as well (Zheng et al., 2018).

Current models of cancer pattern formation include simplified fluid dynamics equations. For instance, the Hagen-Poiseuille flow and FKPP equation \( \frac{\partial c}{\partial t} = D \nabla^2 c + \rho c (1 - \frac{c}{c_{\text{max}}}) \) have been used to model GBM growth, where c is the chemical species’ concentration and $\rho$ is the tumor density (Cai et al., 2017, Altrock, 2015). The Fisher-KPP equation was used as a prognostic indicator in breast cancer patients undergoing neoadjuvant therapy. The reaction-diffusion equations well-described tumor growth in accordance to clinical imaging data (Weis et al, 2015). The model well-characterized the reaction-diffusion based drug delivery to tumors in clinical settings as well (Jarrett et al., 2018).

The FKPP-model of GBM exhibits travelling wave solutions where the tumor invasion wavefronts have a velocity of $v = 2 \sqrt{D \rho}$ (Harko and Mak, 2015). The cell fate trajectories in the FKPP model are depicted as branching diffusion trees (Derrida and Spohn, 1988). The speed of its wavefront is the free energy of the polymer and the minimal speed is the phase-transition into a glassy phase (Derrida and Spohn, 1988). The wavefront velocity is dependent of its initial condition and accepts solutions of travelling wave solutions given by $w(x - ut)$ where the velocity $u = \frac{1}{\beta} \log \left[ k_B \int dV \rho(V) e^{-\beta V} \right]$ and $V$ is the volume, if $\beta > \beta_c$. Here, $\beta_c$ is the critical inverse temperature which is the effective noise parameter in Hopfield spin glasses (Lang et al. (2014)). The probability density $\rho$ is often assumed to be a Gaussian distribution. The result of these mathematics is a bifurcation tree with nodes at different points on the landscape (attractors) where the energies of the random walk change in time. The resemblance of FKPP solutions as branching trees may explain why the above-discussed, MST (minimum spanning trees) and GPLVM were amidst the optimal algorithms in reconstructing cell lineage bifurcations.

Darcy’s law has been used to model GBM patterning as well, where the mass-averaged velocity of solid components (tumor cells) $u_s$ is defined as:

$$u_s = -(\nabla P - \frac{\delta E}{\delta \varphi_T} \nabla \varphi_T)$$

Where $\varphi_T = \text{total tumor volume}$, $E$ the focal adhesion energy given a double-well potential and $P$ is the solid tumor’s pressure (Yan et al., 2017). Darcy’s law can be derived from the 3D Navier-Stokes equations for a shallow flow between two plates.

Electric cell impedance recordings were performed in rat’s prostate cancers. The time-series, Fourier analysis of cancer micro-motions were assessed by Taken’s theorem (time-delay embedding) to detect patterns distinguishable from a random signal. The attractor reconstruction showed positive Lyapunov exponents in phase portraits (i.e. signature of chaos) (Posadas et al., 1996). Chaotic attractors in cancer tissue pathology are indicative of recurrence and increased aggressiveness (Khajanchi et al., 2018). Fractal dimension analysis can distinguish cancers from healthy tissues, which now is redefined as edge detection and feature extraction in Deep Learning assessed tissue pathology (Lennon et al. 2015).
Chaotic attractors emerge in the reaction-diffusion modelling of tumour growth depending on changes in the control parameters, which vary amidst different phenotypes of a heterogeneous ecosystem (e.g. oxygen concentration, glucose level, tumor volume, diffusion from surface and growth-parameters) (Itik and Banks, 2010). Lyapunov exponents and fractal dimension were calculated for cancer patterning. While the Lorenz attractor has a fractal dimension of 2.06, the cancer models showed ~2.03 indicating a chaotic attractor with Shilnikov-bifurcations (Itik and Banks, 2010). Ivancevic et al. (2008) showed a Lorenz-like, chaotic attractor best describes the reaction-diffusion of cancer cells. Cancer modelling was performed using a three-body ecosystem consisting of host, immune and tumor cells. Assuming initially logistic growth, the Lotka-Volterra dynamics became chaotic as denoted by the period doubling cascades in the bifurcation diagrams (Letellier et al., 2013). The bifurcation analysis revealed Rossler-like attractors with a fractal Lyapunov dimension. The use of reaction-diffusion systems to predict cancer dynamics has been compared to forecasting weather patterns (Yankeelov et al., 2013; 2015; Tang et al., 2014). The emergence of Lorenz-like and Rossler-like attractors indicate cancer cell fates are strange-attractors of the Waddington landscape.

TURBULENCE

Lorenz (1963) first demonstrated strange attractors can be solutions to approximate weather turbulence, considering a Benard-Rayleigh convection model. Whether the Navier–Stokes equations allow solutions that develop singularities in finite time remains unresolved, known as the Navier-Stokes smoothness and existence problem. While this remains a complexity problem, there is an immense body of experimental and theoretical works confirming the geometric structures formed by turbulent flows consist of strange attractors (Ruelle, 1980; Landford, 1982; Miles, 1984; Brandstater and HL Swinney, 1987). Strange attractors are spectral signatures of turbulence in experimental hydrodynamics (Ruelle, 1995).

In experimental fluid dynamics, turbulence arises in sudden large bursts, followed by a period of relative quiescent behaviours, defined as intermittency. A continuous frequency spectrum is observed in all cases of experimental fluid turbulence (Ruelle, 1995). The frequency spectra (i.e. square amplitude of each frequency of a fluid) can be used to detect strange attractors through attractor reconstruction methods and peak analysis. There is no direct test to ‘sensitive dependence on initial condition’ in hydrodynamics. Hence, a frequency analysis of the fluid must be performed (Ruelle, 1973; 1995). These heuristics have been implemented on cancer morphology as discussed above. However, they remain to be applied in identifying complex cell fate bifurcations during the onset of chemical turbulence in cancer (stem) cells.

Recent data suggests the roles of chemical turbulence in the emergence of ordered structures during cellular pattern formation and protein folding. While conventionally cells are assumed to be highly viscous structures with laminar protein flows, phase-transition to chemical turbulences can occur in reaction-diffusion systems. Chemical turbulence in protein-mediated cell patterning was recently confirmed experimentally (Brauns et al., 2018; Denk et al., 2018). Paradoxically, following turbulence transition, coherent patterns emerged through diffusively coupled local equilibria. The chemical turbulences were shown to occur in very rapid bursts followed by laminar-like flows (i.e. intermittency).

Cell division is orchestrated by intracellular protein patterning, mainly from EMT pathways, cytoskeletal filaments and cell polarity complexes. Abnormal cell division is the primary signature of cancer. Usually the dynamics corresponding to low concentrations of intracellular proteins are modelled as stochastic
fluctuations. Classical theory states these chemical systems are close to equilibrium and inertial effects are negligible, given a highly crowded, viscous cytoplasm. However, a recent theory by Halatek and Frey (2018) challenged the dogma. Using the finite elements method (FEM), a method used to approximate complex fluid flows in CFD, simulations predicted chemical turbulence at the onset of the pattern-forming instabilities in tissues. Chemical turbulence was qualitatively used to characterize the spatiotemporal chaos observed at the onset of pattern formation. Cytosolic diffusion constants $D_c$ were in the order of $60 \mu m^2 s^{-1}$, where the MinD-ATP/ADP in bulk were given by the Turing reaction-diffusion equations:

$$
\partial_t u_D(z,t) = D_c \nabla^2 u_D - F_{u_D}
$$

$$
\partial_t u_T(z,t) = D_c \nabla^2 u_T + F_{u_D}
$$

Where, $u_D$ and $u_T$ define the cytosolic density of MinD-ATP/ADP conformations and the MinD-ATP was assumed to bind to the membrane via nonlinear coupling rate constants. The resultant kymograms demonstrated turbulent flow patterns at low MinE/MinD ratios. As mentioned, low-levels of gene expressions or protein fluctuations are filtered out in the single cell datasets in the pre-processing step for cell lineage tracking algorithms. Herein, low levels of the cell polarity complex MinE/MinD ratios were shown to produce the chemical turbulences. The theoretical predictions of Halatek and Frey (2018) were confirmed experimentally (Denk et al., 2018).

The PAR cell polarity complex is the mammalian equivalence of the Min proteins which coordinate cancer cell division. According to these findings, chemical turbulence in pattern formation is used as a synonym for spatio-temporal chaos, i.e. a broad distribution in the power spectrum and a low spatial correlation length reminiscent of the Kolmogorov spectrum. But none of these terms are strictly/unambiguously defined in the literature. The term was adapted from the work by Nobel laureate Gerhard Ertl on reactions of heterogeneous catalysis (Kim et al., 2001). According to this work, during chemical turbulence, both the amplitude and the phase of local concentration oscillations are strongly fluctuating creating spiral waves as seen in the findings of Denk et al. (2018) with Min proteins pattern formation. Hence, turbulent chemical oscillations can give rise to both, patchy, multi-fractal structures and ordered patterns in cellular reaction-diffusion systems (Ouyang and Swinney, 1991; Mecke, 1996; Ruelle, 1995).

To further illustrate the paradoxical role of turbulence within cells, recent evidences show protein folding may best be described by the turbulence of Kolmogorov-Richardson’s energy cascade, where turbulent eddies and vortices breakdown into smaller fractal hierarchies. Computational simulations of SH3-domain protein folding exhibited a fractal nature and the flow patterns are filled with 3D eddies containing strange attractors, where the tracer flow paths behaved as saddle-node bifurcations (Kalgin and Chekmarev, 2011; Andryushchenko and Chekmarev, 2017). The mathematical modelling best-fitted Kolmogorov’s energy spectrum for turbulent flows (Kolmogorov, 1941). SH3 domains are critical in the focal adhesion complexes regulating cancer metastasis and EMT transitions. Likewise, the folding dynamics of villin subdomain HP-35 protein in a FRET (Forster resonance energy transfer) experiment was shown to obey the $\beta$-model of turbulence with many orders (scales) of turbulent flow transitions for the eddies in the 3D conformational space (Andryushchenko and Chekmarev, 2016; Chekmarev, 2018).

The indirect role of fluid turbulence in cancer metastasis is quasi-established (Koumoutsakos et al., 2013; Goetz, 2018). The complexity of tumor vasculature and tissue organization can be quantified in
terms of fractal dimensions, where blood vessels showed only intermittent flows (Jain and Baish, 1998; Nasu et al., 1999). As mentioned, multi-fractality is a characteristic feature of both, turbulent flows and tumor complexity (Sreenivasan and Meneveau, 1991; Coffey, 1998). Huang et al. (2018) showed that non-laminar shear stress may increase the adhesive ability of cancer cells in metastatic invasion. Higher systemic flows facilitate EMT transitions in cancer cells during metastasis (Rizvi et al., 2013). The fluid shear stresses experienced by CTC (circulating tumor cells) cells consist of turbulence, where the Navier-Stokes equations are used to compute their systemic flow maps (Rejniak et al., 2016). With appropriate boundary conditions, the steady, incompressible Navier-Stokes and continuity equations are solved using a finite-volume method to model circulating melanoma cells (Behr et al., 2015). Moreover, hematopoietic iPSC (induced pluripotent) stem cells were shown to increase their production rates by many orders of magnitude under turbulent flows (Ito et al., 2018).

Furthermore, active turbulence is an emerging field in the study of soft matter systems. The hydrodynamic interactions of active fluids give rise to the emergence of meso-scaled vortex patterns reminiscent of two-dimensional turbulence (Wensink et al., 2012; Sumino et al., 2012). Cell tissues and reconstituted cytoskeletal solutions exhibit active turbulence, where emergent scaling behaviours exhibiting power-variants of the Navier-Stokes equations were observed. Self propelled particles like cytoskeletal flocking, follow continuum mechanics of a fluid and can exhibit turbulence even at low Reynolds numbers (Marchetti et al., 2013; James et al. 2018; Doostmohammadi et al., 2018).

Furthermore, the emerging field of ‘quantum biology’ further blurs scaling behaviours in complexity science. Although fundamentally all processes are quantum dynamic, ‘quantum biology’ states that scale-invariant, quantum behaviours may occur in biosystems fine-tuned at the edge of chaos (Vattay et al., 2017). That is, certain biosystems exhibit extended quantum coherence and transport of macromolecules, suggesting the emergence of chaotic dynamics at the classical-quantum transition regime and quantum wavefunction-like characteristics at ambient conditions (i.e. entanglement, superposition, tunneling, wave interferences, non-locality, etc.) define meso-macroscopic emergence in soft-matter systems (Lambert et al., 2012; Vattay et al., 2017). For example, quantum coherent beats are detected in photosynthetic systems for femtoseconds in the FMO (Fenna-Matthews-Olson) complex, suggesting energy transfer is best described as a linear superposition of chromophores’ energy states (Engel et al., 2007). Craddock et al. (2014) propose a longer coherence time occurs for the Tryptophan (Trp) networks in tubulin via Forster dipole–dipole energy transfer, forming coherent beats at 600 femtoseconds. Hence, active fluids such as microtubules have been considered as a candidate for such quantum effects and proposed to behave like coherent vibrational condensates, with lacking evidence to support the claim (Craddock et al., 2014). Recently, using the Talbot-Lau interferometer, Fein et al. (2019) experimentally confirmed the extended quantum coherence up to few milliseconds of the matter wave-interference patterns from a >25 kDa assembly of oligoporphyrins. Such findings of quantum superposition of macromolecular structures provide support for alternative, fluid dynamical interpretations of quantum mechanics such as de Broglie-Bohm theory and Madelung hydrodynamics, further blurring the quantum-classical regime.

Interestingly, Trp networks of microtubules are candidates for another emerging complexity within soft matter physics known as liquid-phase condensation. For instance, reconstituted SH3-domain proteins were shown to form liquid droplets of active matter, suggesting they may form metastable condensates with the wetting features of fluids. Proteins involved in the reaction diffusion systems coordinated by PAR cell polarity complexes, centrosomes and microtubules orchestrating the cell-division, spindle network are emerging as intracellular liquid-like structures exhibiting liquid-
condensate phase transitions during mesoscale structural organization and cell fate remodelling (Shin and Brangwyne, 2017).

While the above-mentioned experimental evidences suggest turbulences can occur at low-Reynolds number, it must be noted, that the Reynolds parameter is an ill-defined quantity; a dimensional analysis argument. For instance, in a cylindrical tube, whether the length scale is considered as the radius or diameter will change the Reynolds number by a factor of 2. The scaling problem was addressed on a mathematical framework by Ruelle (2012). Turbulence was treated as a heat flow problem using the nonequilibrium statistical mechanics of moving fluid particles in three-dimensional lattice boxes (Ruelle, 2012). The macroscopic fluid transitions were characterized by Boltzmann-Gibbs distributions. Hence, such a description well applies to the metastable cell states of the Waddington attractor landscape. The critical Reynolds number was shown to be in the order of ~100 for Taylor-Couette and Rayleigh-Benard convection systems, which is relatively feasible at biologically relevant scales (Ruelle, 2012; 2014).

**ALGORITHMIC PROSPECTS**

While time-series datasets will model cancers as complex dynamical systems, the algorithms for their biological networks and Waddington-landscape reconstruction must also be adapted to time-series analysis. Delay-embedding algorithms are the state-of-art for chaotic attractors reconstruction. Domain translation methods (i.e. image-to-image translation) are appropriate embedding algorithms to quasi-map chemical turbulences observed within cells. Fluorescence reporters-based intracellular protein labelling can help track the concentration (signal intensity) changes within cells in different time points. This may apply to the scRNA-Seq reads of fluorescent-labelled gene-expression profiles across different cell sample distributions. The time-lapse imaging (and dynamics of gene expression matrices) can then be mapped with domain adaptation, which trains a neural network for residual mapping between time-points. For instance, cycle GANs (general adversarial networks) are machine-learning algorithms using similar methods on unlabelled data sets, based on image-pixel differences or gene expression residual classification. A pertinent example is GibbsNet, an iterative adversarial network used for image-to-image translation inferences in complex datasets such as traffic flows (Lamb et al., 2017). Saliency maps are also used by Deep Learning networks to accomplish similar feats in image pattern recognition.

Object edge detection and tracking algorithms can classify multiple arrays of fluorescence-labelled gene expression or protein flows in time lapse imaging of cells. This is a counting problem. Image density maps can be predicted using supervised learning methods optimizing the loss based on the MESA-distance (Maximum Excess over SubArrays). Trajectories are optimally learned by applying a Gaussian peak to label the centroid of cells and predict the flow/location of cells by mapping the center from frame to frame (Lempitsky and Zisserman, 2010). Regression networks with convolutional redundant counting are currently used to tackle counting problems in complex dynamical systems such as traffic flows, crowds, and cells. Generic Matching Network (GMN) architectures can count any objects in a class of complex dynamical systems, including the tracking of cancer cells (Cohen et al., 2017; Lu et al., 2018). Hence, such simple machine-learning algorithms can be used to track intracellular protein patterning to test for the presence of strange attractors in cell fate transitions. While it may quasi-map differential gene-expression patterns within cancer stem cells, the high dimensionality of gene expression datasets indicates Deep Learning networks (DNN) are better suited.

Moreover, the optimization algorithms (e.g. Bellman-Ford, Dijkstra's, fast marching) discussed in the graph-theoretic flow networks herein, can be replaced by Hamilton-Jacobi-Bellman (HJB) equations...
pertaining to fluid models (Farsikov et al., 2000). The solution of the Hamilton-Jacobi-Bellman equation is a partial differential equation that gives the optimal (minimum) cost flow for a dynamical decision problem. HJB is a necessary and sufficient condition to find optimal time paths (i.e. finding local minima/attractors) of control variables in feedback loop systems (i.e. GRN). It is analogous to the Hamilton-Jacobi equation with the energy term minimized with respect to a weight or control parameter \( u(t) \), given as:

\[
\frac{\partial V(q, t)}{\partial t} + \min_u H(q, p, t; u) = 0
\]

where, \( p = \nabla V, \) \( H \) is the analogous Hamiltonian (energy), and \( q(t) \) is the state vector of the system. This is a complexity equation since \( \min () \) adaptations are often (currently) intractable. Optimal control problems are generally nonlinear and without analytic solution, hence, falling under the P vs. NP problem (i.e. cannot be resolved in reasonable time). Therefore, BHJ is often handled with stochastic methods as seen with the cost flow optimization algorithms currently assessing cell fate decision making. Machine learning algorithms treat fluid dynamics problems as regression or flow optimization control tasks (Brunton et al., 2019). Hence, there are fundamental limitations imposed by the machine learning algorithms discussed for cell lineage networks reconstruction in terms of modelling complex fluid transitions.

Furthermore, K-complexity (i.e. Kolmogorov-Chaitin/algorithmic complexity) is a robust measure of network complexity vastly unutilized in current cell lineage reconstruction approaches. As the implications of quantum computing remain at infancy, classical algorithms of network complexity alone are discussed herein. While the uncomputability of the shortest program \( P \) for the output string \( s \) (matrix) of a Turing machine \( T \) remains a central problem in Algorithmic Information Theory, the question of interest herein (i.e. gene expression network reconstruction) lies within the following NP-hard, inverse problem. Is there a program which takes the string/matrix \( s \) as an input and produces the K-complexity \( K_T \) of the network as output? (Toscano et al., 2014). Note, a network is essentially a set of strings (graphs) (Zenil et al., 2018).

\[ K_T = \min \{|P|, T(P) = s\} \]

The expected probability that a random program \( P \) running on the universal Turing machine with binary data inputs produces the string \( s \) upon halting is given by the Solomonoff-Levin algorithmic probability measure: \( P(s) = \sum_{P: T(P) = s} \frac{1}{2^{|P|}} \). As seen, this measure becomes non-computable with increasing number of possible programs and lack of imposed restrictions on the length of program. Numerical approximation methods exist for indirectly measuring the K-complexity, namely, using the Algorithmic Coding Theorem, which states \( K_T \approx -\log_2 P(s) \). Block decomposition methods, compression algorithms and dimensional algorithms are tools to compute the algorithmic probability and the information dynamics of larger datasets such as biological networks (Toscano et al., 2014; Zenil et al., 2018). Hence, information-theoretic measures of K-complexity and algorithmic graph theory are emergent tools to be explored in the network science of single-cell, time series datasets.

Most importantly, algorithms pertaining to time-series analysis are the critical determinants for distinguishing stochasticity/noise, quasi-periodicity and deterministic chaos in the dynamical system. As mentioned, the lack of time-series datasets in cancer networks reconstruction remains a fundamental roadblock in oncology. Key examples of such algorithms include Kolmogorov-Sinai (metric) entropy (i.e. measure of information production rate and unpredictability), Lyapunov exponents (measure of
divergence/closeness of trajectories), frequency/power spectra, auto-correlation functions and regularity measures of the corresponding phase space portraits (e.g. multi-fractality). Such measures can be applied on the time-lag embedding of the datasets (Taken’s theorem) and should be assessed along with K-complexity approximations to infer causal information dynamics in cancer networks reconstruction. As such, the pool of algorithms for estimating chaotic emergence is immensely unexplored in biological networks reconstruction. For example, the Kolmogorov-Sinai entropy of a time-series dataset can be approximated using redundancy methods, Cohen-Procaccia algorithm, etc. (Palus, 1997).

In proposition, cancer network reconstruction algorithms can be improved by adopting fluid dynamical grid schemes. CFD algorithms such as the finite elements method (FEM) and Lattice Boltzmann method (LBM), compatible with neural networks remain unexplored in biological networks reconstruction (Takeuchi and Kosugi, 1994; Chen and Doolen, 1998; Hennigh, 2017). For instance, Boltzmann machines, a type of stochastic neural networks are tools used in modelling complex fluid dynamics. Cellular automata have been exploited as computational systems for lattice fluid dynamics (Wolfram, 1986).

Recurrent neural networks can accurately map the delay-coordinate embedding signals and phase-portraits of oscillatory, chaotic systems (Cestnik and Abel, 2019). While time-delay embedding is effective for the time-series predictions of low-dimensional systems, Reservoir computing (RC) is a machine-learning algorithm that trains recurrent neural networks to find the Lyapunov exponents of high-dimensional dynamical systems. Reservoir computing permits the chaotic attractor reconstruction in complex fluid flows (Nakai and Saiki, 2018). For instance, the reservoir computer well-predicted the short-term time-series forecasting of the Kuramoto-Sivashinsky (KS) equation to several multiples of Lyapunov time. The KS equation is a chaotic system whose pattern formation closely resembles that of fluid turbulence (Pathak et al., 2018a,b).

Deep learning architectures can classify-predict irregular patterns observed in complex dynamical systems, such as turbulent fluid flows (Kutz et al., 2017). Even simple Hopfield networks can map strange hidden attractors. Using an activation function of tanh(x), with as little as three or four neurons, strange attractors emerged in the Hopfield network’s Lyapunov spectra (Yang and Huang, 2016). However, Deep Neural Networks (DNNs) can learn complex spatio-temporal patterns between large datasets of raw data inputs. Due to the multi-layered architecture, DNNs capture the complex interactions seen in multi-scaled, complex systems. Ling et al. (2015, 2016) first-demonstrated the applicability of DNNs to predict the turbulent flows of the Reynolds-averaged Navier-Stokes equations. A Galilean-invariance embedded, DNN network architecture (Tensor Basis Neural Network) underwent training on various turbulent flow datasets followed by the Bayesian optimization for the neural network’s hyper-parameters (i.e. the number of hidden layers, the number of nodes per hidden layer, and the gradient descent algorithm’s learning rate). The DNN showed accurate predictions of Reynolds stress anisotropy in newly presented datasets. As seen, scDeepCluster outperforms other currently existent machine learning algorithms in biological networks reconstruction. Hence, in principle, Deep learning algorithms and delay coordinate embedding algorithms trained for mapping complex fluid flows can be optimized for detecting strange attractors within cancer gene expression datasets.

CONCLUSION

To conclude, a Waddington landscape can be reconstructed from any tumor sample’s single-cell gene expression datasets. Amidst all currently employed machine learning algorithms, Deep Learning
architectures (DNNs) are the state-of-the-art to accurately map the distinct phenotype clusters and its single-cell fate trajectories. DNNs are emerging as tools for modelling complex dynamical systems, such as turbulent fluid flows, wherein the physics are not well understood. Paradoxically, recent findings suggest the role of ‘chemical turbulence’ in cellular pattern formation and the emergence of chaotic attractors in cancer morphogenesis. Hence, in principle, DNNs can be trained to recognize complex fluid flows during cancer pattern formation and thereby, capture strange attractors emerging in cancer cell fate decision-making. For instance, double cell-state reporters can map cell fate switches observed in cancer stem cells (e.g. EMT states), chemically-induced stem cell reprogramming or exosome-mediated phenotype transitions in time-lapse videomicroscopy or flow cytometry. The discussed attractor reconstruction methods and time-series analyses of such reporters’ expression in correspondence to the underlying gene/protein expression profiles such as scRNA-Seq, may reveal the presence of strange attractors and their associated signaling pathways. Delay-embedding algorithms such as reservoir computing and computational fluid models are further proposed as alternatives to detect strange attractors in single-cell, time-series datasets. Hence, a better understanding of the fluid dynamics governing cell fate transitions and the adoption of the discussed AI algorithms, will pave the future of reconstructing cancer interactomes and forecasting its gene expression dynamics.

DECLARATIONS: I, Abicumaran Uthamacumaran, am the sole author of the presented paper.

ACKNOWLEDGEMENTS: Thanks to Dr. Mario D’Amico (Concordia University), Dr. Mohan Malleshaiah (UdeM) and Dr. Rolando Del Maestro, Dr. Arena Goffredo, Dr. Mohamed Abdouh, and Dr. Lorenzo Ferri of McGill University for granting me the knowledge and tools discussed herein.

REFERENCES


14) Li, Q. et al. Dynamics inside the cancer cell attractor reveal cell heterogeneity, limits of stability, and escape. PNAS. 113 (10) 2672-2677 (2016)


20) Font-Clos, F. Topography of epithelial-mesenchymal plasticity. PNAS. 1-6 (2018)


67) Yuan, Jinzhou et al. “Single-cell transcriptome analysis of lineage diversity in high-grade glioma.” Genome medicine vol. 10,1 57. 2018
84) Shin, H. et al, Correlation between Cancerous Exosomes and Protein Markers Based on Surface-Enhanced Raman Spectroscopy (SERS) and Principal Component Analysis (PCA). ACS Sens., 2018, 3 (12), pp 2637–2643
95) BossoMaier, T.R.J. and Green, D.J. Complex Systems. Cambridge Univ. Press, 2000


James, M. et al., Turbulence and turbulent pattern formation in a minimal model for active fluids, Phys Rev. Fluids 3, 061101(Rapid Communication) (2018)

James, M. and Wilczek, M., Vortex dynamics and Lagrangian statistics in a model for active turbulence, EPJE 41, 21 (2018)


