Multiple trajectory alignment reconstructs disease dynamics for discovery and clinical benefit

Amit Frishberg[1,2,5], Neta Milman[1], Ayelet Alpert[1], Joachim L. Schultze[2,3,4], Fabian J. Theis[5,6,7], Shai S. Shen-Orr[1,*]

[1] Department of Immunology, Faculty of Medicine, Technion-Israel Institute of Technology, Haifa, Israel

[2] Systems Medicine, Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE), Bonn, Germany

[3] Genomics and Immunoregulation, Life & Medical Sciences (LIMES) Institute, University of Bonn, Bonn, Germany

[4] Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE). PRECISE Platform for Genomics and Epigenomics at DZNE and University of Bonn, Germany

[5] Institute of Computational Biology, Helmholtz Center Munich, 85764, Neuherberg, Germany

[6] Department of Mathematics, Technical University of Munich, 85748, Garching, Germany

[7] Technical University of Munich, TUM School of Life Sciences Weihenstephan, 85354, Freising, Germany

* Corresponding author. Email: shenorr@technion.ac.il

Abstract

Biological processes are changes occurring concomitantly over time. Therefore, any comparison ignorant of temporal changes is greatly deficient. Even in longitudinal experimental designs, data is usually collected at fixed intervals, whereas process dynamics vary between individuals. We developed TimeAx, which brings time into the equation by building a comparative framework for capturing biological processes dynamics from 'omic data'. We used TimeAx to study influenza infection dynamics and Urothelial bladder cancer tumorigenesis, discovering molecular mechanisms that drive disease progression as well as promoting clinical symptoms, for short and long term biological processes. Specifically, we detected an "inflection point", where the tumor progresses into an advanced, pro-metastatic state. Overall, we present a powerful framework for studying high resolution temporal dynamics, providing improved molecular interoperability and clinical benefit.

Introduction

Biological processes are dynamic, therefore changing over time. From intra-cellular molecular state changes which occur following cellular differentiation and activation to multiple molecular, cellular and physiological changes during disease progression and even in health. Identifying the underlying dynamics of biological processes at high-resolution and enabling their quantitative comparison can provide us the tools to alter them to improve health (**Figure 1A**). A striking example of this radiocarbon (carbon-14) dating to archaeology, which introduced a quantitative framework to model time, revolutionizing the entire field by bringing resolution and clarity of the temporal dimension where it was murky at best prior [1]. More recently, in biology, the heterogeneity of single-cell mRNA and protein data has been broadly used for trajectory inference that describes and compares the dynamics of a biological process occurring at the cell, but not organismal, level [2,3].

Time-series experimental designs provide an opportunity for studying biological process dynamics. However, while disease progression is patient-specific, time-series data is usually collected at fixed intervals, thus reducing the efficiency of comparing progression dynamics, as linear models fail to capture inter-individual variance, and forcing the need for clustering of many time points together in order to observe some extent of shared dynamics. Dimensionality reduction methods, such as Principal Component Analysis (PCA), often used to discover this hidden variance, present low interpretability, as the inferred dynamics initiation point is usually unknown, and the calculation of progression states, based on a multi-dimensional space, may be driven by artifacts [4]. Therefore, even when large time-series data are available, forming a robust model for disease progression is challenging.

Here, we present TimeAx, a method that models disease dynamics over time based on time-series data from multiple individuals, thus tracing the dynamics and progression rates associated with molecular mechanisms of the disease and patients' clinical outcomes. Similar to computational multiple sequence alignment approaches, which provide biologists the capability to understand DNA and protein function through a quantitative comparison [5], TimeAx performs multiple alignment of shared dynamics across a common trajectory, thus allowing the modeling of any type of continuous process and its quantitative comparison (**Figure 1B).** We applied TimeAx to model the disease progression of both short- and long- term diseases, revealing novel molecular mechanisms and providing major clinical benefits.

Results

Overview of TimeAx

To model the dynamics of a biological process, TimeAx relies solely on an input of longitudinal sampling from multiple subjects. Subjects may have 3 or more time points from which samples were collected and assayed by 'omic-level' (i.e., multiplex) feature measurement (here, bulk gene expression profiles). Subject time points may differ in number and in collection time. Broadly, the TimeAx process consists of three steps (**Figure S1A**, See **Methods**): First, a feature selection step, in which TimeAx selects a set of features (*"conserved-dynamics-seed")* as the backbone of the trajectory of the biological process we aim to infer. Second, TimeAx builds a consensus trajectory, which approximates the dynamic process the subjects being studied are undergoing (**Figure 1B**). Finally, to identify the biological state a particular individual at a particular time point is in, in relation to the entirety of the dynamics of the biological process being studied, TimeAx leverages the inferred set of consensus trajectories to predict a set of positions, which are then averaged to infer the final position on the consensus trajectory, referred to as 'pseudotime'. In addition, TimeAx provides a robustness measure for its inferred consensus trajectory, to assess the accuracies in which they captured patient dynamics across time (See **Methods**).

To test TimeAx's ability to capture the dynamics of a disease, we first performed a simulation study. We simulated bulk gene expression measurements from different individuals sampled at multiple time points and undergoing a common biological process with different dynamics (see **Methods**). In this scenario the pseudotime position on the consensus trajectory of each simulated sample is known. Next, we tested the correlation between simulated pseudotime (i.e. gold standard) and the simulated time points and TimeAx predicted pseudotime. TimeAx predicted pseudotime highly overlapped with the simulated pseudotime within each individual, across different levels of added noise, while the simulated time points failed capturing the heterogeneity between samples (**Figure 1C,** see **Methods**). One important result of TimeAx alignment is an improved ability to discover genes whose expression is closely related to disease progression dynamics (See **Methods**; **Figure 1D**). The accuracy of pseudotime inference

increased when more samples were available per individual (**Figure S1B**). Finally, the number of consensus trajectories had only a marginal effect on TimeAx's efficiency, as five consensus trajectories were often sufficient (**Figure S1C**).

TimeAx modeling of acute disease reveals molecular mechanisms and clinical benefits

To showcase the benefit of a quantitative framework to understand biological process dynamics, we studied the chronological progression of influenza infection, an acute disease which majorly affects the immune-system over short periods of time. We applied TimeAx on an influenza challenge study (H3N2/Wisconsin strain) in which 17 healthy adults were infected and then profiled longitudinally for whole blood gene expression by RNA-seq at 13-15 distinct time points within the first 108 hours following infection (total of 268 samples) [6] (**Figure 2A**). Original analysis of this gene expression data discovered a clinically relevant gene module and presented different gene expression dynamics over time, without accounting for inter-patient variation. Based on this data, TimeAx composed a model for whole blood influenza infection dynamics over time, inferring fine-tuned pseudotime positions for all samples (robustness score of R = 0.99; **Figure S2A**; conserved-dynamics-seed genes listed in **Table S1**), with only partial resemblance to sampling times, predominantly collected at fixed intervals for all individuals (R = 0.65; **Figure S2B,C**). Moreover, we observed different disease progression patterns over time across patients, suggesting utility for unmasking signals by harmonizing individuals via trajectory alignment (**Figure 2B**). We then calculated the association between pseudotime and gene expression (i.e., by Pearson correlation), and found it to be stronger compared to a model based on sampling times (**Figure 2C** and **Table S2**), allowing the identification of molecular processes changing over time that would otherwise be missed (**Figure 2D;** See **Methods**). Specifically, we detected 1,888 significantly associated genes (16%) solely based on pseudotime and 860 genes (7%) as detectable by both pseudotime and a sampling times based model, but only 113 genes (1%) solely based on the sampling time model (significance threshold of $p<10^{-5}$; **Figure 2E**). Taken together, these results suggest that TimeAx dynamics modeling enables molecular interpretability at a higher resolution.

We postulated that higher pseudotime positions, as inferred by TimeAx, are related to host response magnitude. Two lines of evidence support this hypothesis: At the molecular level, the conserved-dynamics-seed genes included interferon induced genes, such as RTP4, IFI27 and PSME2, as well as heme metabolism, such as ALAS2, AHSP, HBD, BPGM and BLVRA (**Table S1**). Similarly, genes with positive associations with the pseudotime were highly enriched for the interferon and heme metabolism pathways (**Figure S2D** and **Table S3**), with less prominent associations for sampling time (**Figure 2F**), supporting previous findings emphasizing these pathways as related to augmented immune responses, with strong application on disease

severity and patients clinical outcomes, during influenza infections [7–10]. Clinically, while all patients were challenged with influenza, only some of them developed clinical symptoms, while others were asymptomatic. Pseudotime demonstrated strong ties with the appearance of symptoms, as asymptomatic patients had significantly lower pseudotime position compared to symptomatic ones ($p<10^{-14}$; **Figure 2G**).

To validate our results, we used the TimeAx generated model to predict pseudotime for samples from an additional challenge study of an influenza infection (H1N1 strain) in 24 human adults, sampled for whole blood RNA-seq at 13-15 distinct time points within the first 108 hours after the infection (total of 382 samples, denoted as 'Cohort 2') [11] (**Figure 2A**). Similarly, pseudotime in Cohort 2 displayed a strong association with the appearance of symptoms ($p<10^{-16}$; **Figure 2H**). Last, we checked the validity of the model to infer pseudotime in a different population and measurement platform. Specifically, we used our adult-generated pseudotime model to predict the pseudotime position of influenza infected and healthy control children (n=19, 33 respectively) from whole blood microarray [12] (**Figure 2A**, denoted as 'Cohort 3', see **Methods**). Influenza infected children had significantly higher pseudotime positions compared to controls, suggesting the general utility of the model ($p<10^{-5}$; **Figure 2I**). Taken together, TimeAx inference of acute disease progression dynamics yielded a clinically useful robust model that yields novel molecular mechanisms in influenza infection.

<u>TimeAx allows assessment of cancer outcomes by modeling tumor dynamics over years</u>

Biological processes often advance slowly, making it difficult to track their dynamics and understand their molecular drivers. This often leads to the need for clustering patient samples into disease subtypes, losing the continuous aspect of disease dynamics modeling [13,14]. To test TimeAx's ability in capturing long temporal processes, and in particular disease progression, we modeled Urothelial bladder cancer (UBC), a tumor with high recurrence rates, that initially appears in a non-muscle invasive state with higher probability of transforming into its muscle-invasive state with each recurrence, increasing the risk to develop metastasis [15]. We trained TimeAX using time series microarray data from 18 patients with recurring non-muscle invasive bladder cancer, sampled longitudinally during the incidence of tumor recurrence (4-6 samples per patient, collected up to 15 years apart from first to last recurrence) [13] (denoted as Cohort 4; **Figure 3A,** see **Methods**). Original analysis of this dataset predominantly identified tumor subtypes and discovered an increase in the expression of mesenchymal cells markers between late and early recurrences [13].

TimeAx composed a model for UBC dynamics over time, inferring pseudotime positions for all samples (robustness score of R = 0.89; **Figure S3A**; conserved-dynamics-seed genes listed in **Table S4**), with only partial resemblance to tumor recurrence times (R = 0.67; **Figure S3B,C**). Similar to

the TimeAx model of influenza infection, we observed different disease progression patterns over time across patients (**Figure 3B**), discovering stronger associations between pseudotime and gene expression (i.e., by Pearson correlation) compared to tumor recurrence times (**Figure 3C** and **Table S5**). Specifically, 1962 genes (15%) were significantly associated with pseudotime, while only 1 gene (TMEM55B) was significantly associated with a model based on tumor recurrence times (significance threshold of $p<0.05$; **Figure 3D**). This signal enhancement was also observed at the pathway level where a TimeAx based analysis identified stronger associations for known cancer-related processes such as EMT ($q=6\cdot10^{-49}$) [16,17], and novel ones such as TNFa signaling, interferon gamma and G2M checkpoint ($q=6\cdot10^{-26}$, 10^{-19}, $3\cdot10^{-10}$, respectively) (**Figure 3E** and **Table S6,** see **Methods**). Importantly, the TimeAx inferred pseudotime was associated with known clinical markers of disease progression such as CCL2 and IFITM2 and conversely negatively associated with SGPL1, a marker that was shown to be related to positive outcomes [18–20] (**Figure S3D**). Interestingly, classifying samples based on the known UBC tumors clinical staging [14] showed high variation in pseudotime positions for each clinical subtype (**Figure S3E)**. Yet, higher pseudotime was significantly associated with advanced tumor staging ($p=0.03$; **Figure 3F**). To validate this, we used our TimeAx generated model to predict pseudotime for samples from two additional cohorts: of 276 UBC patients, each sampled once and assayed by microarray [21] and 430 UBC RNA-seq samples from the Cancer Genome Atlas (TCGA) Program (denoted as Cohort 5 and 6 respectively) (**Figure 3A**). Predicting each subject's pseudotime position, we observed similar gene and tumor stage associations (**Figure S3F, Figure 3G, Table S5**), emphasizing the pseudotime as a molecular and clinical axis of tumor progression. As high variation in pseudotime was observed within each tumor stage (**Figure 3F,G**), we next leveraged cohort 6 to study the associations of tumor stage versus pseudotime with the clinical outcome of the patients. Whereas tumor staging was only marginally associated with patient death ($p=0.03$; **Figure 3H, left**), TimeAx pseudotime presented improved accuracy in discriminating between final outcomes ($p=6\cdot10^{-5}$; **Figure 3H, right**). Thus, TimeAx can capture long-term dynamics for improved discovery and clinical application purposes.

<u>TimeAx identifies an inflection point in tumor progression promoting a pro-metastatic state</u>

The TimeAx model, based on whole transcriptomes, represents a continuous shift in a complex multicellular program in which both cell compositions and cell-states may alter along the positions of the pseudotime. Specifically, low tumour purity, a measure for the fraction of cancer cells in a tumor sample, was found to be related to increase in immune infiltration and overall poor prognosis [22,23]. Indeed, we identified a position along the pseudotime where a sharp decrease in tumor purity occured, a trend undetectable when ordering samples by recurrence ($r = -0.54$ and $r=-0.1$, respectively; **Figure 4A**, see **Methods**), showcasing the benefit of TimeAx

modeling and suggesting the existence of an inflection point, where a change in the tumor occurs, affecting the immune-stroma composition at the tumor site [23,24].

To understand how cellular composition and regulatory programs changes along pseudotime and how these relate to the decrease in tumor purity at the inflection point (**Figure 4C**), we deconvolved the cell composition of all samples in the three cohort, predicting the compositions of seven major cell types: urothelial cells, T cells, muscle cells, basal tumor cells, endothelial cells, macrophages and fibroblasts, as well as different immune cell subtypes (See **Methods**). We found cell composition changes associated with high pseudotime positions, located post-inflection point. Specifically, we observed a significant decrease in the abundance of urothelial cells and naive CD4+ T cells and an increase in the abundance of activated macrophages and fibroblasts ($p<0.006$; **Figure 4B** and **Figure S4A**), with a noticeable increase in CD4+ memory T-cells ($p<0.1$; **Figure 4B** and **Figure S4A**) but no significant change in basal tumor cells ($p>0.1$; **Figure S4A,B**), supporting previous findings regarding the association of the infiltration of activated immune cells and cancer associated fibroblasts into the tumor's stroma with tumor progression and a poor clinical outcome [25,26]. Next, we contrasted gene abundance and co-expression pre and post inflection point. We identified four main co-regulatory gene clusters, whose co-regulation was broadly disrupted post-inflection point, as well as reversal of gene associations with the pseudotime ($r=-0.62$**,** by Pearson's correlation, slope of changes pre vs post; **Figure S4C** and **Table S7**), suggesting a major switch in regulation during tumor progression dynamics that can be investigated via TimeAx.

To pinpoint the biological processes driving this major shift in tumor progression dynamics we characterized pathway behavior in samples whose pseudotime positions were ascribed pre versus post inflection point (**Figure 4C,** see **Methods**). We observed strong positive enrichments for cell cycle and DNA repair pathways ($q=8\cdot10^{-16}$ and $9\cdot10^{-6}$, respectively), including an increase in the pathways' genes expression along the pseudotime, pre-inflection point, supporting the notion of uncontrolled cell replication and higher mutation rates as hallmarks of cancer progression [15,27] (**Figure 4E** and **Table S8**). In contrast, ECM activity was only positively associated with pseudotime post-inflection point ($q=10^{-21}$; **Figure 4E** and **Table S8**), supporting previous findings pointing out this pathway's role in cancer migration and metastasis, and suggesting the inflection point as the pseudotime state of metastasis initiation, as tumors tend to remodel their surrounding matrix to promote a microenvironment supporting metastasis [28]. Lastly, we found positive enrichment of epithelial mesenchymal transition (EMT) both pre and post inflection point ($q=0.003$ and $9\cdot10^{-17}$, respectively; **Figure 4E** and **Table S8**), suggesting that EMT is a continuous process starting early in the tumor and increasing along the pseudotime.

Surprisingly, we also noted a striking positive enrichment post-inflection point for neuronal differentiation and neuronal signaling pathways (e.g., Synaptic membrane; **Figure 4E** and **Table S8**), stemming from genes with positive associations with the pseudotime, encoding for receptors

for small molecule neurotransmitters, their downstream G coupled receptors and ion channels, which were spread across different co-regulatory clusters, suggesting a global phenomena of tumor progression (**Figure 4F**). Specifically, we discovered associations for genes encoding for subunits of purinergic receptors (ADORA1, P2RX6), biogenic amines receptors (ADRA2C, ADRB1, HTR1B), acetylcholine receptors (CHRNA2, CHRNA4, CHRNB2), amino acids receptors (GABRB3, GABRD, GABRR2, GRIK3, GRIK4, GRIN1, GRIN2C, GRIN3B, GRM2, GRM5) and opioid receptors (SIGMAR1). Tumors are often innervated by nerve fibers that release factors promoting cancer cells proliferation and migration [29,30]. Yet, an alternative explanation may be that neuro-developmental programs are activated as part of tumorigenesis, as has recently been described for small-cell lung cancer, where cellular protrusions that resemble axons allow the tumor to gain the ability to metastasize [31]. Taken together, the TimeAx deduced model suggests the inflection point as the state where the tumor proceeds into an advanced pro-metastatic phase, which is accompanied by a sharp decrease in tumor purity, due to rapid immune-cell and cancer associated fibroblast infiltration and a loss of regulation over cell cycle, DNA repair and ECM remodeling mechanisms, with the elevation of the last, and a strong activation of neuronal programs, which may mark the high metastatic potential of the tumor (**Figure 4G**).

Discussion

We present TimeAx, a robust framework for biological processes reconstruction and quantitative comparison from omic-level data. TimeAx models biological processes over time, revealing the shared dynamics, which was initially hidden due to population heterogeneity. This allows the inference of pseudotime positions, quantitative units that represent the samples' true biological states, which can be rigorously compared and used to build a high-resolution understanding of the temporal process. We present TimeAx's utility for inference and downstream analysis of both short and long term biological process dynamics. The assembled consensus trajectories allowed high-resolution time-dependent characterization of system state which resulted in the discovery of novel disease mechanisms and an improved clinical outcome prediction.

The TimeAx framework can be further extended. Specifically, at present, TimeAx models dynamics as a one dimensional consensus trajectory, whereas biological processes may display more complex dynamics (such as forking). In addition, an extension of TimeAx would allow the inference of biological process dynamics based on data from less than 3 time points per individual. TimeAx can be also used to model other types of dynamics, other than disease progression as presented here, such as disease recovery over time and non-disease biological processes, such as immune age [32]. In addition, while exemplified for gene expression data, TimeAx can be applied to other molecular data types, such as protein and epigenetics, as well as non-molecular data.

We present a first-of-its kind powerful framework for studying time-dependent biological processes at high resolution. TimeAx modeling frees researchers from relying on chronological time in experimental designs and analyses. Rather it makes the passing of biological time a comparable unit that researchers can discuss quantitatively. Akin to the insight introduced by sequence alignment, and single cell cell-state trajectory inference, TimeAx opens the door to high resolution comparative understanding of organismal biological processes dynamics.

Acknowledgements

We thank Y. Abraham for help with designing and creating figure illustrations. The results here are in part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga. This research was supported by the ISRAEL SCIENCE FOUNDATION (grant No. 1626/20), within the Israel Precision Medicine Partnership program. This work was supported in part by the German Research Foundation (DFG) to J.L.S. under Germany's Excellence Strategy (DFG) – EXC2151 – 390873048); the HGF grant sparse2big, the EU H2020 projects SYSCID (Grant Agreement No. 733100) and ImmunoSep (Grant Agreement No. 847422). J.L.S was further supported by the BMBF-funded excellence project Diet–Body–Brain (DietBB) (grant number 01EA1809A), iTREAT (FKZ: 01ZX1902A), and by NaFoUniMedCovid19" (FKZ: 01KX2021, project acronym „COVIM"). This study was funded in part by the European Union's Horizon 2020 Research and Innovation Program under the ERA-Net Cofund action no. 727565; the Joint Programming Initiative, A Healthy Diet for a Healthy Life (JPI-HDHL; project 529051018) and under the ERA-CVD non-cofunded action JTC2017 (Mechanisms of early atherosclerosis and/or plaque instability in Coronary Artery Disease)) awarded to J.L.S.

Author contributions

SSO conceived the idea, AF and SSO developed the TimeAx method, AA and FJT contributed to method development, AF performed the analysis. NM and JLS helped with biological interpretation. AF and SSO wrote the manuscript and all authors reviewed and revised it.

Competing interests

SSO holds equity and is a consultant of CytoReason. FJT reports receiving consulting fees from ImmunAI and ownership interest in Dermagnostix.

Code availability

TimeAx is publically available as an R package at Github: https://github.com/amitfrish/TimeAx.

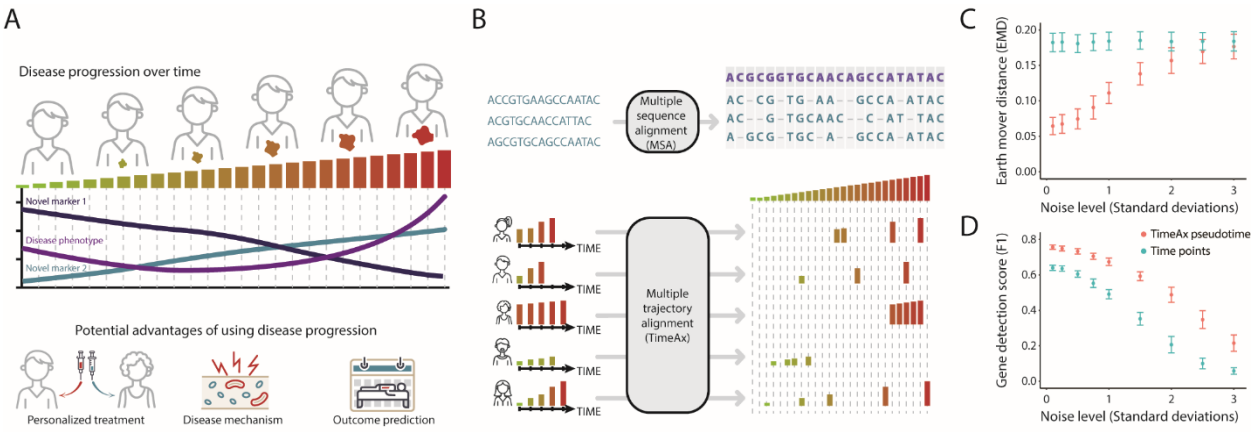

Figure 1: Overview of TimeAx. A. An illustration of the power of disease dynamics in research and clinical applications. **B.** An illustration presenting the similarity between TimeAx trajectory alignment and multiple sequence alignment (MSA). **C-D.** Simulated data analysis. Superior results for TimeAx predicted pseudotime (red), compared to simulated time points (blue), across different levels of noise (*x*-axis), calculating the EMD distance to the simulated psedutime **(C)** and the discovery rate of progression-related genes **(D)** (*y*-axis; See **Methods**). For both panels, error bars represent mean ± std.

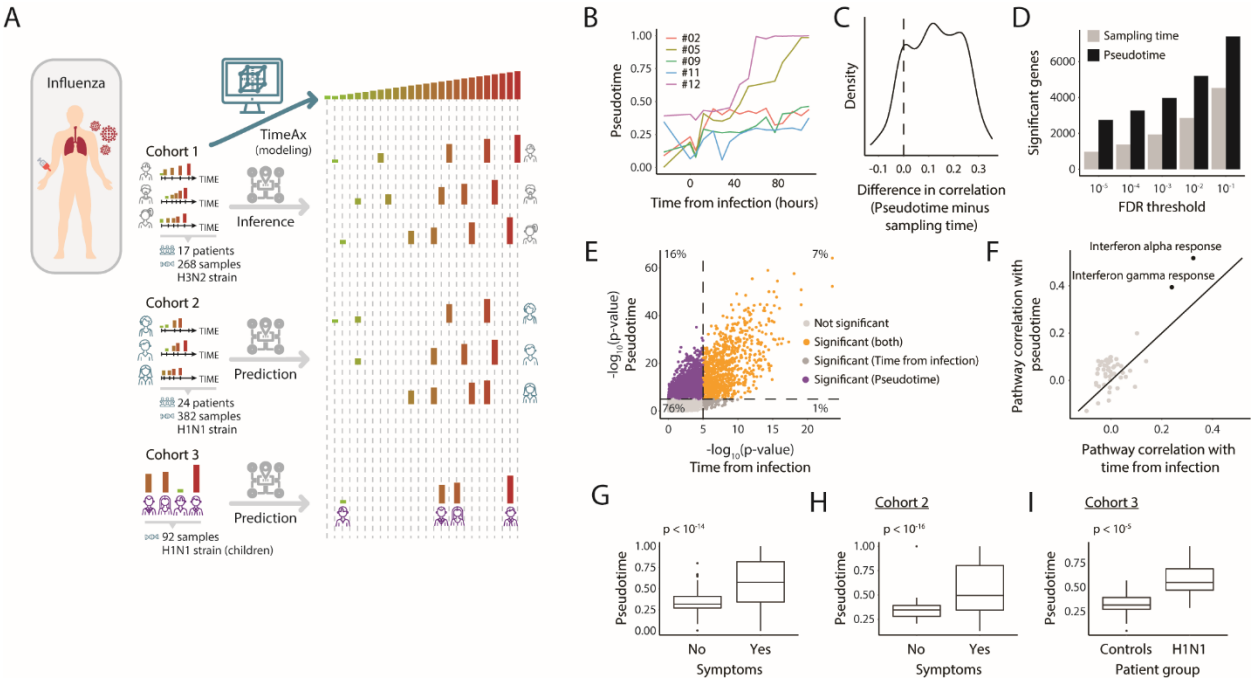

Figure 2: TimeAx modeling of acute disease reveals molecular mechanisms and clinical benefits. **A.** An illustration of the influenza infection dynamics TimeAx modeling and pseudotime inference, based on longitudinal cohort 1, as well as pseudotime predicted for validation cohorts 2 and 3. **B.** TimaAx pseudotime (*y*-axis) is different from sampling time (time from infection, *x*-axis), exemplified in five patients in cohort 1. **C.** Psuedotime demonstrates higher 'Pearson' correlation coefficients with gene expression, compared to sampling time, presented as the distribution of differences between the two measures, across all genes. **D.** Number of progression-related genes (*y*-axis), across different Q-value thresholds (*x*-axis), is higher using pseudotime (black), compared to sampling time (gray) (See **Methods**). **E.** Gene associations (-log10 transformed, FDR-corrected, Q-values) with sampling time (*x*-axis) and pseudotime (*y*-axis), using a Q-value threshold of 10^{-5} (Dashed lines) (See **Methods**). Genes are colored based on their association with the two time axes. **F.** Median gene association levels in biological pathways from MSigDB Hallmarks (gray dots), calculated with either sampling time (*x*-axis) or pseudotime (*y*-axis). Emphasized in black are pathways with improved scores for pseudotime. **G-I.** Pseudotime clinical applicability. Shown are pseudotime associations with clinical symptoms in cohorts 1 (**G**) and cohort 2 (**H**) and with the differences between infected and control patients in cohort 3 **(I)**. Boxes represent the 25th, 50th, and 75th percentiles; whiskers show maxima and minima.

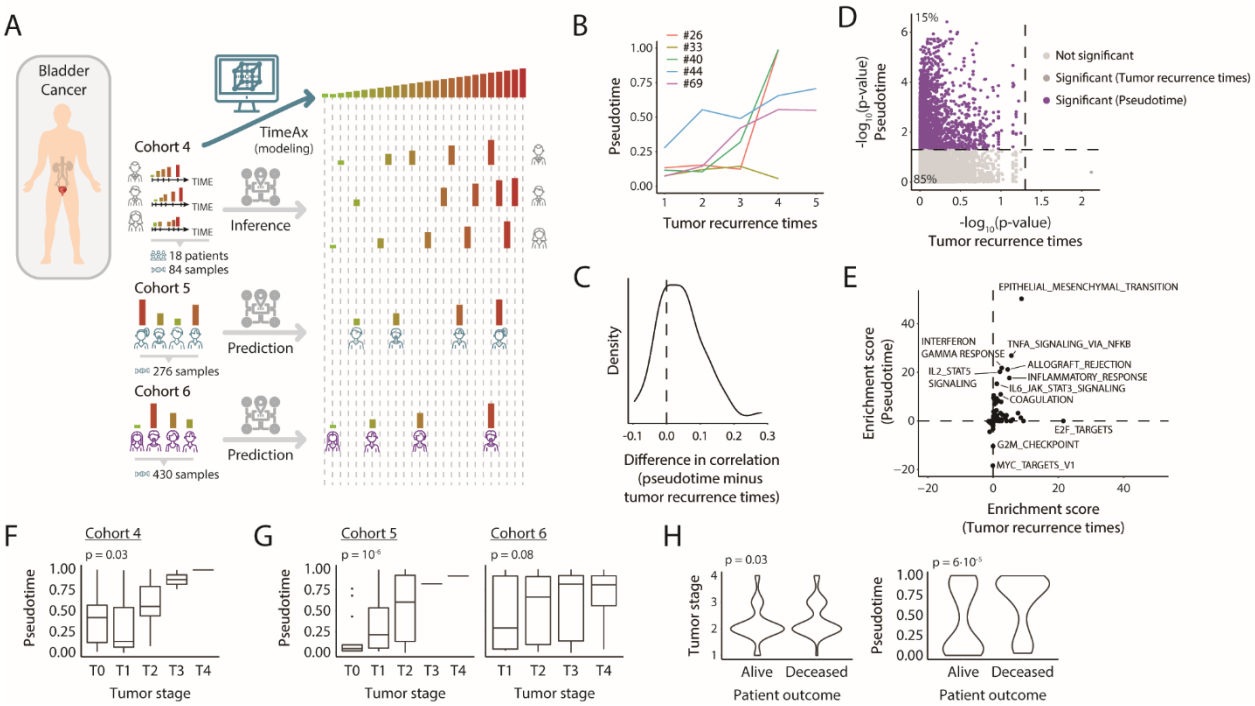

Figure 3: TimeAx allows assessment of cancer outcomes by modeling tumor dynamics over years. A. An illustration of the UBC dynamics TimeAx modeling and pseudotime inference, based on longitudinal cohort 4, as well as pseudotime predicted for validation cohorts 5 and 6. **B.** TimaAx pseudotime (*y*-axis) is different from sampling time (time from infection, *x*-axis), exemplified in five patients in cohort 4. **C.** Pseudotime demonstrates higher 'Pearson' correlation coefficients with gene expression, compared to sampling time, presented as the distribution of differences between the two measures, across all genes. **D.** Gene associations (-log10 transformed, FDR-corrected, Q-values) with sampling time (*x*-axis) and pseudotime (*y*-axis), using a Q-value threshold of 10^{-5} (Dashed lines) (See **Methods**). Genes are colored based on their association with the two time axes, displaying significant associations almost entirely only with the pseudotime. **E.** Enrichment scores of biological pathways from MSigDB Hallmarks (black dots), displaying improved scores while calculated using pseudotime (*y*-axis) compared to tumor recurrence times (*x*-axis). **F-H.** Pseudotime clinical applicability. Shown are pseudotime relation with tumor stage in cohort 4 (**F**) and cohorts 5-6 (**G**), and its improved association, over tumor stage, with patients clinical outcomes (**H**). In **F-G**, boxes represent the 25th, 50th, and 75th percentiles; whiskers show maxima and minima.

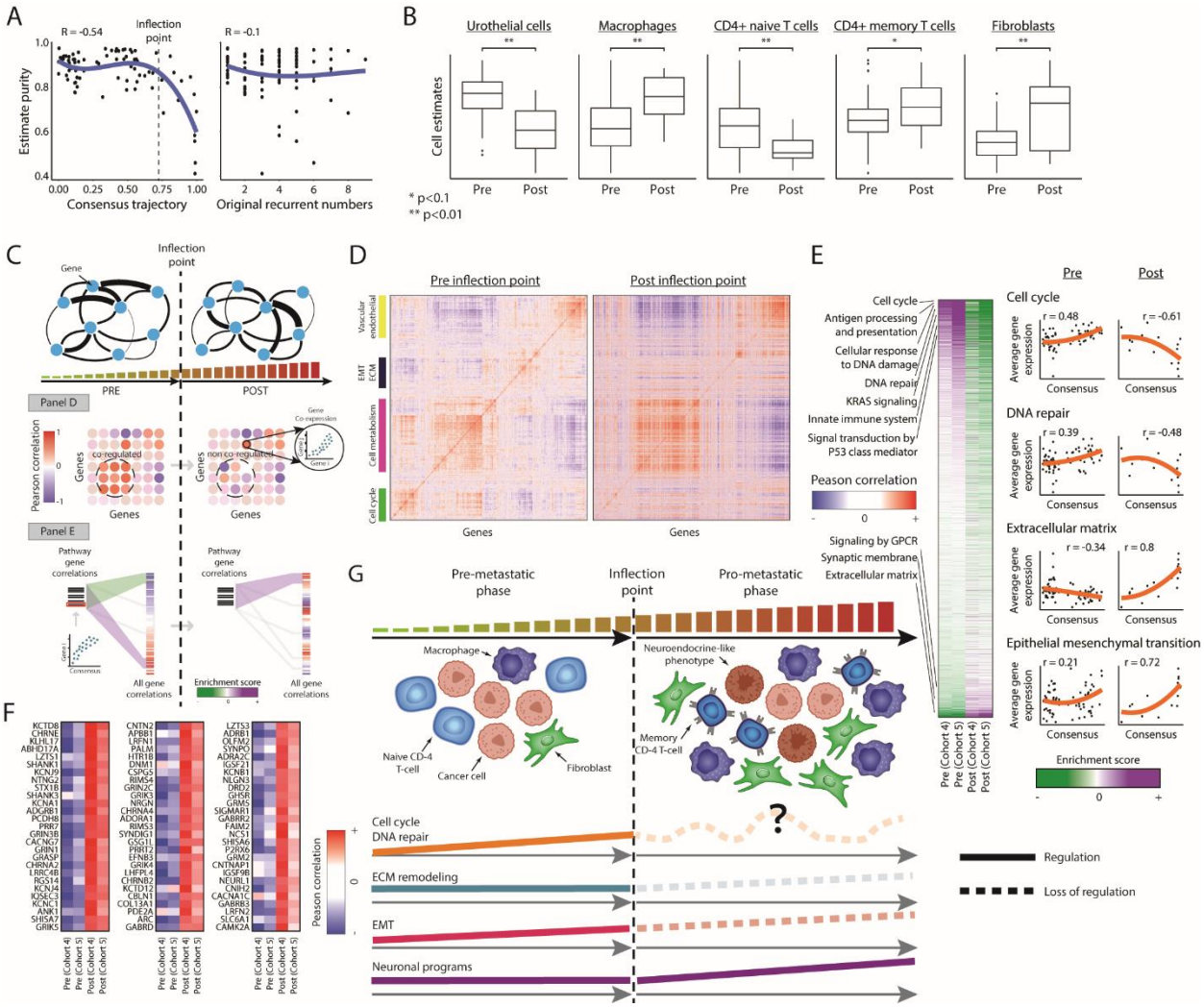

Figure 4: TimeAx identifies an inflection point in tumor progression promoting a pro-metastatic state. A. Tumor estimate purity scores (*y*-axis) along the pseudotime (left) and the tumor recurrence times (right) (*x*-axis), displaying a sharp decrease in high pseudotime positions (inflection point; dashed line). **B.** Cell type deconvolved compositions (*y*-axis), displaying major differences between pre- and post- inflection point samples. * p<0.1, ** p<0.01. **C.** An illustration of the co-expression and enrichment analysis (panel D and E, respectively), for discovering molecular regulatory changes between pre- and post- inflection point samples (See **Methods**). **D.** Gene co-expression matrix for samples pre- (left) and post- (right) inflection point. Four co-expression clusters were found pre-inflection point (demonstrated by colored lines) and were largely dysregulated post-inflection point. **E.** Enrichment scores of different biological pathways, pre- and post- inflection point, in cohorts 4 and 5. Here, negative to positive enrichments are colored in a green to purple scale. For four different pathways, results are demonstrated (scatter plots) as the average pathway genes expression (*y*-axis) along the pseudotime (*x*-axis). **F.** Strong associations for different neurotransmitter genes with the pseudotime, differing between pre- and post- inflection point samples, in cohorts 4 and 5. In **D** and **F**, negative to positive correlations

are colored in a blue to red color scale. **G.** An overview of the cellular and molecular changes during UBC progression.

References

1. Libby, W. F. Atmospheric Helium Three and Radiocarbon from Cosmic Radiation. *Phys. Rev.* **69**, 671–672 (1946).
2. Alpert, A., Moore, L. S., Dubovik, T. & Shen-Orr, S. S. Alignment of single-cell trajectories to compare cellular expression dynamics. *Nat. Methods* **15**, 267–270 (2018).
3. Tritschler, S. *et al.* Concepts and limitations for learning developmental trajectories from single cell genomics. *Development* **146**, (2019).
4. Chari, T., Banerjee, J. & Pachter, L. The Specious Art of Single-Cell Genomics. *BioRxiv* (2021) doi:10.1101/2021.08.25.457696.
5. Feng, D. F. & Doolittle, R. F. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J. Mol. Evol.* **25**, 351–360 (1987).
6. Huang, Y. *et al.* Temporal dynamics of host molecular responses differentiate symptomatic and asymptomatic influenza a infection. *PLoS Genet.* **7**, e1002234 (2011).
7. Killip, M. J., Fodor, E. & Randall, R. E. Influenza virus activation of the interferon system. *Virus Res.* **209**, 11–22 (2015).
8. Wu, W. & Metcalf, J. P. The role of type I ifns in influenza: antiviral superheroes or immunopathogenic villains? *J. Innate Immun.* **12**, 437–447 (2020).
9. Espinoza, J. A., González, P. A. & Kalergis, A. M. Modulation of Antiviral Immunity by Heme Oxygenase-1. *Am. J. Pathol.* **187**, 487–493 (2017).
10. Canesin, G., Hejazi, S. M., Swanson, K. D. & Wegiel, B. Heme-Derived Metabolic Signals Dictate Immune Responses. *Front. Immunol.* **11**, 66 (2020).
11. Woods, C. W. *et al.* A host transcriptional signature for presymptomatic detection of

infection in humans exposed to influenza H1N1 or H3N2. *PLoS ONE* **8**, e52198 (2013).

12. Herberg, J. A. *et al.* Transcriptomic profiling in childhood H1N1/09 influenza reveals reduced expression of protein synthesis genes. *J. Infect. Dis.* **208**, 1664–1668 (2013).
13. Sjödahl, G. *et al.* Molecular changes during progression from nonmuscle invasive to advanced urothelial carcinoma. *Int. J. Cancer* **146**, 2636–2647 (2020).
14. Kamoun, A. *et al.* A Consensus Molecular Classification of Muscle-invasive Bladder Cancer. *Eur. Urol.* **77**, 420–433 (2020).
15. Sanli, O. *et al.* Bladder cancer. *Nat. Rev. Dis. Primers* **3**, 17022 (2017).
16. Baumgart, E. *et al.* Identification and prognostic significance of an epithelial-mesenchymal transition expression profile in human bladder tumors. *Clin. Cancer Res.* **13**, 1685–1694 (2007).
17. Franzen, C. A. *et al.* Urothelial cells undergo epithelial-to-mesenchymal transition after exposure to muscle invasive bladder cancer exosomes. *Oncogenesis* **4**, e163 (2015).
18. Kitamura, T. *et al.* CCL2-induced chemokine cascade promotes breast cancer metastasis by enhancing retention of metastasis-associated macrophages. *J. Exp. Med.* **212**, 1043–1059 (2015).
19. Xu, L. *et al.* IGF1/IGF1R/STAT3 signaling-inducible IFITM2 promotes gastric cancer growth and metastasis. *Cancer Lett.* **393**, 76–85 (2017).
20. Nasiri, A. R., Rodrigues, M. R., Li, Z., Leitner, B. P. & Perry, R. J. SGLT2 inhibition slows tumor growth in mice by reversing hyperinsulinemia. *Cancer Metab.* **7**, 10 (2019).
21. Sjödahl, G., Eriksson, P., Liedberg, F. & Höglund, M. Molecular classification of urothelial carcinoma: global mRNA classification versus tumour-cell phenotype classification. *J.*

Pathol. **242**, 113–125 (2017).

22. Mao, Y. *et al.* Low tumor purity is associated with poor prognosis, heavy mutation burden, and intense immune phenotype in colon cancer. *Cancer Manag. Res.* **10**, 3569–3577 (2018).
23. Zhang, C. *et al.* Tumor purity as an underlying key factor in glioma. *Clin. Cancer Res.* **23**, 6279–6291 (2017).
24. Majidpoor, J. & Mortezaee, K. Steps in metastasis: an updated review. *Med. Oncol.* **38**, 3 (2021).
25. Pan, S., Zhan, Y., Chen, X., Wu, B. & Liu, B. Bladder cancer exhibiting high immune infiltration shows the lowest response rate to immune checkpoint inhibitors. *Front. Oncol.* **9**, 1101 (2019).
26. Boström, M. M. *et al.* Tumor-Associated Macrophages Provide Significant Prognostic Information in Urothelial Bladder Cancer. *PLoS ONE* **10**, e0133552 (2015).
27. Glaser, A. P., Fantini, D., Shilatifard, A., Schaeffer, E. M. & Meeks, J. J. The evolving genomic landscape of urothelial carcinoma. *Nat. Rev. Urol.* **14**, 215–229 (2017).
28. Winkler, J., Abisoye-Ogunniyan, A., Metcalf, K. J. & Werb, Z. Concepts of extracellular matrix remodelling in tumour progression and metastasis. *Nat. Commun.* **11**, 5120 (2020).
29. Cervantes-Villagrana, R. D., Albores-García, D., Cervantes-Villagrana, A. R. & García-Acevez, S. J. Tumor-induced neurogenesis and immune evasion as targets of innovative anti-cancer therapies. *Signal Transduct. Target. Ther.* **5**, 99 (2020).
30. Jung, E., Alfonso, J., Monyer, H., Wick, W. & Winkler, F. Neuronal signatures in cancer. *Int. J. Cancer* **147**, 3281–3291 (2020).

31. Yang, D. *et al.* Axon-like protrusions promote small cell lung cancer migration and metastasis. *eLife* **8**, (2019).

32. Alpert, A. *et al.* A clinically meaningful metric of immune age derived from high-dimensional longitudinal monitoring. *Nat. Med.* **25**, 487–495 (2019).

33. Tin Kam Ho. Random decision forests. in *Proceedings of 3rd International Conference on Document Analysis and Recognition* 278–282 (IEEE Comput. Soc. Press, 1995). doi:10.1109/ICDAR.1995.598994.

34. Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).

35. Lee, H. W. *et al.* Single-cell RNA sequencing reveals the tumor microenvironment and facilitates strategic choices to circumvent treatment failure in a chemorefractory bladder cancer patient. *Genome Med.* **12**, 47 (2020).