# Contextualized Networks Reveal Heterogeneous Transcriptomic Regulation in Tumors at Sample-Specific Resolution

Caleb N. Ellington*[1], Benjamin J. Lengerich[2,3], Thomas B.K. Watkins[4], Jiekun Yang[2,3], Hanxi Xiao[5], Manolis Kellis†[2,3], and Eric P. Xing†[1,6,7]

[1]Carnegie Mellon University
[2]Massachusetts Institute of Technology
[3]Broad Institute of MIT and Harvard
[4]Cancer Institute, University College London
[5]University of Pittsburgh
[6]Petuum, Inc
[7]Mohamed bin Zayed University of Artificial Intelligence

December 1, 2023

**Abstract**

Cancers are shaped by somatic mutations, microenvironment, and patient background, each altering gene expression and regulation in complex ways, resulting in heterogeneous cellular states and dynamics. Inferring gene regulatory network (GRN) models from expression data can help characterize this regulation-driven heterogeneity, but network inference requires many statistical samples, traditionally limiting GRNs to cluster-level analyses that ignore intra-cluster heterogeneity. We propose to move beyond cluster-based analyses by using *contextualized* learning, a multi-task learning paradigm which allows us to infer sample-specific models using phenotypic, molecular, and environmental information pertinent to the model, encoded as the model's "context" to be conditioned on. We unify three network model classes (Correlation, Markov, Neighborhood) and estimate context-specific GRNs for 7997 tumors across 25 tumor types, with each network contextualized by copy number and driver mutation profiles, tumor microenvironment, and patient demographics. Contextualized GRNs provide a structured view of expression dynamics at sample-specific resolution, which reveal co-expression modules in correlation networks (CNs), as well as cliques and independent regulatory elements in Markov Networks (MNs) and Neighborhood Regression Networks (NNs). Our generative modeling approach allows us to predict GRNs for unseen tumor types based on a pan-cancer model of how somatic mutations affect gene regulation. Finally, contextualized networks enable GRN-based precision oncology, explaining known biomarkers in terms of network-mediated effects, and leading to novel subtypings for thyroid, brain, and gastrointestinal tumors that improve survival prognosis.

## Introduction

Tumors are heterogeneous, developing through clonal evolution that accumulates mutations, including cancer-driving single-nucleotide variants (SNVs) and somatic copy number alterations (SCNAs). In addition to tumor cell intrinsic changes, tumors develop in and are shaped by a microenvironment that includes immune cells, the extracellular matrix, blood vessels and surrounding cells. This extensive heterogeneity necessitates hetereogeneous treatments targeted

*cellingt@cs.cmu.edu
†Corresponding Authors: eric.xing@mbzuai.ac.ae, manoli@mit.edu

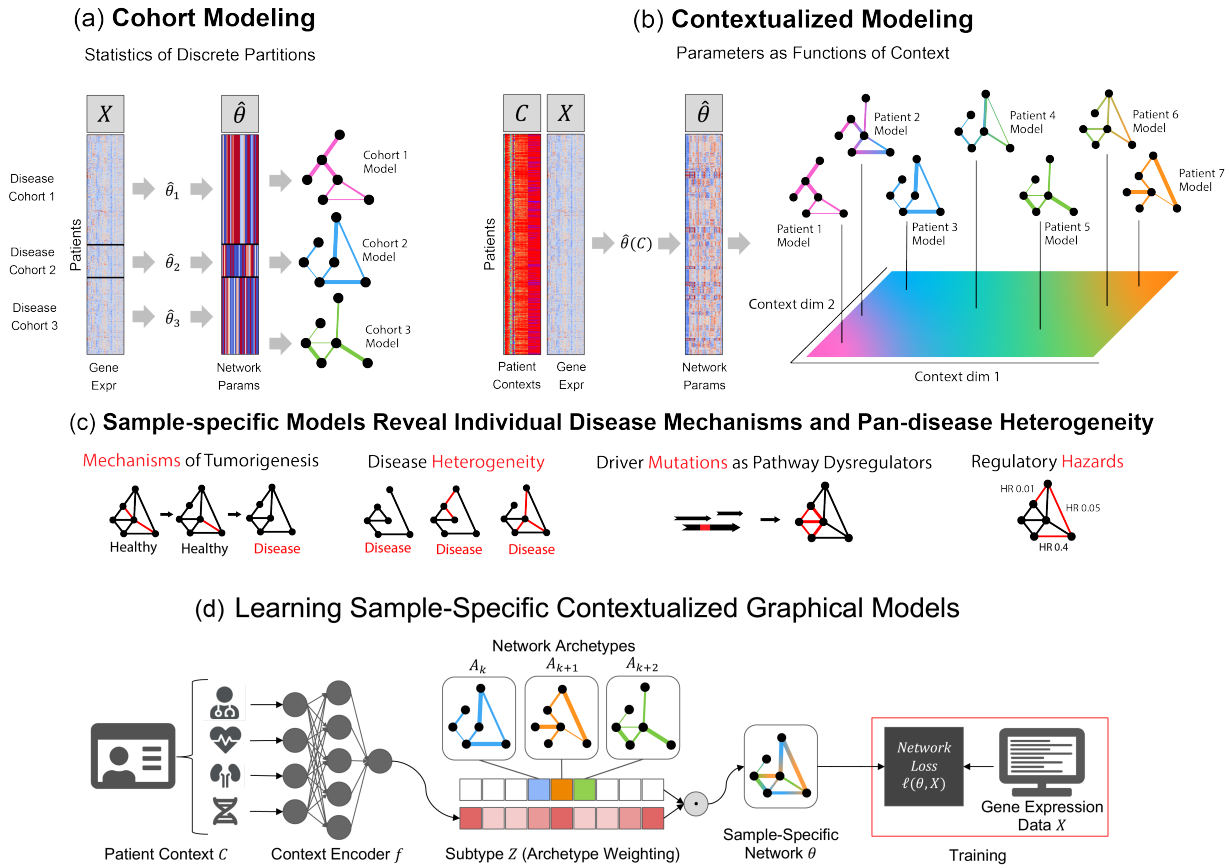



Figure 1: (a) Traditional modeling approaches assume each training cohort or (sub)population is homogeneous and samples are identically distributed. Cohorts must be large enough to allow robust inference, presenting a tradeoff between personalization and power. (b) Contextualization assumes model parameters are a function of context, allowing powerful context-specific inference without *a priori* clustering of subpopulations or assuming homogeneity. Contexts can be unique to each sample, permitting sample-specific model inference. (c) Sample-specific models reveal population heterogeneity, relate rare pathological mechanisms to more common ones, and provide new data views for prognosis and biomarker identification. (d) Graphical depiction of the deep learning framework. Sample context is used to predict weights on each of the model archetypes, which we call the subtype. The sample-specific network is estimated as the tensor dot product of archetypal networks and subtype weights. The network archetypes are learned simultaneously alongside the context encoder using backpropagation.

to individual patients. However, estimating treatment effects and patient prognosis at patient-specific resolution implies an n-of-1 approach to treatment that is technically and temporally infeasible. Instead, methods have historically sought to identify prognostic biomarkers that stratify patients into tumor subtype cohorts, and predictive biomarkers that identify patients who generally respond to treatment. The Cancer Genome Atlas[1] (TCGA) derives prognostic subtypes via cluster analysis on clinical and molecular data, including cancer-driving SNVs, SCNAs, DNA methylation, mitochondrial DNA, RNA-seq, miRNA, protein abundance arrays, histology images, patient demographics, and/or immunological data, and further identifies prognostic biomarkers as features that differentiate these clusters [1–23]. While clusters can be analyzed in terms of feature stratification, clustering ignores the latent feature interactions and hierarchical feature relationships that define biological systems. Biomarkers identified by cluster analysis have no mechanistic interpretation, and require further experimentation to validate their role in tumorigenesis and

[1] https://www.cancer.gov/tcga

tumor pathology. Consequently, the identification of biomarkers using somatic DNA alterations or gene expression patterns has proved challenging [24], but biological dogma and notable exceptions (e.g. HER2 amplification in breast cancer) motivate us to find a systematic way to search for differentiating regulatory factors that reflect cellular states and foreshadow cellular responses to treatments. In our view, biomarker discovery should directly inform the development of novel treatments, revealing molecular features that relate to the robustness or fragility of molecular systems in individual tumors. Addressing the shortcomings of cluster analysis, we focus on three questions: (1) how do we model the mechanisms of molecular interactions as it relates to tumorigenesis and treatment efficacy, (2) how do we identify prognostic biomarkers for rare diseases and outlier patients that are too sparsely sampled to cluster, and (3) how can we quantify the heterogeneity of tumor pathology, which is widely acknowledged but poorly understood, and encode or represent the myriad of phenotypic, molecular, and environmental factors driving this heterogeneity from observational data alone?

Toward representing interactions, gene regulatory networks (GRNs) represent the functional circuitry within cells that simultaneously respond to biomolecular stimulus and drive tumorigenesis. We intuit that many interactions between disparate biomolecular features can be identified at the cellular level through transcriptomic regulation, both directly and indirectly. Further, tumor-specific GRNs capture regulatory redundancy and fragility in individual cancers, whereby multi-omic features relate to GRN structure and organization, and GRN organization reveals the functional mechanisms of tumor pathology and the robustness of therapeutic targets. Single-cell and multi-omic profiling have advanced the potential for studying highly context-specific regulatory relationships in GRNs, but computational methods of inferring GRNs continue to rely on partitioning samples into homogeneous sets of samples [25–28]. As such, existing methods for high-resolution network inference either impose strong biological priors based on known transcription factor-gene regulation [29], or apply a sample-left-out approach that lacks statistical power [30, 31]. Partition-based modeling is insufficient to capture high-resolution or continuously rewiring GRNs, a problem for precision oncology because some types of cancer neither form discrete clusters [32] nor cluster by tissue of origin [33].

More generally, the exponential increase of data set complexity, heterogeneity, and size, has motivated the need for sample-specific inference in many application areas [34–38]. Contextualized modeling [39] addresses this by representing the heterogeneity in data as driven by sample-specific models, and explaining variation among sample-specific models in terms of sample context encodings. These contexts can be any information that may explain heterogeneity in the data (e.g. age, genotype, medical images, environmental factors). More traditionally, context-driven heterogeneity might be controlled by performing probabilistic inference on context-specific data sets, but this fails to scale to high-dimensional and continuously-varying contexts, common in biomedical data, where context data splits have as few as one sample and most context conditions are missing entirely. The simplest and most classic version of the contextualization paradigm are varying-coefficient models [40], which account for the effects of a univariate and continuously-varying context on a linear model's parameters.

Modern contextualized models, proposed in [41] and generalized in [39], are a combination of statistical modeling and deep learning, where context encoders are typically neural networks that can utilize any multi-modal contextual information. This framework also introduces the concept of model archetypes (Figure 1d), whereby all sample-specific models are spanned by the set of model archetypes, constraining and explaining their variation through the context encoding which parameterizes this space (See Methods). Thus, these archetypes, also learned from data, link the heterogeneity of sample-specific models to variation in the context encoding and enable sharing information between sample-specific model inference tasks. Many notable works on heterogeneous linear effects use this framework [41–46], but contextualized models have yet to be extended to the more general graphical modeling regime.

To infer tumor-specific GRNs that account for patient-to-patient heterogeneity, we propose to reframe GRN inference within the contextualized modeling paradigm, thereby sharing information among tumor-specific inference tasks by relating these tasks through their clinical and molecular contexts. By recasting networks as the output of a learnable function, our approach shares statistical power between samples while also permitting fine-grained variation to capture the complexity of sample-specific contexts such as tissue-of-origin, somatic mutation landscape, tumor microenvironment and clinical measurements. We formulate three types of GRNs (Markov, Neighborhood, and Correlation networks) under this paradigm, and estimate sample-specific GRNs which enable sample-specific analyses of latent regulatory processes. By applying this computational framework to over 7000 samples, we find that contextualized networks improve prediction of held-out expression data and reveal latent heterogeneity which has previously been obscured by partition-based methods of network inference.

# Results

Graphical model inference is a canonical task in the life sciences, whereby unstructured observational data is transformed into graph revealing the latent structure, strength, and direction of interactions between biological entities. However, for heterogeneous populations where latent models change from individual to individual, such as how tumorigenesis is driven by patient-specific environmental and molecular factors, traditional modeling approaches are ill-defined. We introduce contextualized GRNs, which learn to model the effect of individual clinical and molecular contexts on GRN parameters, revealing latent model-based drivers of GRN dysregulation and tumor heterogeneiety.

While GRNs are commonly interpreted as adjacency graphs [25, 27, 29], existing methods for GRN inference can be categorized as variants of four probabilistic models: Markov networks, which represent pairwise dependencies, Pearson's correlation networks, which represent pairwise correlations, neighborhood regression networks, which represent each node as a linear combination of its neighbors, and Bayesian networks, and which represent directed and acyclic interactions. We focus on Markov, correlation, and neighborhood networks, unifying these models through a reparameterization trick, and thus enable them to be contextualized uniformly within our framework.

The power gained from contextualization allows us to estimate highly accurate sample-specific network models without incorporating any prior knowledge of network structure. We apply contextualized graph estimation to infer context-specific networks for 7997 patients in TCGA, utilizing molecular contexts including cancer-driving single nucleotide variations (SNVs), somatic copy number alterations (SCNAs), biopsy composition metrics, and patient demographics. Across all network classes, contextualized networks confer significant improvements for network accuracy and likelihood, generalizing to held-out populations and held-out disease types. We evaluate our 7997 patient-specific GRNs for new clinical and biological insights, discovering robust state-of-the-art prognostic subtypes for thyroid carcinoma (THCA) and brain lower grade glioma (LGG), as well as cross-tissue disease groups like the gastointestinal tract that includes four distinct tumor types. Finally, patient-specific networks relate prognostic biomarkers to changes in specific regulatory modules and gross GRN organization, and identify candidate biomarkers for further investigation.

## Contextualization Recovers Latent Variation from Heterogeneous Observational Data

Observational data is cheap and abundant in comparison to experimental data, but suffers from complex environmental confounders, contexts, and conditions. To recover latent processes of data generation through model estimation with observational data, the user must artificially control for on sample contexts and conditions to emulate a controlled experimental environment. However, controlling for all conditions and contexts simultaneously, especially on biomedical data with high-dimensional contexts, leads to conditions with as few as just one sample – too small to infer accurate context-specific models. Ignoring contextual effects (i.e. population modeling) is similarly ill-advised for heterogeneous data, leading to spurious results from models that are misspecified and inaccurate (e.g. Simpson's paradox).

Contextualization [39] addresses this by applying deep learning to the meta-relationship between contextual information and context-specific model parameters. Contextualization unifies previous approaches such as varying-coefficient modeling [40], cluster analysis, and cohort analysis by introducing two simple concepts: a context encoder which translates sample context into model parameters, and a sample-specific model which represents the latent context-specific mechanisms of data generation. By learning how models change in response to context, contextualization enables powerful control over high-dimensional and continuously varying contexts, discovering dynamic latent structure underlying data generation in heterogeneous populations.

## Contextualization Enables Estimation of Sample-Specific Correlation, Markov, and Neighborhood Networks

To contextualize Markov, correlation, and neighborhood networks, we first unify them with linear parameterizations equivalent to each models' unique constraints (See Methods). Linear parameterization provides a differentiable objective for optimizing each model and the linear residual errors are proportional to the negative log likelihood of each network model under the data. Our unifying linearization of these models allows us to apply contextualization uniformly to each network class, and further enables us to benchmark and test the effects of common model personalization paradigms against contextualization in terms of model likelihood and modeling errors.

| | Markov | Neighborhood | Correlation |
|---|---|---|---|
| Population | $0.985 \pm 0.006$ | $0.984 \pm 0.004$ | $0.963 \pm 0.000$ |
| Cluster-specific | $0.365 \pm 0.014$ | $0.349 \pm 0.012$ | $0.683 \pm 0.052$ |
| Disease-specific | $0.368 \pm 0.003$ | $0.351 \pm 0.003$ | $0.673 \pm 0.002$ |
| **Contextualized** | $\mathbf{0.322 \pm 0.014}$ | $\mathbf{0.296 \pm 0.013}$ | $\mathbf{0.529 \pm 0.019}$ |
| Error Reduction | $14.6\% \pm 3.4\%$ | $18.1\% \pm 3.3\%$ | $20.2\% \pm 3.4\%$ |

Table 1: Error of inferred networks to match held-out gene expression profiles. For all three types of networks (Markov, Neighborhood, and Correlation), we report mean-squared error (MSE) of gene expression predicted by the network. Reported values are mean ± std over 30 runs with a bootstrapped training set and randomly initialized model weights. Error reduction is reported relative to the best baseline, which in all cases is disease-specific modeling.

## Contextualized Networks Improve Likelihood of Held-Out Expression Profiles

Contextualization improves the fit of networks models to gene expression data (Table 1). We benchmark the contextualized networks by comparing against several granularities of partition-based models: (1) a population network model which estimates the same network for all samples, (2) cluster-specific networks that are estimated independently for each cluster of contextual information, and (3) disease-specific networks that are estimated independently for each cancer type (Fig. 7). For all three network models, we evaluate the fit of the network model to actual expression data by measuring the predictive performance of the network graphical model. These predictive performances are measured as mean-squared errors between predicted and observed expression data which are inversely proportional to the model likelihood under the probabilistic interpretation of the network graphical model. Relative to disease-specific model inference (the best baseline method), contextualized networks reduce modeling error on average by 14.6% for Markov networks, 18.1% for neighborhood selection, and 20.2% for correlation networks. Contextualized graphical models achieve this improved predictive performance by accounting for contextual dependencies in model parameters without imposing prior assumptions on the form of these dependencies. As a result, contextualized graphical models capture context-specific effects that can be overlooked by group-level modeling approaches (e.g. cluster-specific, disease-specific models).

## Contextualized Networks Share Power Between All Cancer Types and Infer Models for Unseen Diseases

Contextualization relates transcriptional regulation to genomic variation through a context encoder (Fig. 1). During training, the encoder learns to modify the parameters of a downstream network model in response to contextual signals. At test time, the encoder uses learned context signals to generalize between sparsely sampled contexts. Rare or undersampled diseases like Kidney Chromophobe (KICH) and Glioblastoma multiforme (GBM) can benefit from contextual signals learned from well-sampled diseases in similar tissues (Figure 2b). In disease-specific modeling, these smaller subpopulations must either be lumped within a larger tissue group, ignoring subpopulation heterogeneity, or modeled individually, sacrificing statistical power in a "large $p$ small $n$" regime. For example, there are $n = 75$ training samples from KICH patients, while each disease-specific network has $50 \times 50$ edges, or $p = 2500$ parameters; estimating a disease-specific network from such limited data would be prohibitively high-variance.

Furthermore, contextualization adapts models to unseen contexts at test time, responding to even extreme distribution shift (Fig. 2a). For completely unseen contexts, the context encoder can still leverage learned relationships between contexts and models to infer *zero-shot* network models on-demand. We evaluate model performance through a disease-fold cross vaidation, where we hold out each of the 25 disease types in turn and learn to contextualize networks on the remaining 24. Notably, disease-specific modeling cannot be applied in this regime. In contrast, contextualized networks improve model performance and reduce error on 22 or 25 hold-out diseases, even when generalizing to an entirely new disease type.

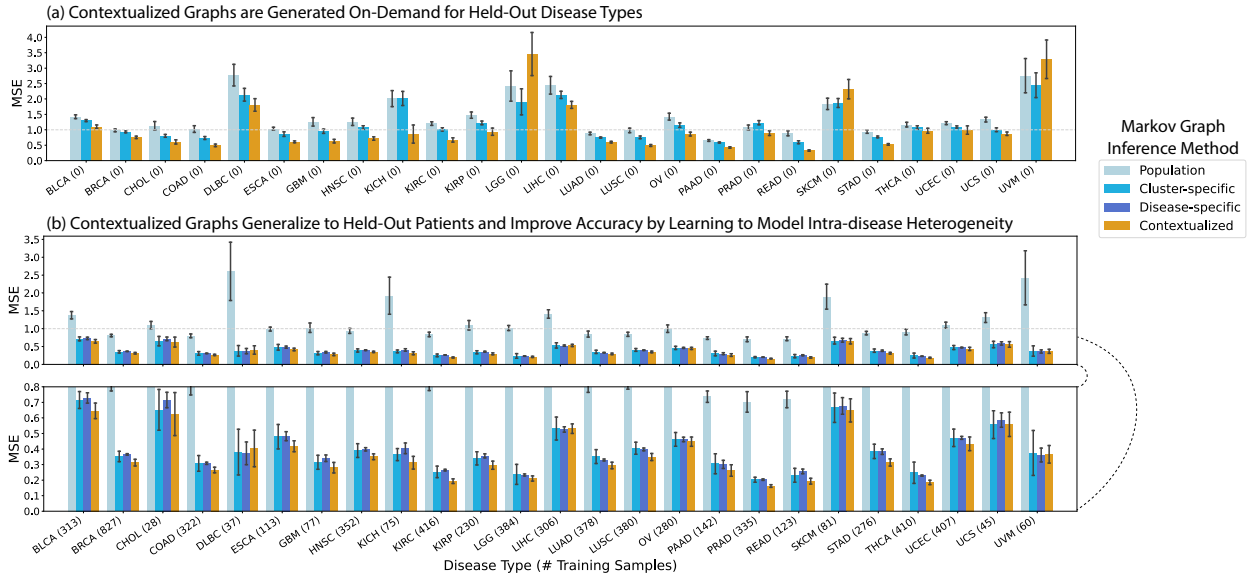



Figure 2: Performance of Contextualized Markov Networks. **(a)** Disease-fold cross-validation, in which each of the 25 disease types are held out from training and evaluated only at testing time. Disease-specific network inference cannot be applied in this regime. **(b)** Testing on held-out patients. Results are from 30 bootstrapped runs for each hold-out disease type and the hold-out patient set. Bar height is the group-averaged mean squared-error of the bootstrap-averaged network models. Error bars are the standard deviation over bootstraps of the group-averaged mean squared-error of the network models.

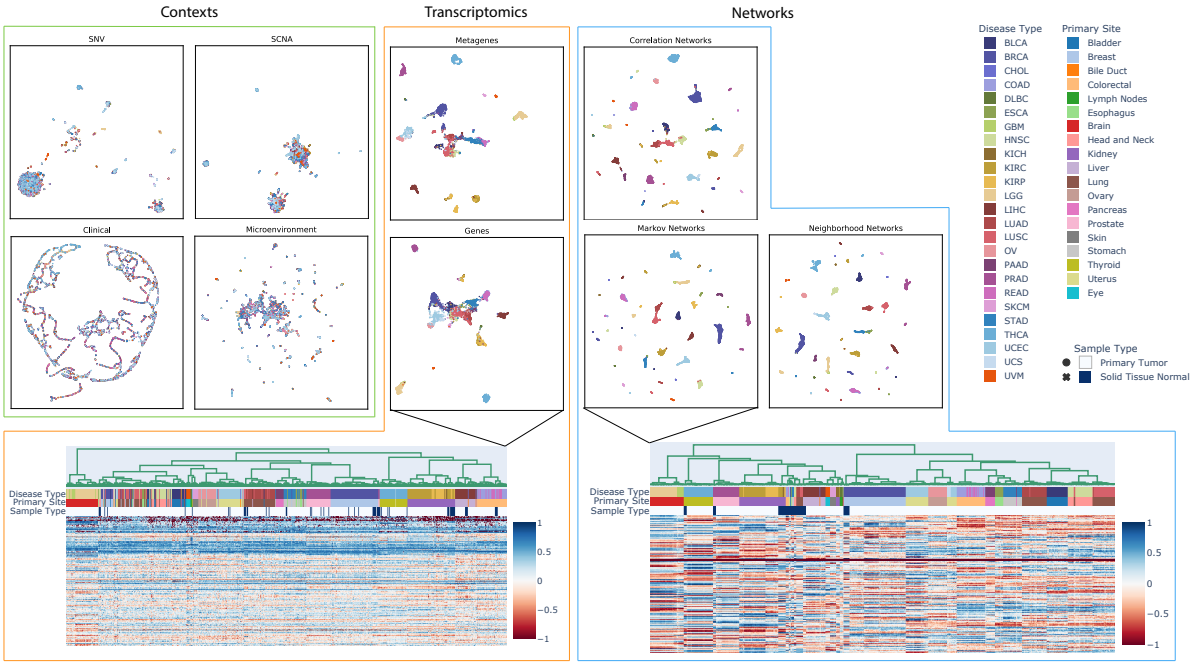



Figure 3: Embeddings, colored by disease type, reveal the organization of different disease views. Context views alone cannot capture tumor disease types. Transcriptomic views recapitulate disease types. Contextualized networks discover new separations and similarities, revealing disease subtypes and cross-disease relationships.

## Contextualized Networks Reveal Tissue-Specific Regulatory Modules

Contextualization produces context-specific network models, resulting in patient-specific networks for all 7997 patients in our TCGA dataset. Organizing patients according to their network models reveals that tissue type is a primary driver, but not the sole factor in determining gene-gene interactions (Fig. 3). In particular, diseased networks differ drastically from healthy networks, while gene and PCA-derived metagene expression profiles are still largely tissue-derived. Additionally, intra-disease (Fig. 5a) and inter-disease (Fig. 6a) subtypes are visible even at pan-cancer resolution (Fig. 3), making obvious common tumorigenesis mechanisms that underly noisy gene expression dynamics.

## Contextualized Networks Reveal Regulatory Modules Conserved Across Tissues in Cancer

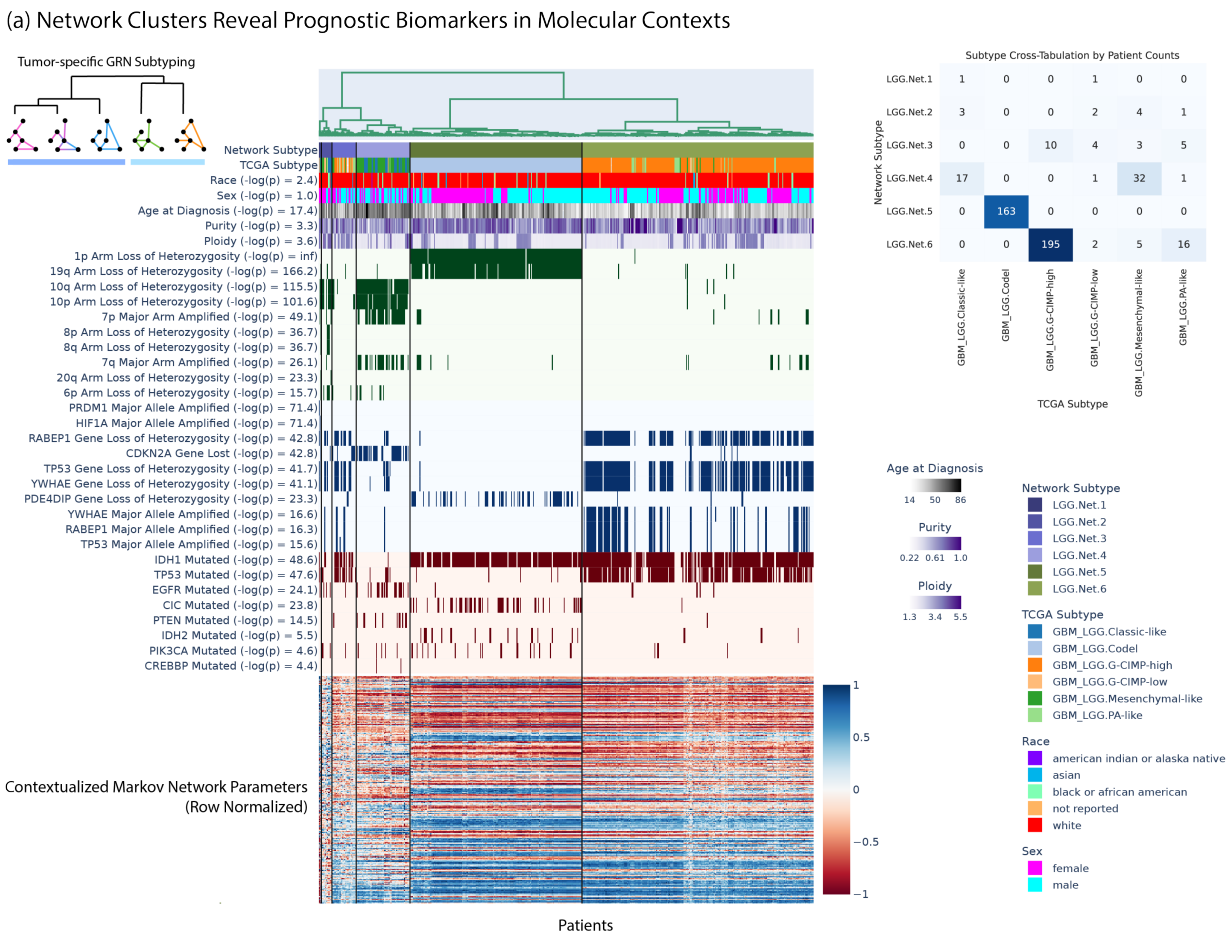
Contextualized networks reveal that tumors of the GI tract display a continuum of GRN dysregulation (Figure 6). While this continuum cannot be captured by existing TCGA subtypes [47], contextualized networks form clusters that relate existing subtypes to inter-disease and intra-disease heterogeneity via conserved regulatory motifs and shared dysregulation motifs. Contextualized networks reveal that tumors of the GI tract display a continuum of GRN dysregulation (Figure 6). While this continuum cannot be captured by existing TCGA subtypes, contextualized networks form clusters that relate existing subtypes to inter-disease and intra-disease heterogeneity via conserved regulatory motifs and shared dysregulation motifs. Finally, contextualized networks discover disparate types of GRN dysregulation within patients assigned to the SCNA-derived GI.CIN subtype, comprising the majority of GI tract tumors (Fig. 6a). Re-assigning patients based on GRN-derived subtypes improves prognosis (Fig. 6b) and reveals biomarkers of these dysregulation subtypes (Fig. 6a) including SNV-SCNA interactions such as HRAS mutations with chromosome 18 arm p loss of heterozygosity.

## Contextualized Networks Discover Novel Prognostic Subtypes

| | Expression | TCGA CoCA | Networks |
|---|---|---|---|
| Multivariate log-rank test (-log(p)) | 8.53 | 9.65 | 11.24 |
| Minimum Pairwise log-rank test (-log(p)) | 8.27 | 9.55 | 11.71 |

Table 2: Stratification disease subtypes in terms of survival. Survival tests quantify the difference in survival distributions between groups as a p-value. Contextualized networks improve on both tests on average by several orders of magnitude compared with other subtyping methods. The multivariate log-rank test quantifies overall stratification of survival distributions across all subtypes. The minimum pairwise result is the minimum p-value of all pairwise subtype tests, showing the maximum survival stratification between prognostic subtypes.

For each of the 25 tumor types, we cluster patients by their contextualized networks to identify network-based tumor subtypes by flattening the network parameters and applying hierarchical ward clustering. To compare the prognostic utility of network-based subtypes against the prognostic utility of state-of-the-art TCGA subtypes and expression subtypes, we use the same number of clusters for each disease as subtypes annotated in TCGA. We find that network-based subtypes are more prognostic on average than both expression-derived subtypes and TCGA subtypes (Table 2). In general, we find that network-based subtypes either recapitulate or refine prognostic subtypes produced by TCGA, which often utilize additional data types including DNA methylation, miRNA, and histopathological imaging. All subtype comparisons by disease are available in Appendix S4. For 10 of 25 tumor types, contextualized networks reduce one of the survival function p-values by at least an order of magnitude, and in some cases, as much as 9 orders of magnitude on KIRC, 4 orders of magnitude on LGG, 2.4 orders of magnitude on THCA, and 2.3 orders of magnitude on HNSC. On KICH, both network subtypes and TCGA subtyapes are outperformed by expression subtypes by 13.5 orders of magnitude. In the second and third worst cases for contextualized networks, network subtypes are outperformed by TCGA subtypes on GBM and UVM in terms of survival prognosis by about 1.5 and 1.3 orders of magnitude, respectively.



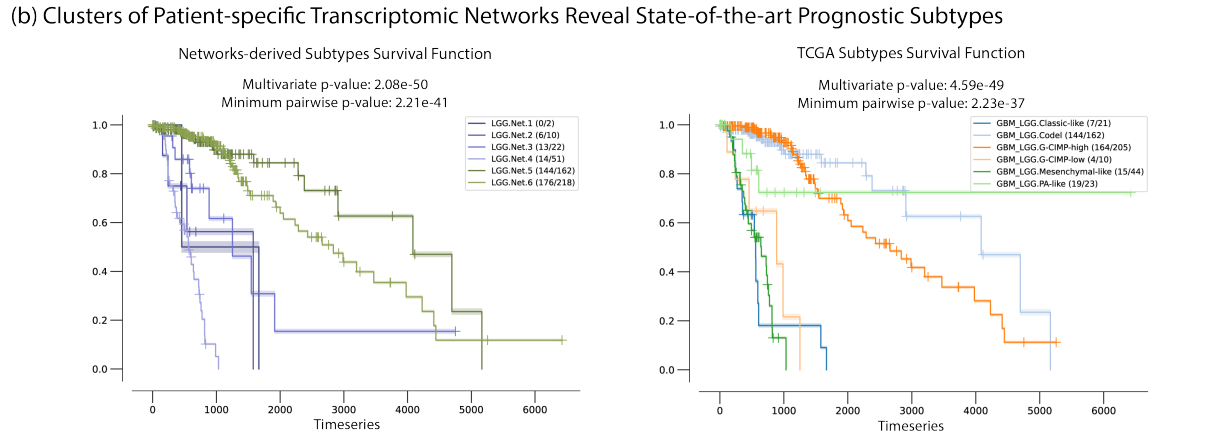




Figure 4: Exploration of network subtypes for LGG, looking at correlated clinical information, arm-level copy alterations, gene-level copy alterations, and gene-level single nucleotide variations.

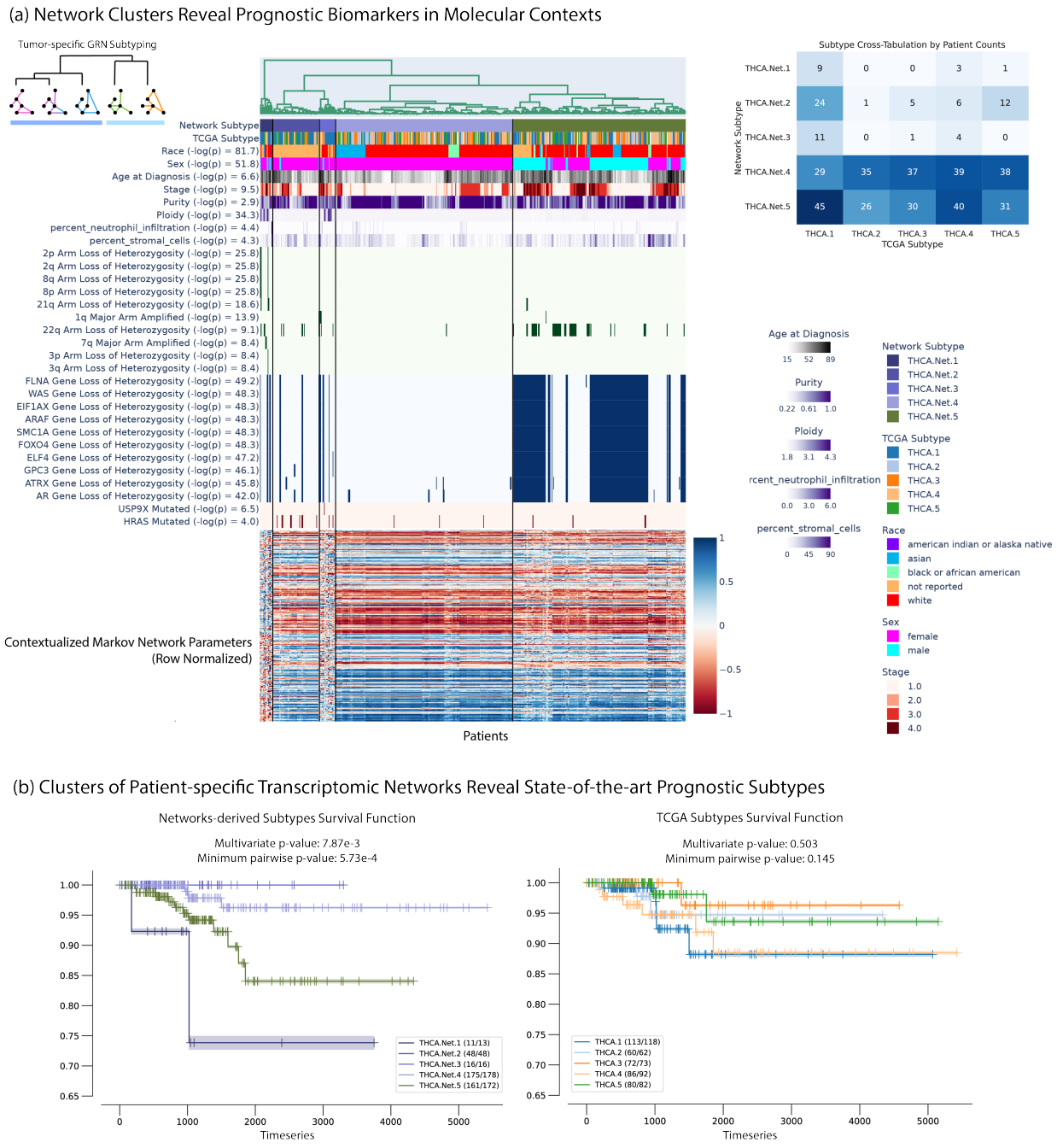



Figure 5: Exploration of network subtypes for THCA, looking at correlated clinical information, arm-level copy alterations, gene-level copy alterations, and gene-level single nucleotide variations.

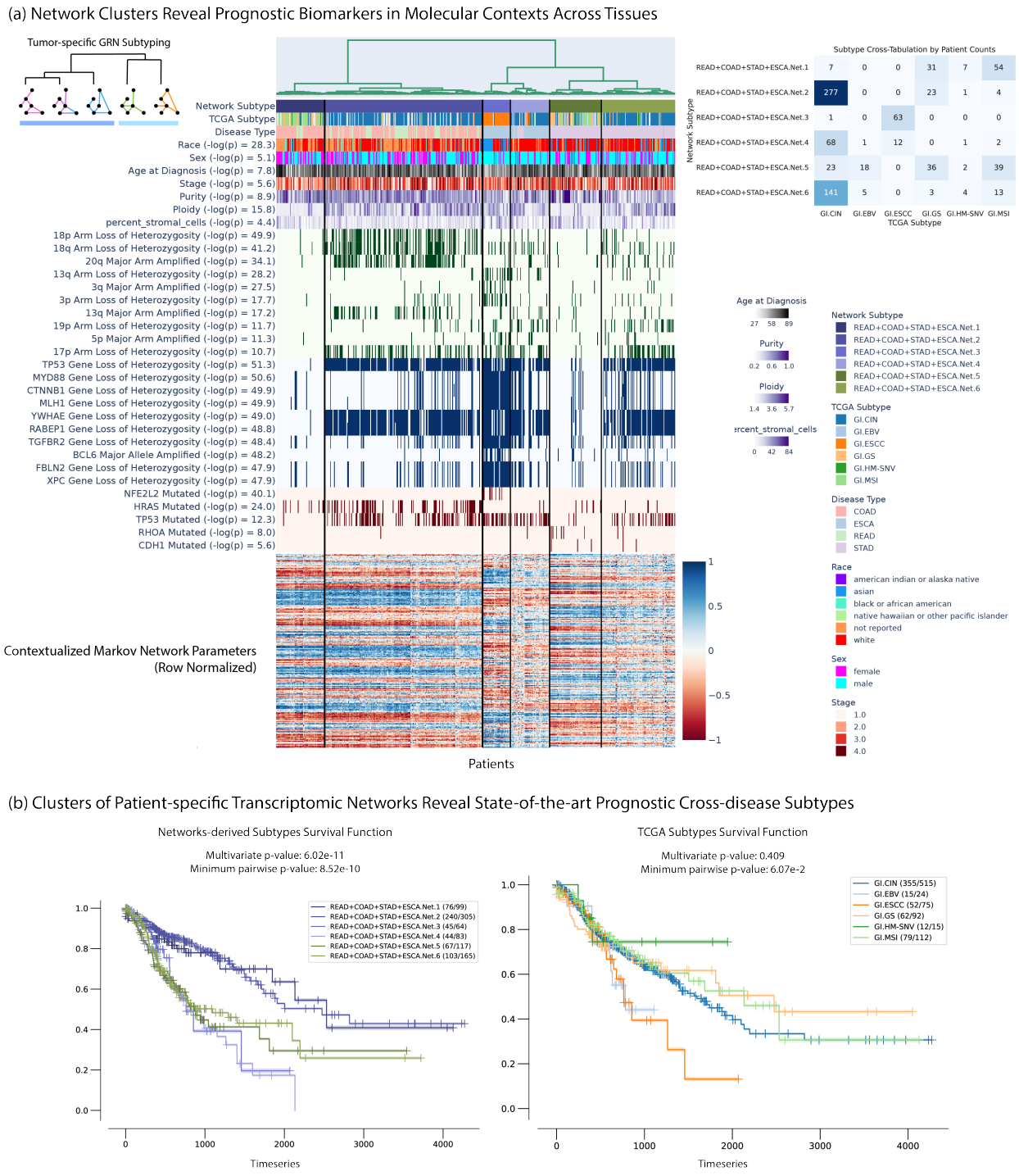



Figure 6: Exploration of cross-disease network subtypes for cancers of the GI tract, including READ, COAD, STAD, and ESCA, looking at correlated clinical information, arm-level copy alterations, gene-level copy alterations, and gene-level single nucleotide variations.

# Discussion

In this study, we propose contextualized GRNs as cohesive sample-specific representations of latent tumor states underlying disease progression and survival. Our models reveal new insights about cancer heterogeneity by relating transcriptomic, genetic, immune, and clinical factors to through tumor regulatory network topology.

The importance of context in cancer development and treatment is well recognised with treatment decisions frequently determined by a tumor's tissue of origin. The frequency of mutations in specific driver mutations varies substantially between tumors of different tissues and likely reflects the importance of distinct signaling pathways within distinct cellular contexts [48]. For example, BRAF(V600E) driver mutations vary substantially in frequency across cancer types and drugs that target the BRAF(V600E) mutant product are less effective in colorectal cancers than in skin cutaneous melanoma and non-small cell lung cancers with this mutation [49]. Further emphasizing the importance of context beyond the tissue-level, considerable variation in terms of aggressiveness, drug sensitivity, and genomic mutations, is also observed between tumors arising from the same cell type and tissue [50]. These heterogeneous genetic contexts likely hinder efforts to define tumor subgroups based on specific mutations with epistasis, which involves the action of one gene on another, having been shown to affect treatment efficacy in acute myeloid leukemia where NPM1 mutations confer a favorable prognosis only in the presence of a co-occurring IDH1 or IDH2 mutation [51].

Although genetic heterogeneity between tumors from the same tumor type is known to be widespread, it has long been thought that heterogeneity at the phenotype level may not be so marked, with the same cellular pathways often affected [52]. For example, dysregulation of the G1–S transition is observed in almost all cancers, and may occur through multiple routes, both promoting proliferation and overriding cellular senescence [53]. However, in spite of the evidence for functional convergence, it is challenging with current statistical methods to identify biomarkers that define similar phenotypes on genetically diverse contexts in order to guide treatment.

Many promising expression-based biomarkers use the level of expression across gene pathways or multiple genes rather than identifying specific somatic mutations [24]. Contextualized GRNs provide an intuitive way of identifying both subpopulations with differential transcriptomic regulation and the pathway-level cohorts of genes that should be studied as potential biomarkers, as well as the likely effect size of pathway dysregulation. Contextualized GRNs further identify associated contextual signals with these subpopulations, providing new leads for traditional classes of genomic biomarkers.

More broadly, contextualized modeling seeks to estimate context-specific models beyond context-specific sampling constraints. By sharing information among samples while also allowing sample-specific variation, our framework models complex and dynamic distributions despite physical and technical barriers that would typically prohibit sample-specific inference. Context-dependent models models naturally account for non-identically distributed data and provide a principled method for performing statistical inference on data that would traditionally be too small or too heterogeneous. While it is generally believed that biological observations are a product of latent cellular states and tumors exhibit extreme patient-to-patient heterogeneity, these ideas are orthogonal in traditional modeling regimes. Contextualized GRNs are the first to effectively unite the two: networks are a useful latent representation, relating biomarkers to pathology through systems of molecular interactions, and accounting for network heterogeneity allows us to explore both population-level and per-patient tumor pathology in terms of latent representations of molecular systems.

# Materials and methods

## Contextualized Networks

We seek a context-specific density of network parameters $\mathbb{P}(\theta \mid C)$ such that

$$\mathbb{P}(X \mid C) = \int_{\theta} d\theta \mathbb{P}_M(X \mid \theta)\mathbb{P}(\theta \mid C)$$

is maximized, where $\mathbb{P}_N(X \mid \theta)$ is the probability of gene expression $X \in \mathbb{R}^p$ under network model class $M$ with parameters $\theta \in \mathbb{R}^{p \times p}$ and context $C$, which can contain both multivariate and real features. To overcome $\theta$ being a

high-dimensional, structured latent variable, we assume that all contextualized networks lie on a subspace spanned by a set of $K$ network archetypes $\mathcal{A} \coloneqq \text{span}(\{A_k \in \mathbb{R}^{p\times p} : A_1, ..., A_K\})$, i.e. $\theta \in \mathcal{A}$. Further, the space spanned by $\mathcal{A}$ is parameterized by a latent variable ("subtype") $Z \in \mathbb{R}^K$ such that $Z$ is a deterministic function of context $Z = f(C)$ and the context-specific network model $\theta$ (and subsequently the gene expression observations $X$) are independent of context given $Z$, i.e. $C \perp (X, \theta) \mid Z$. In this way, we constrain $\theta$ as a convex combination of network archetypes via latent mixing.

$$\begin{aligned}
\mathbb{P}(X \mid C) &= \int_{\theta,Z} d\theta dZ \mathbb{P}_M(X \mid \theta)\mathbb{P}(\theta|Z)\mathbb{P}(Z \mid C) \\
&= \int_{\theta,Z} d\theta dZ \mathbb{P}_M(X \mid \theta)\delta(\theta - \sum_{k=1}^{K} Z_k A_k)\delta(Z - f(C)) \\
&= \mathbb{P}_M(X|\phi(C; f, \mathcal{A})) \\
\phi(C; f, \mathcal{A}) &= \sum_{k=1}^{K} Z_k A_k = \sum_{k=1}^{K} f(C)_k A_k
\end{aligned}$$

Where the context encoder $\phi(C; f, \mathcal{A})$ is parameterized by a learnable context-to-subtype mapping $f$ and the set of archetypes $\mathcal{A}$. This architecture is shown in Figure 1d, and is learned end-to-end with backpropagation. While the archetypal networks $A_k$ could use prior knowledge for initialization or regularization, no prior knowledge is required. In all experiments reported here, we do not use any prior knowledge of network structure or parameters.

This framework unites three different perspectives of GRNs: (1) Correlation networks, in which network edges are the pairwise Pearson's correlation between nodes, (2) Markov networks, in which edges are the pairwise precision values representing conditional dependencies between nodes, and (3) Neighborhood regression networks, in which edges represent directed linear relationships between nodes. The key challenge for each network class is to define a differentiable loss function $\ell_M$ that is proportional to the negative log probability of our contextualized network model.

$$\begin{aligned}
\widehat{f}, \widehat{\mathcal{A}} &= \underset{f,\mathcal{A}}{\text{argmax}} \sum_{n=1}^{N} \log\left(\mathbb{P}_M(X_n|\phi(C_n; f, \mathcal{A}))\right) \\
&= \underset{f,\mathcal{A}}{\text{argmin}} \sum_{n=1}^{N} \ell_M\left(\phi(C_n; f, \mathcal{A}), X_n\right)
\end{aligned}$$

The loss objective can be used in the end-to-end optimization, solving for the context encoder and the network archetypes simultaneously, and subsequently inferring the context-specific parameters $\theta$. Below, we outline a unifying linear parameterization of each network loss. Implementation details are discussed in Appendix S1.

## Contextualized Neighborhood Regression

We first apply contextualization to the graph variable selection algorithm proposed by Meinhausen and Buhlmann [54]. The direct relationship of this model to lasso regression links contextualized neighborhood regression to original works on contextualized linear models [41], making it a convenient stepping stone toward the graphical models in the sequel. The model is a Gaussian graphical model where $X \sim N(0, \Sigma)$ and $\Sigma$ has sparse off-diagonal entries. The algorithm, neighborhood regression, recovers edges between nodes with non-zero partial correlations by solving the lasso regression for every feature $X_i$, given every other feature $X_{-i}$, where regression maximizes $P(X_i|X_{-i})$ via the loss

$$\widehat{\theta}_i = \underset{\theta}{\text{argmin}} \|X_i - X_{-i}\theta\|_2^2 + \lambda\|\theta\|_1$$

resulting in edges with source $X_j$ for every $j \neq i$ and sink $X_i$ and strength $\theta_{ij}$, or no edge if $\theta_{ij} = 0$. Equivalently, we parameterize the neighborhood selection objective using the square matrix of network edge parameters $\theta \in \mathbb{R}^{p\times p}$.

$$\widehat{\theta} = \underset{\theta}{\operatorname{argmin}} \|X - X\theta\|_F^2 + \lambda \sum_i \|\theta_i\|_1 \text{ s.t. } \operatorname{diag}(\theta) = [0]$$

Where the contextualized neighborhood network objective replaces $\theta$ for each sample with a context-specific $\theta_n = \phi(C_n; f, \mathcal{A})$. Finally, we define a function $\phi'$ to mask the diagonal of $\theta$, presenting the loss function $\ell_{NN}$ for contextualized neighborhood regression networks

$$\ell_{NN}(\phi(C; f, \mathcal{A}), X) = \|X - X\phi'(C; f, \mathcal{A})\|_2^2 + \lambda \sum_i \|\phi'(C; f, \mathcal{A})_i\|_1$$
$$\phi'(C; f, \mathcal{A}) = (1 - \mathbb{I}) \otimes \phi(C; f, \mathcal{A})$$

where $\otimes$ is the hadamard product.

## Contextualized Markov Networks

Linear regression and Gaussian graphical models are intrinsically related, allowing us to extend work on contextualized linear models to various graphical representations of the Gaussian graphical model. To estimate sample-specific precision matrices representing the conditional dependency structure of an undirected graphical model or Markov network, we assume the data is drawn from $X \sim N(0, \Omega^{-1})$ where $\Omega = \Sigma^{-1}$ and estimate pairwise partial correlation coefficients. Using an equivalence defined by Peng et al. [55], the partial correlation coefficient is defined by regression as

$$\rho_{ij} = \operatorname{sign}(\beta_{ij})\sqrt{\beta_{ij}\beta_{ji}} = -\frac{\omega_{ij}}{\omega_{ii}\omega_{jj}}$$

Where the precision matrix $\Omega$ has elements $\omega_{ij}$ and $\beta$ is the ordinary least squares solution to multivariate linear regression $\beta_i = \operatorname{argmin}_\beta \|X_i - X_{-i}\beta\|_2^2$. Critically, the precision matrix directly encodes conditional independence between features in $X$, and thus precision encodes the Markov network.

$$\omega_{ij} = 0 \iff X_i \perp X_j \mid X_{-\{i,j\}}$$

Following [56], we assume constant diagonal precision $w_{ii} = w_{jj} \forall i, j$ and therefore achieve proportionality between the regression and the precision matrix.

$$\omega_{ij} \propto -\operatorname{sign}(\widehat{\beta}_{ij})\sqrt{\widehat{\beta}_{ij}\widehat{\beta}_{ji}}$$

Assuming unit diagonal precisions $\omega_{ii} = 1$, the proportionality becomes exact equivalence. Further, proportionality induces symmetry in the regression, i.e. $\beta_{ij} = \beta_{ji}$. We encode this in the objective by requiring our estimate for $\theta$ to be a symmetrically augmented matrix based on $\gamma$, i.e. $\beta = \gamma + \gamma^T$

$$\widehat{\gamma} = \underset{\gamma}{\operatorname{argmin}} \|X - X(\gamma + \gamma^T)\|_F^2 \text{ s.t. } \operatorname{diag}(\gamma) = [0]$$

If $\Omega$ is sparse, we can apply lasso regularization to the multivariate regression objective [54]. Given the similarity between this differential Markov network objective and the neighborhood regression objective, we follow the exact contextualization procedure from above to contextualize $\gamma$ and arrive at a loss function $\ell_{MN}$

$$\ell_{MN}(\phi(C; f, \mathcal{A}), X) = \|X - X(\phi'(C; f, \mathcal{A}) + \phi'(C; f, \mathcal{A})^T)\|_2^2 + \lambda \sum_i \|\phi'(C; f, \mathcal{A})_i\|_1$$

where $\phi'$ is defined identically for masking the diagonal. The resulting precision matrix estimate is $\widehat{\Omega} = -(\phi'(C; \widehat{f}, \widehat{\mathcal{A}}) + \phi'(C; \widehat{f}, \widehat{\mathcal{A}})^T)$. In practice we do not threshold the estimated precision to non-zero values, instead using the exact precision matrix to represent the Markov network, retaining information about dependency strength as well as dependency structure in the network.

## Contextualized Correlation Networks

Correlation networks are simple to estimate and often state-of-the-art for gene regulatory network inference [27]; contextualized correlation expand this utility to the granularity of sample-specific network inferences. To estimate sample-specific correlation networks, we assume the data was drawn from $X \sim N(0, \Sigma)$ and use the well known univariable regression view of Pearson's marginal correlation coefficient:

$$\rho_{ij}^2 = \frac{\sigma_{ij}}{\sigma_{ii}\sigma_{jj}} = \beta_{ij}\beta_{ji}$$

where the covariance matrix $\Sigma$ has elements $\sigma_{ij}$, and $\beta_{ij} = \text{argmin}_\beta (X_j - X_i\beta)^2$. This form converts correlation into two separable univariate least-squares regressions that maximize the marginal conditional probabilities $P(X_i|X_j)$ and $P(X_j|X_i)$. Contextualizing this differentiable objective, we get the contextualized correlation network loss

$$\ell_{CN}(\phi(C; f, \mathcal{A}), X) = \|X - X \otimes \phi(C; f, \mathcal{A})\|_F^2$$

where the context-specific correlation matrix is reconstructed as $\widehat{\rho}^2 = \phi(C; \widehat{f}, \widehat{\mathcal{A}}) \otimes \phi(C; \widehat{f}, \widehat{\mathcal{A}})^T$.

## Baselines

We compare contextualized modeling with several traditional approaches for context-controlled and context-agnostic inference, including population modeling, cluster modeling, and cohort modeling (Fig. 7). A population model assumes that the entire sample population is identically distributed. As a result, population modeling infers a single model representing all observations. In reality, sample populations often contains two or more uniquely distributed subpopulations. If we expect that there are several subpopulations with many observations each, and that these subpopulations can be stratified by context, it may be appropriate to cluster the data by context to identify these subpopulations and then infer a model for each context-clustered subpopulation. This assumes that all context features are equally important and therefore does not tolerate noise features well. Alternatively, when subpopulation groupings are known to be determined by a few important features, cohort modeling is more appropriate. Sample cohorts can be identified based on prior knowledge about important context features (e.g. disease type).

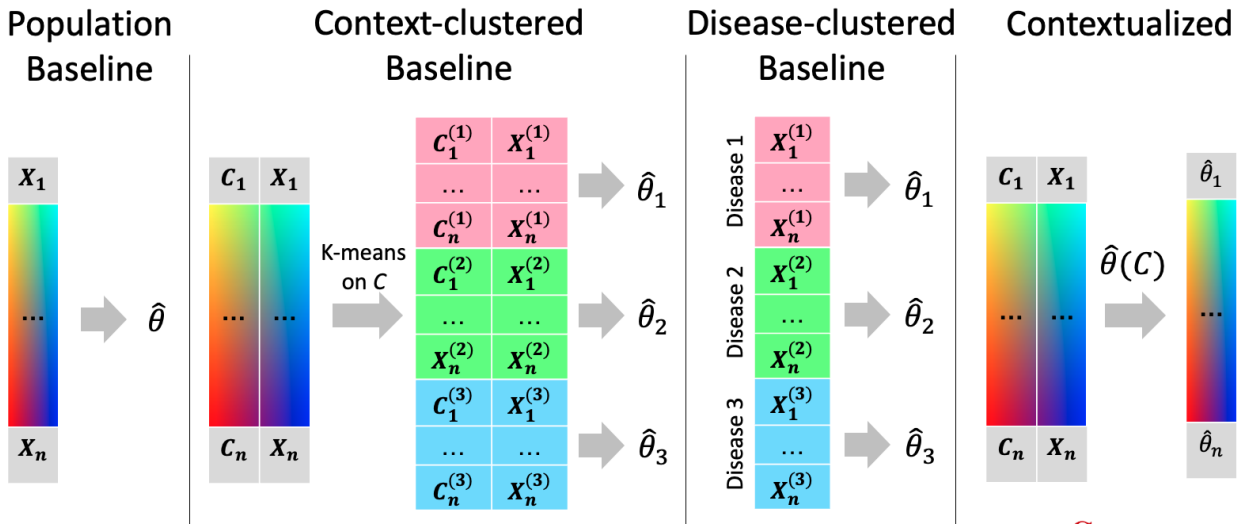



Figure 7: Modeling regimes for personalized inference.

The baseline modeling regimes enjoy the benefits of traditional inference methods (i.e. identifiability, convergence) by relying on the assumption there are a discrete number of subpopulations underlying the observed data that are each defined by a latent model, and each of these subpopulations is well-sampled. This assumption is rarely, if ever, satisfied in a real-world setting. We develop contextualized modeling as a synthesis between traditional statistical inference and modern deep learning to enable model-based analysis of heterogeneous real data. Contextualized modeling assumes a functional dependency between models, but unlike prior methods makes no assumption about the form or complexity of this dependency. As such, contextualized models permit context-informed inference even when contexts are sparsely sampled and high dimensional.

## Data

Our dataset is constructed from The Cancer Genome Atlas[2] (TCGA) and related studies, covering 7997 samples from 7648 patients with 6397 samples for training and validation and 1600 as testing. For context, we use clinical information, biopsy composition, SCNAs and cancer-driving SNVs (Appendix S2). Gene expression data was log-transformed and compressed to a set of cancer driver genes, then transformed using PCA into 50 metagenes. Networks were learned on the metagene expression data.

## Code availability

All methods are available in ContextualizedML, an open-source `PyTorch` library for contextualized modeling. Contextualized graphical models, as well as contextualized regressors can be estimated using an intuitive sklearn-style import-fit-predict workflow.

```
from contextualized.easy import ContextualizedCorrelationNetworks
model = ContextualizedCorrelationNetworks()
model.fit(C_train, X_train)
err = model.measure_mses(C_test, X_test)
r = model.predict_correlation(C)
```

We provide demos and tutorials for network inference at `https://contextualized.ml`. Our code for generating the figures in this manuscript is available at `https://github.com/cnellington/CancerContextualized`.

## Data availability

The TCGA data used is public and available for download via the Genomic Data Commons Data Portal. Data processing is detailed in Appendix S2.

[2] www.cancer.gov/tcga

# References

[1] Muzny, D. M. *et al.* Comprehensive molecular characterization of human colon and rectal cancer **487**, 330–337. URL https://www.nature.com/articles/nature11252. Number: 7407 Publisher: Nature Publishing Group.

[2] Berger, A. C. *et al.* A comprehensive pan-cancer molecular study of gynecologic and breast cancers **33**, 690–705.e9. URL https://www.cell.com/cancer-cell/abstract/S1535-6108(18)30119-3. Publisher: Elsevier.

[3] Lawrence, M. S. *et al.* Comprehensive genomic characterization of head and neck squamous cell carcinomas **517**, 576–582. URL https://www.nature.com/articles/nature14129. Number: 7536 Publisher: Nature Publishing Group.

[4] Comprehensive molecular characterization of papillary renal-cell carcinoma **374**, 135–145. URL https://doi.org/10.1056/NEJMoa1505917. Publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJMoa1505917.

[5] Collisson, E. A. *et al.* Comprehensive molecular profiling of lung adenocarcinoma **511**, 543–550. URL https://www.nature.com/articles/nature13385. Number: 7511 Publisher: Nature Publishing Group.

[6] Creighton, C. J. *et al.* Comprehensive molecular characterization of clear cell renal cell carcinoma **499**, 43–49. URL https://www.nature.com/articles/nature12222. Number: 7456 Publisher: Nature Publishing Group.

[7] Hammerman, P. S. *et al.* Comprehensive genomic characterization of squamous cell lung cancers **489**, 519–525. URL https://www.nature.com/articles/nature11404. Number: 7417 Publisher: Nature Publishing Group.

[8] Robertson, A. G. *et al.* Comprehensive molecular characterization of muscle-invasive bladder cancer **171**, 540–556.e25. URL https://www.cell.com/cell/abstract/S0092-8674(17)31056-5. Publisher: Elsevier.

[9] Ally, A. *et al.* Comprehensive and integrative genomic characterization of hepatocellular carcinoma **169**, 1327–1341.e23. URL https://www.sciencedirect.com/science/article/pii/S0092867417306396.

[10] Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma **513**, 202–209.

[11] Network, T. C. G. A. R. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas **372**, 2481–2498. URL https://www.nejm.org/doi/10.1056/NEJMoa1402121. Publisher: Massachusetts Medical Society.

[12] Farshidfar, F. *et al.* Integrative genomic analysis of cholangiocarcinoma identifies distinct IDH-mutant molecular profiles **18**, 2780–2794. URL https://www.sciencedirect.com/science/article/pii/S2211124717302140.

[13] Cabassi, A. & Kirk, P. D. W. Multiple kernel learning for integrative consensus clustering of omic datasets **36**, 4789–4796. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7750932/.

[14] Robertson, A. G. *et al.* Integrative analysis identifies four molecular and clinical subsets in uveal melanoma **32**, 204–220.e15. URL https://www.cell.com/cancer-cell/abstract/S1535-6108(17)30295-7. Publisher: Elsevier.

[15] Raphael, B. J. *et al.* Integrated genomic characterization of pancreatic ductal adenocarcinoma **32**, 185–203.e13. URL https://www.cell.com/cancer-cell/abstract/S1535-6108(17)30299-4. Publisher: Elsevier.

[16] Sohn, B. H. *et al.* Clinical significance of four molecular subtypes of gastric cancer identified by the cancer genome atlas project **23**, 4441–4449. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5785562/.

[17] Cherniack, A. D. *et al.* Integrated molecular characterization of uterine carcinosarcoma **31**, 411–423. URL https://www.cell.com/cancer-cell/abstract/S1535-6108(17)30053-3. Publisher: Elsevier.

[18] Kim, J. *et al.* Integrated genomic characterization of oesophageal carcinoma **541**, 169–175. URL https://www.nature.com/articles/nature20805. Number: 7636 Publisher: Nature Publishing Group.

[19] Ceccarelli, M. *et al.* Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma **164**, 550–563. URL https://www.sciencedirect.com/science/article/pii/S009286741501692X.

[20] The molecular taxonomy of primary prostate cancer **163**, 1011–1025. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4695400/.

[21] Davis, C. F. *et al.* The somatic genomic landscape of chromophobe renal cell carcinoma **26**, 319–330. URL https://www.sciencedirect.com/science/article/pii/S1535610814003043.

[22] Levine, D. A. Integrated genomic characterization of endometrial carcinoma **497**, 67–73. URL https://www.nature.com/articles/nature12113. Number: 7447 Publisher: Nature Publishing Group.

[23] Cohen, A., Holmen, S. & Colman, H. IDH1 and IDH2 mutations in gliomas **13**, 345. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4109985/.

[24] Ben-Hamo, R. *et al.* Predicting and affecting response to cancer therapy based on pathway-level biomarkers **11**, 3296.

[25] Badia-i Mompel, P. *et al.* Gene regulatory network inference in the era of single-cell multi-omics 1–16. URL https://www.nature.com/articles/s41576-023-00618-5. Publisher: Nature Publishing Group.

[26] Stone, M. *et al.* Identifying strengths and weaknesses of methods for computational network inference from single cell RNA-seq data. URL https://www.biorxiv.org/content/10.1101/2021.06.01.446671v1. Publication Title: bioRxiv.

[27] Pratapa, A., Jalihal, A. P., Law, J. N., Bharadwaj, A. & Murali, T. M. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data **17**, 147–154.

[28] Thompson, D., Regev, A. & Roy, S. Comparative analysis of gene regulatory networks: from network reconstruction to evolution **31**, 399–428. URL http://dx.doi.org/10.1146/annurev-cellbio-100913-012908.

[29] Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering **14**, 1083–1086. URL https://www.nature.com/articles/nmeth.4463. Number: 11 Publisher: Nature Publishing Group.

[30] Kuijjer, M. L., Tung, M. G., Yuan, G., Quackenbush, J. & Glass, K. Estimating sample-specific regulatory networks **14**, 226–240. URL http://dx.doi.org/10.1016/j.isci.2019.03.021.

[31] Lazareva, O. *et al.* DysRegNet: Patient-specific and confounder-aware dysregulated network inference. URL https://www.biorxiv.org/content/10.1101/2022.04.29.490015v1. Pages: 2022.04.29.490015 Section: New Results.

[32] Ursu, O. *et al.* Massively parallel phenotyping of coding variants in cancer with perturb-seq **40**, 896–905. URL https://www.nature.com/articles/s41587-021-01160-7. Number: 6 Publisher: Nature Publishing Group.

[33] Hoadley, K. A. *et al.* Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer **173**, 291–304.e6. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5957518/.

[34] Ageenko, I. I., Doherty, K. A. & Cleave, A. P. V. Personalized lifetime financial planning tool. URL https://patents.google.com/patent/US20100161467/en.

[35] Buettner, F. *et al.* Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells **33**, 155–160. URL https://www.nature.com/articles/nbt.3102. Number: 2 Publisher: Nature Publishing Group.

[36] Fisher, A. J., Medaglia, J. D. & Jeronimus, B. F. Lack of group-to-individual generalizability is a threat to human subjects research **115**, E6106–E6115. URL https://www.pnas.org/doi/10.1073/pnas.1711978115. Publisher: Proceedings of the National Academy of Sciences.

[37] Hart, S. Precision education initiative: Moving towards personalized education **10**, 209–211. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5476312/.

[38] Ng, K., Sun, J., Hu, J. & Wang, F. Personalized predictive modeling and risk factor identification using patient similarity **2015**, 132–136.

[39] Lengerich, B., Ellington, C. N., Rubbi, A., Kellis, M. & Xing, E. P. Contextualized machine learning. URL http://arxiv.org/abs/2310.11340. 2310.11340[cs,stat].

[40] Hastie, T. & Tibshirani, R. Varying-coefficient models **55**, 757–796. URL https://www.jstor.org/stable/2345993. Publisher: [Royal Statistical Society, Wiley].

[41] Al-Shedivat, M., Dubey, A. & Xing, E. Contextual explanation networks **21**, 1–44. URL http://jmlr.org/papers/v21/18-856.html.

[42] Lengerich, B. J. *et al.* Discriminative subtyping of lung cancers from histopathology images via contextual deep learning URL https://www.medrxiv.org/content/10.1101/2020.06.25.20140053v1.abstract. Publisher: Cold Spring Harbor Laboratory Press.

[43] Lengerich, B., Aragam, B. & Xing, E. P. Learning sample-specific models with low-rank personalized regression. URL http://arxiv.org/abs/1910.06939. 1910.06939[cs,stat].

[44] Lengerich, B. J., Nunnally, M. E., Aphinyanaphongs, Y., Ellington, C. & Caruana, R. Automated interpretable discovery of heterogeneous treatment effectiveness: A COVID-19 case study 104086. URL https://www.sciencedirect.com/science/article/pii/S1532046422001022.

[45] Deuschel, J. *et al.* Contextualized policy recovery: Modeling and interpreting medical decisions with adaptive imitation learning. URL http://arxiv.org/abs/2310.07918. 2310.07918[cs,stat].

[46] Lengerich, B., Ellington, C., Aragam, B., Xing, E. P. & Kellis, M. NOTMAD: Estimating bayesian networks with sample-specific structures and parameters URL http://arxiv.org/abs/2111.01104. 2111.01104.

[47] Liu, Y. *et al.* Comparative molecular analysis of gastrointestinal adenocarcinomas **33**, 721–735.e8. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5966039/.

[48] Vogelstein, B. *et al.* Cancer genome landscapes **339**, 1546–1558.

[49] Kopetz, S. *et al.* Encorafenib, binimetinib, and cetuximab in BRAF v600e–mutated colorectal cancer **381**, 1632–1643. URL https://doi.org/10.1056/NEJMoa1908075. Publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJMoa1908075.

[50] Burrell, R. A., McGranahan, N., Bartek, J. & Swanton, C. The causes and consequences of genetic heterogeneity in cancer evolution **501**, 338–345.

[51] Patel, J. P. *et al.* Prognostic relevance of integrated genetic profiling in acute myeloid leukemia **366**, 1079–1089.

[52] Hanahan, D. Hallmarks of cancer: New dimensions **12**, 31–46.

[53] Rubin, S. M., Sage, J. & Skotheim, J. M. Integrating old and new paradigms of g1/s control **80**, 183–192.

[54] Meinshausen, N. & Bühlmann, P. High-dimensional graphs and variable selection with the lasso **34**, 1436–1462. URL https://projecteuclid.org/journals/annals-of-statistics/volume-34/issue-3/High-dimensional-graphs-and-variable-selection-with-the-Lasso/10.1214/009053606000000281.full. Publisher: Institute of Mathematical Statistics.

[55] Peng, J., Wang, P., Zhou, N. & Zhu, J. Partial correlation estimation by joint sparse regression models **104**, 735–746. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2770199/.

[56] Wang, Z. *et al.* Bayesian edge regression in undirected graphical models to characterize interpatient heterogeneity in cancer **117**, 533–546.

[57] Zack, T. I. *et al.* Pan-cancer patterns of somatic copy number alteration **45**, 1134–1140.

[58] Calabrese, C. *et al.* Genomic basis for RNA alterations in cancer **578**, 129–136. URL https://www.nature.com/articles/s41586-020-1970-0. Number: 7793 Publisher: Nature Publishing Group.

[59] Van Loo, P. *et al.* Allele-specific copy number analysis of tumors **107**, 16910–16915. URL https://www.pnas.org/doi/full/10.1073/pnas.1009843107. Publisher: Proceedings of the National Academy of Sciences.

[60] Drews, R. M. *et al.* A pan-cancer compendium of chromosomal instability **606**, 976–983. URL https://www.nature.com/articles/s41586-022-04789-9. Number: 7916 Publisher: Nature Publishing Group.

[61] Tokheim, C. & Karchin, R. CHASMplus reveals the scope of somatic missense mutations driving human cancers **9**, 9–23.e8.

[62] Steele, C. D. *et al.* Signatures of copy number alterations in human cancer **606**, 984–991. URL https://www.nature.com/articles/s41586-022-04738-6. Number: 7916 Publisher: Nature Publishing Group.

[63] Sondka, Z. *et al.* The COSMIC cancer gene census: describing genetic dysfunction across all human cancers **18**, 696–705. URL https://www.nature.com/articles/s41568-018-0060-1. Number: 11 Publisher: Nature Publishing Group.

[64] Liberzon, A. *et al.* The molecular signatures database (MSigDB) hallmark gene set collection **1**, 417–425.

[65] Kim, E., Gheorge, V. & Hart, T. Dynamic rewiring of biological activity across genotype and lineage revealed by context-dependent functional interactions. URL https://www.biorxiv.org/content/10.1101/2021.06.25.450004v1. Publication Title: bioRxiv.

[66] Mohammadi, S., Davila-Velderrain, J. & Kellis, M. Reconstruction of cell-type-specific interactomes at single-cell resolution **9**, 559–568.e4. URL http://dx.doi.org/10.1016/j.cels.2019.10.007.

[67] Kolar, M., Song, L., Ahmed, A. & Xing, E. P. Estimating time-varying networks. URL http://arxiv.org/abs/0812.5087. ISBN: 0812.5087 Publication Title: arXiv [stat.ML].

**Supplemental Information for:**

***Contextualized Networks Reveal Heterogeneous Transcriptomic Regulation in Tumors at Sample-Specific Resolution***

# S1 Implementation

Our entire framework (Fig. 1) is implemented in PyTorch using the PyTorch Lightning framework within our open-source software ContextualizedML. The context encoder, network archetypes, and contextualized network models are learned simultaneously using end-to-end backpropagation of the network loss (defined in Methods).

**Context Encoder & Training** The context encoder is implemented as a multi-layer perceptron with 3 hidden layers, each 100 neurons wide with ReLU activations. The context data views (S2.2) are concatenated sample-wise to create a single context feature vector encompassing all views for each patient. We use a batch-size of 10 and our learning rate is chosen automatically using PyTorch Lightning's `auto-lr-find` with an initial state of 1e-3. Model weights are initialized as `Uniform[-0.01, 0.01]`. We split our dataset into 80% training-validation and 20% testing. We create 30 bootstraps of the training-validation set and finally split into 80% training and 20% validation, resulting in a 64-16-20 split for train-validation-test where the train and validation sets are bootstrapped to evaluate model variance. We use early-stopping with a patience of 5 to end training when the minimum validation loss has not been improved for 5 epochs. We retain only the model with the minimum validation loss for each bootstrap. In Table 1, we evaluate these bootstraps individually to get error means and variances. Following this, we apply each of our bootstrapped models to the non-bootstrapped training-validation set and average the outputs of each model to obtain a single graph for each patient in this set, which we evaluate in-depth in Figures 4, 5, and 6, and for all disease types in Appendix S4.

The context encoder is a highly flexible component of our framework and a driving force for future work. This attribute can be used to enforce assumptions about the relationships between contexts and models, between context features, and about the archetype space. For instance, by using a neural additive model instead of a multi-layer perceptron, we provide context-feature-specific archetype weights for interpretability. Similarly, we can augment our context encoder with a convolutional base and include imaging modalities in our context views. At the context encoder head, we currently use an unconstrained output, but applying a softmax activation would require all of the sample-specific models to lie within a polytope defined by the archetypal networks.

# S2 Data

## S2.1 Data sources

The Cancer Genome Atlas[3] (TCGA) is a publicly-available pan-cancer datasource containing genomic, transcriptomic, and clinical profiling of tumors from dozens of landmark studies. We queried TCGA for samples with bulk RNA-sequencing and merged this dataset with two follow-up studies on an overlapping set of patients.

**Somatic copy number alterations (SCNAs)** SCNAs affect a larger fraction of the genome than do any other type of somatic genetic alteration [57] and are a major driver of expression variation in cancer [58]. We used copy number profiles derived from TCGA samples using ASCAT [59] from a pan-cancer study of the role of allele-specific SCNAs in cancer [60].

**Driver single-nucleotide mutations (SNVs)** SNVs can be classified into "driver" mutations thought to provide selective growth advantage and "passenger" mutations thought to have little role in promoting cancer development. We incorporated driver SNVs from the TCGA-derived CHASMplus dataset [61]

[3] www.cancer.gov/tcga

### S2.2 Context data views

**Clinical information** This data view incorporates sample tissue-of-origin, race, age at diagnosis, gender, year of birth, and days to collection provided by TCGA.

**Biopsy Composition** This data view contains the sample's percent tumor cells, percent normal cells, percent tumor nuclei, percent monocyte infiltration, percent lymphocyte infiltration, and percent neutrophil infiltration provided by TCGA. We also incorporate expression-derived estimates of the fraction of a sample consisting of tumor cells from [60].

**Copy Number Alterations** From ASCAT [59], we gather whole genome doubling events as well as gain and loss events for bp-specific regions of hg19 based on data from [62]. We transform these gain and loss events into both arm-level and gene-level events, where arm-level events affect 85% of an entire arm in the same event, while genes-level events affect a single gene. We transform these into number of major and number of minor chromosome arms, and the number of major and minor alleles for the set of 295 genes that overlap between COSMIC [63] and MSigDB [64]. For both gene and arm-level events, we create a separate indicator for loss of heterozygosity on each gene.

**Driver Mutations** From CHASMplus [61] we gather the mutations on all COSMIC [63] oncogenes/tumor suppresor genes and binarize the presence or absence of a mutation in each gene.

### S2.3 Transcriptomic data views

**Transcriptomics** We take the set of known oncogenes/tumor suppressor genes annotated in COSMIC [63] and included in TCGA gene expression panels. We then calculate the variance of each gene in each tumor type and take a weighted sum of these variances according to the total number of samples in each tumor type. We select the top 100 genes by this metric of "intra-disease variance".

**Baselines** We are not aware of any other scalable meta-learning, deep learning, or varying-coefficient methods to produce context-informed correlation, Markov, and Bayesian networks under a universal framework. As such, our baselines apply the network estimators in S1 under several well-known and general paradigms for improving model personalization, broadly relating to cluster analysis. Our population baseline provides no personalization, learning a single model for the entire population of training samples. Our context-clustered baseline takes an unsupervised approach to personalization by first doing a k-means clustering with k=25 on the aggregated context views (S2) and then inferring cluster-specific networks. Our disease-clustered baseline uses a personalization oracle, grouping samples by tumor type and then inferring disease-specific networks.

## S3 Related work

State-of-the-art gene regulatory network estimators are limited to population, cohort, and cluster-based approaches [26, 65, 66]. Other proposals to estimate networks as the difference between a population model and a sample-left-out model lack statistical power [30]. Kolar et. al achieve sample-specific network estimation without sacrificing statistical power by using an approach similar to classic varying-coefficient models that weighs samples by their distance over context [67]. However, this approach inherently assumes smoothness of the parameters over a context, which does not align with our understanding of the non-linear, switch-like changes in biological systems that lead to disease. Contextual estimation networks (CENs) remove this smoothness assumption by inferring the relationship between context and model parameters with a neural network, but the CEN framework is only proposed as an adaptive learning approach for linear models [41]. Context-varying linear models have previously been applied to multi-omic cancer data, where context-varying coefficients inform how epigenetic markers have patient-specific effects on clinical outcomes [42]. Linear models do not inform us of the differential gene-gene interactions that explain changes in cellular behavior. To understand regulatory and metabolic variation at per-sample resolution, we require network models with context-varying structures and parameters.

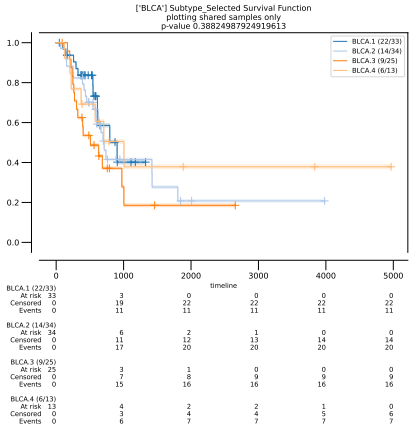
# S4 Extra Results

| Disease Type | TCGA Subtypes | Expression Subtypes | Network Subtypes |
|---|---|---|---|
| BLCA | 0.411 | 0.151 | 0.713 |
| BRCA | 1.484 | 0.616 | 1.558 |
| CHOL | – | – | – |
| COAD | 0.016 | 0.014 | 1.219 |
| DLBC | – | – | – |
| ESCA | 0.044 | 0.884 | 0.049 |
| GBM | 1.644 | 0.958 | 0.101 |
| HNSC | 1.209 | 0.312 | 3.465 |
| KICH | 0.715 | 13.802 | 0.211 |
| KIRC | 5.042 | 6.109 | 13.741 |
| KIRP | 14.538 | 10.582 | 15.205 |
| LGG | 48.338 | 33.438 | 49.681 |
| LIHC | 0.009 | 0.427 | 0.827 |
| LUAD | 0.687 | 1.172 | 0.507 |
| LUSC | 0.123 | 0.105 | 0.249 |
| OV | 0.704 | 0.684 | 0.05 |
| PAAD | 0.439 | 1.104 | 1.494 |
| PRAD | – | – | – |
| READ | 0.221 | 0.203 | 0.117 |
| SKCM | – | – | – |
| STAD | 0.044 | 1.117 | 0.575 |
| THCA | 0.298 | 0.164 | 2.104 |
| UCEC | 6.937 | 3.343 | 7.07 |
| UCS | 0.319 | 0.023 | 0.048 |
| UVM | 4.838 | 2.589 | 3.565 |

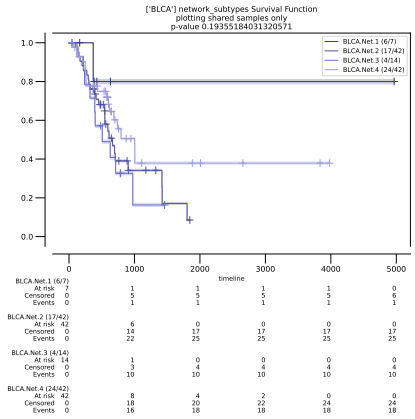
Table S1: Multivariate log-rank test comparison across different subtyping methods in terms of -log(p-value). Only samples shared between all datasets are used. – indicates no samples are shared, or subtypes do not exist for TCGA.

| Disease Type | TCGA Subtypes | Expression Subtypes | Network Subtypes |
|---|---|---|---|
| BLCA | 1.059 | 0.589 | 0.949 |
| BRCA | 2.056 | 1.13 | 2.542 |
| CHOL | – | – | – |
| COAD | 0.162 | 0.159 | 2.026 |
| DLBC | – | – | – |
| ESCA | 0.323 | 1.598 | 0.284 |
| GBM | 1.631 | 1.509 | 0.63 |
| HNSC | 1.855 | 0.853 | 3.307 |
| KICH | 0.715 | 13.802 | 0.211 |
| KIRC | 5.61 | 5.247 | 14.82 |
| KIRP | 19.696 | 9.241 | 18.661 |
| LGG | 36.533 | 25.894 | 40.656 |
| LIHC | 0.105 | 0.619 | 1.263 |
| LUAD | 1.67 | 2.29 | 1.198 |
| LUSC | 0.614 | 0.417 | 0.625 |
| OV | 1.414 | 1.325 | 0.287 |
| PAAD | 0.937 | 1.465 | 2.18 |
| PRAD | – | – | – |
| READ | 0.431 | 0.723 | 0.474 |
| SKCM | – | – | – |
| STAD | 0.469 | 1.777 | 1.428 |
| THCA | 0.837 | 0.831 | 3.242 |
| UCEC | 5.555 | 3.319 | 7.42 |
| UCS | 0.319 | 0.023 | 0.048 |
| UVM | 5.076 | 2.61 | 4.536 |

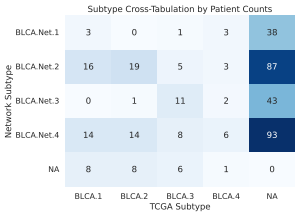
Table S2: Minimum pairwise log-rank test comparison across different subtyping methods in terms of -log(p-value). Only samples shared between all datasets are used. – indicates no samples are shared, or subtypes do not exist for TCGA.




(a) Known subtypes survival function




(b) Network subtypes survival function




(c) Network subtypes crosstabulation with known subtype

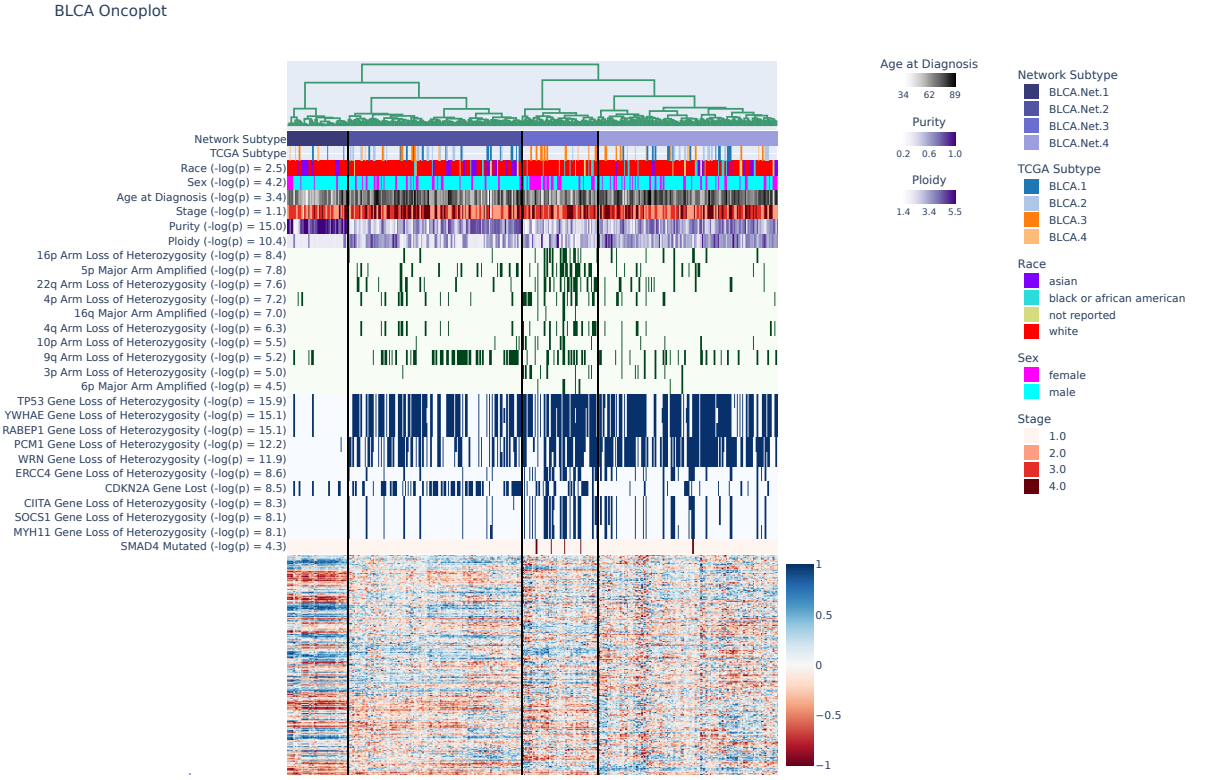
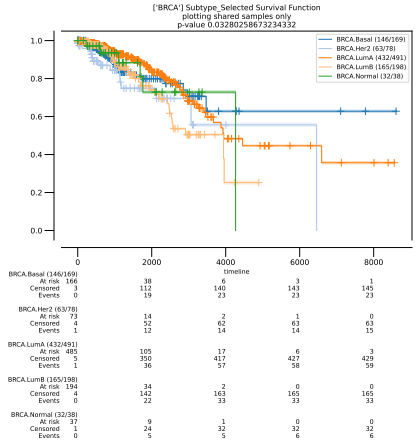
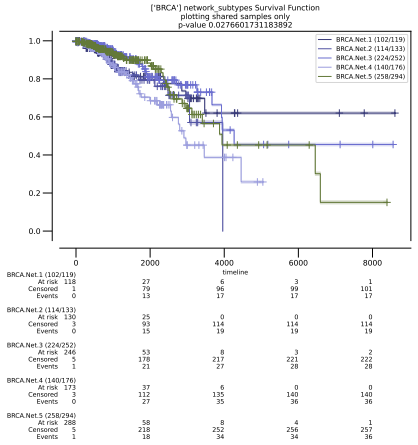


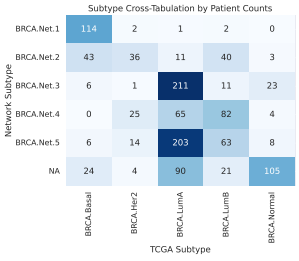

Figure S1: Exploration of network subtypes for Bladder Urothelial Carcinoma (BLCA), looking at correlated clinical information, arm-level copy alterations, gene-level copy alterations, and gene-level single nucleotide variations.




(a) Known subtypes survival function




(b) Network subtypes survival function




(c) Network subtypes crosstabulation with known subtype

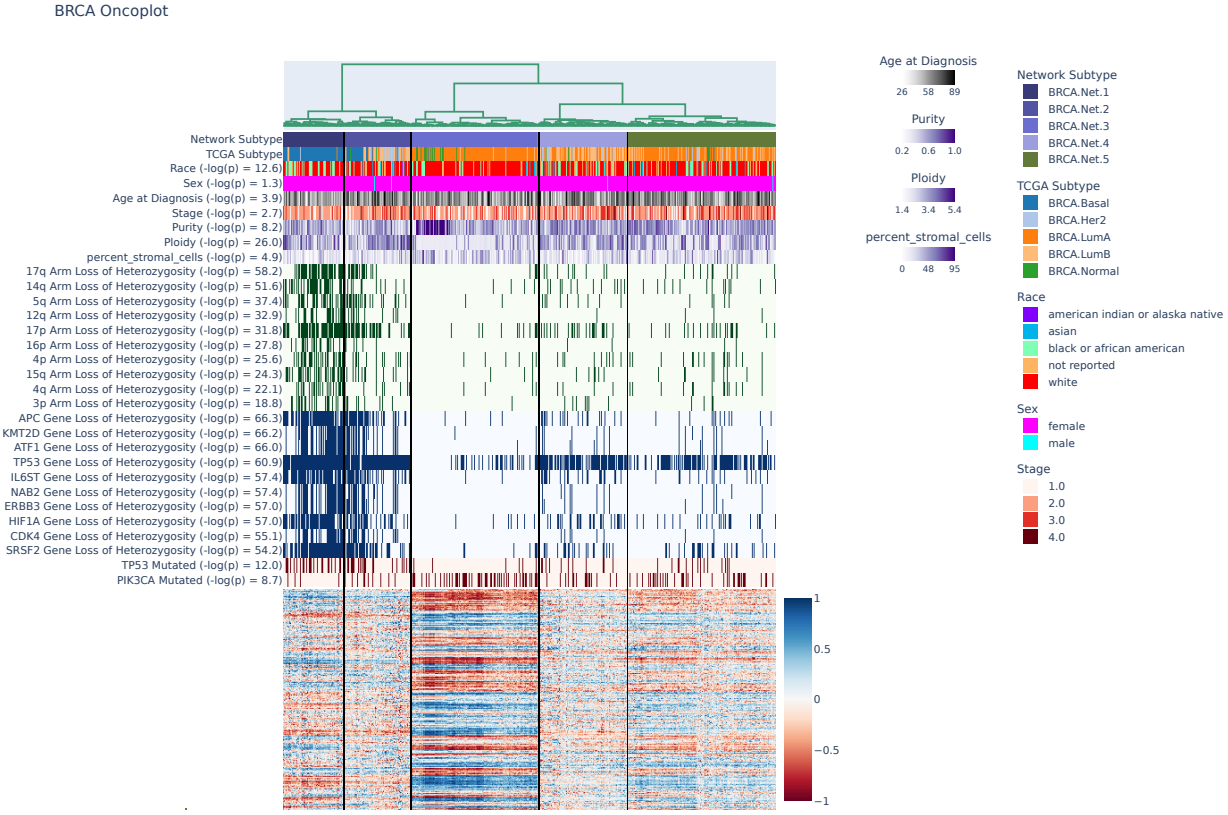
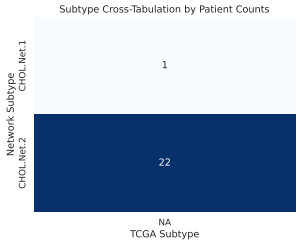



Figure S2: Exploration of network subtypes for Breast invasive carcinoma (BRCA), looking at correlated clinical information, arm-level copy alterations, gene-level copy alterations, and gene-level single nucleotide variations.




(a) Network subtypes crosstabulation with known subtype

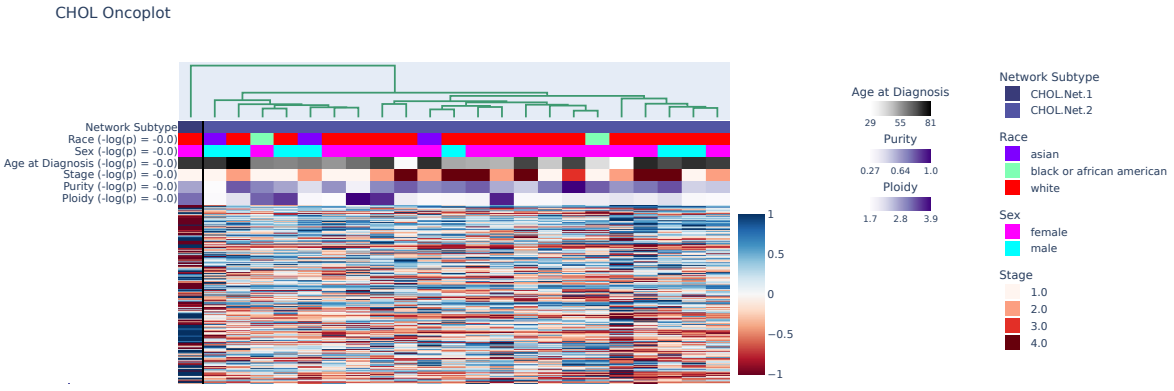
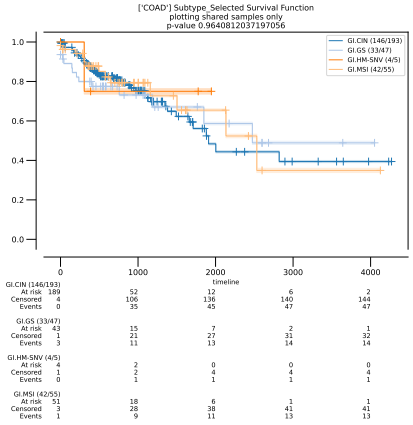
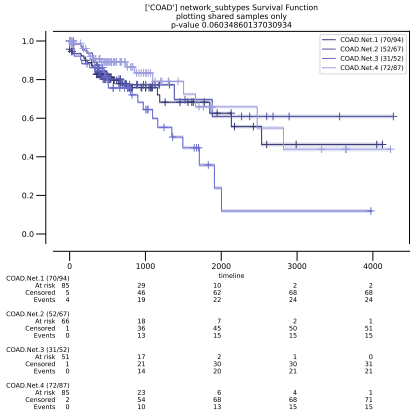


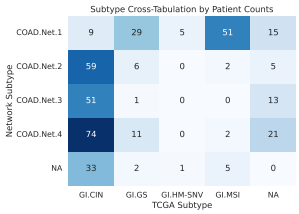

Figure S3: Exploration of network subtypes for Cholangiocarcinoma (CHOL), looking at correlated clinical information, arm-level copy alterations, gene-level copy alterations, and gene-level single nucleotide variations.




(a) Known subtypes survival function




(b) Network subtypes survival function




(c) Network subtypes crosstabulation with known subtype

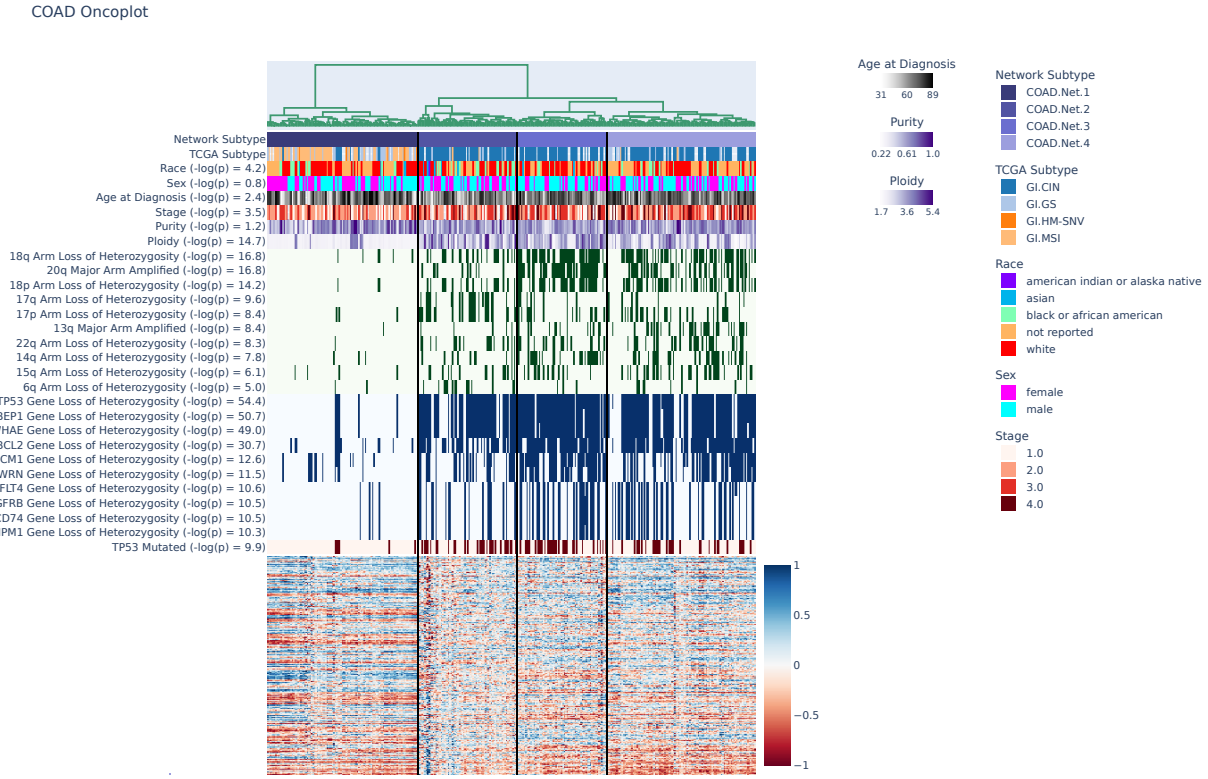
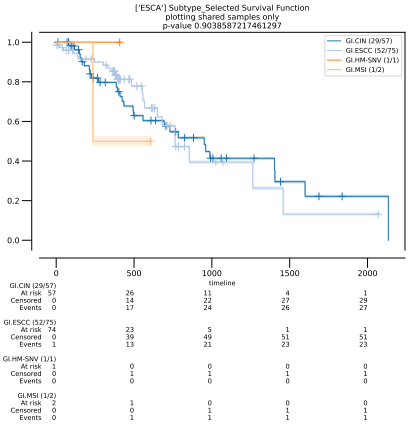
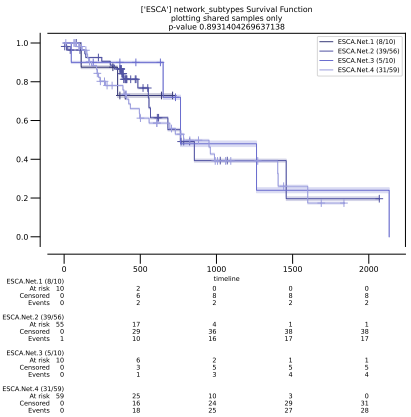


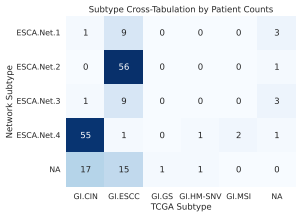

Figure S4: Exploration of network subtypes for Colon adenocarcinoma (COAD), looking at correlated clinical information, arm-level copy alterations, gene-level copy alterations, and gene-level single nucleotide variations.




(a) Known subtypes survival function




(b) Network subtypes survival function




(c) Network subtypes crosstabulation with known subtype

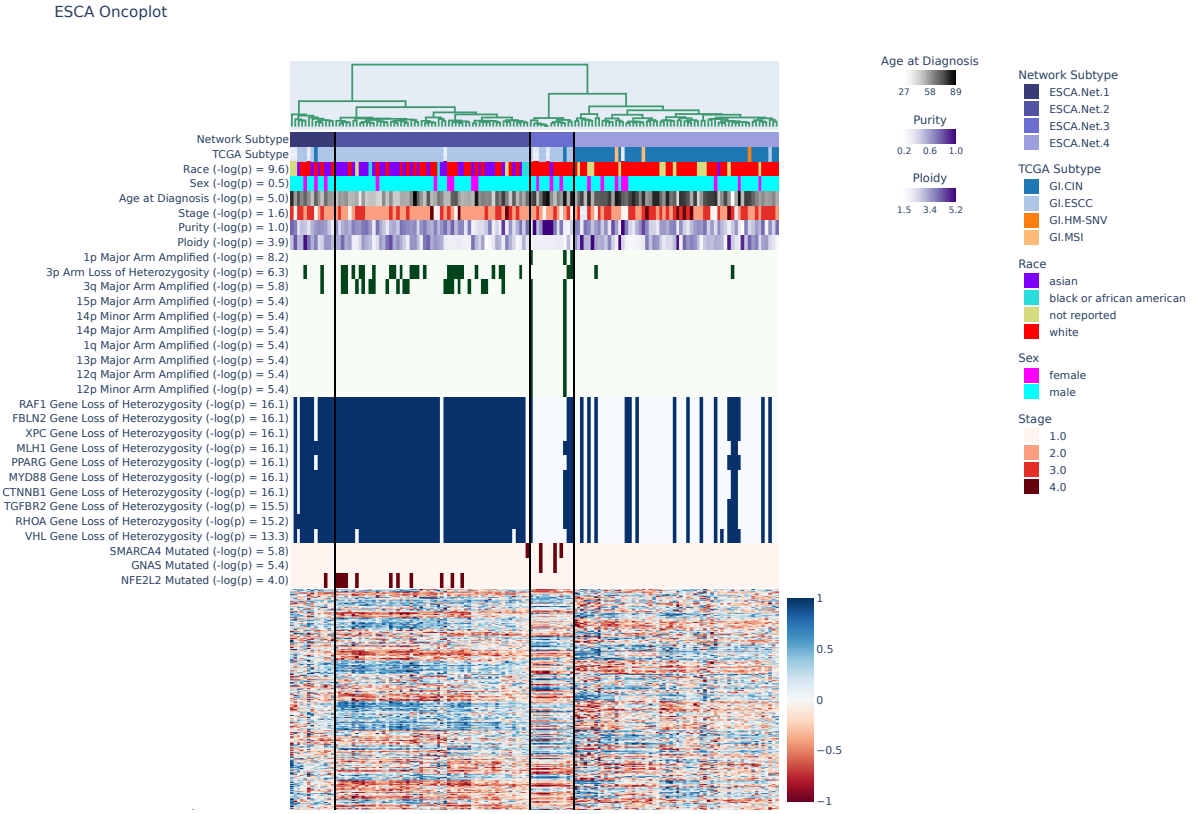
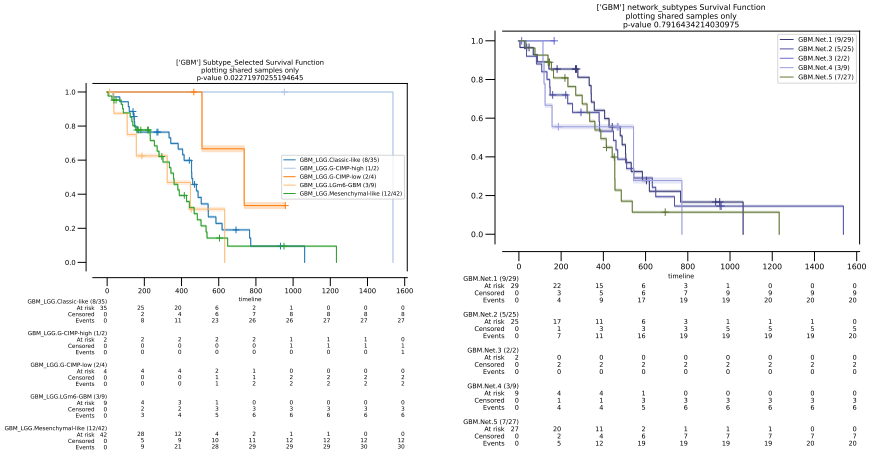



Figure S5: Exploration of network subtypes for Esophageal carcinoma (ESCA), looking at correlated clinical information, arm-level copy alterations, gene-level copy alterations, and gene-level single nucleotide variations.




(a) Known subtypes survival function

(b) Network subtypes survival function

(c) Network subtypes crosstabulation with known subtype

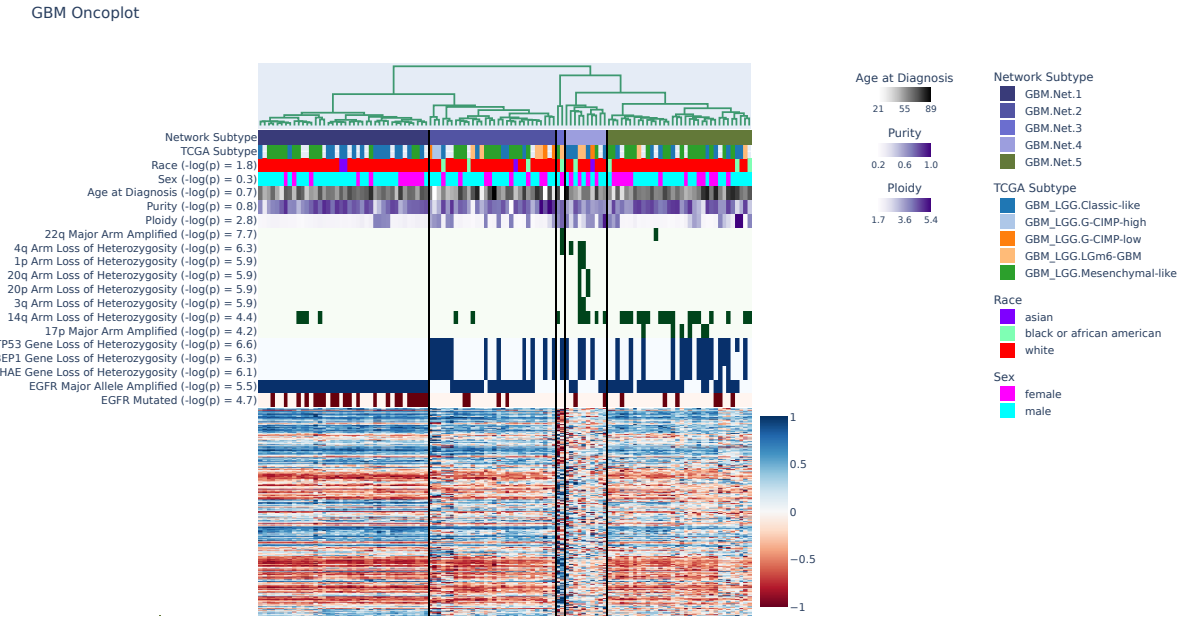
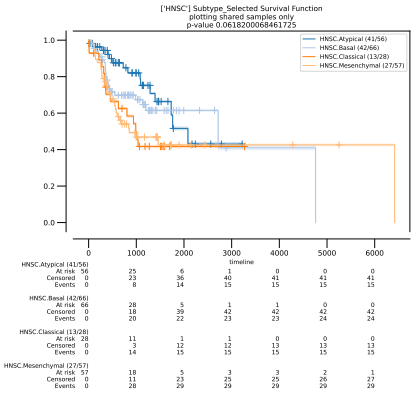
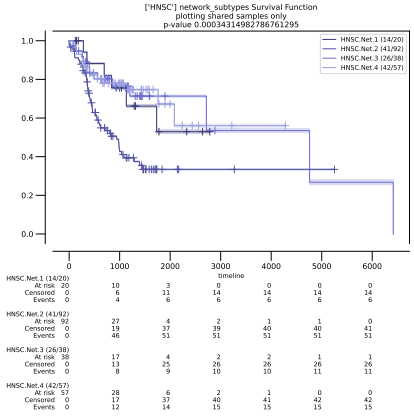


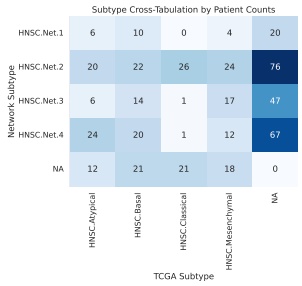

Figure S6: Exploration of network subtypes for Glioblastoma multiforme (GBM), looking at correlated clinical information, arm-level copy alterations, gene-level copy alterations, and gene-level single nucleotide variations.




(a) Known subtypes survival function




(b) Network subtypes survival function




(c) Network subtypes crosstabulation with known subtype

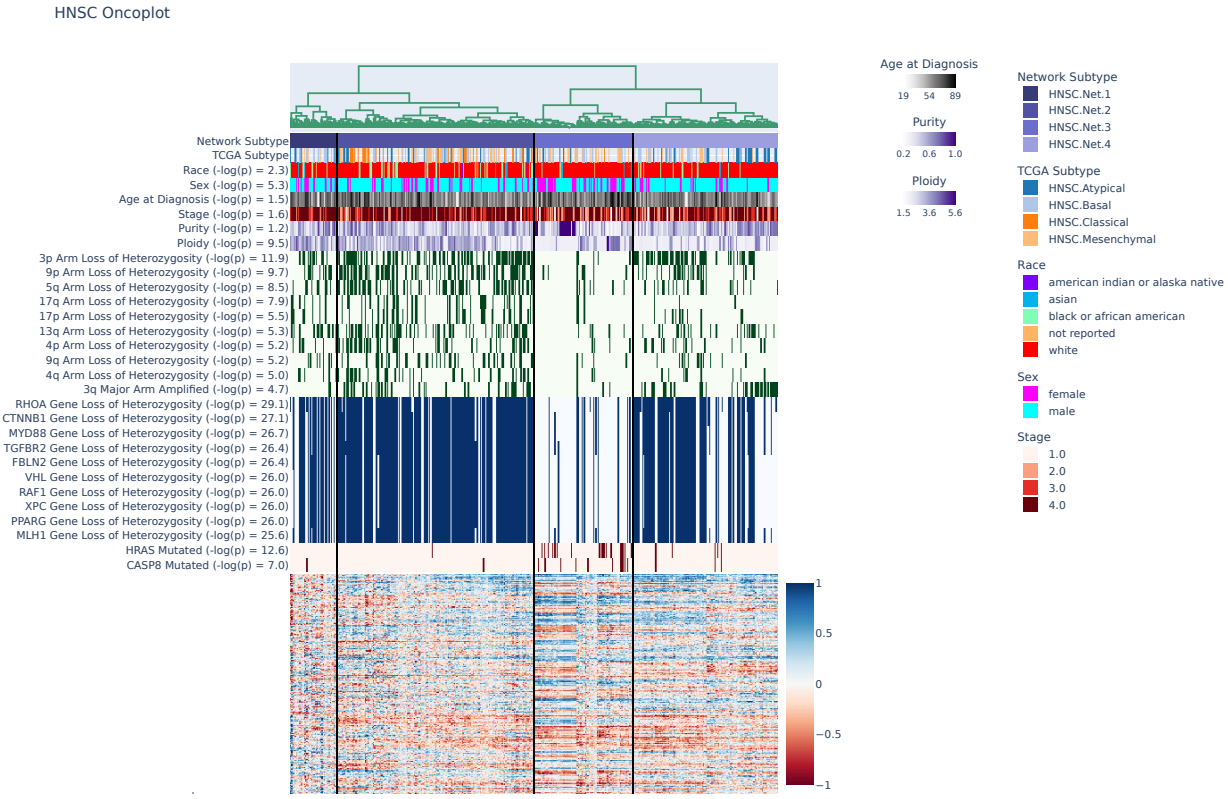
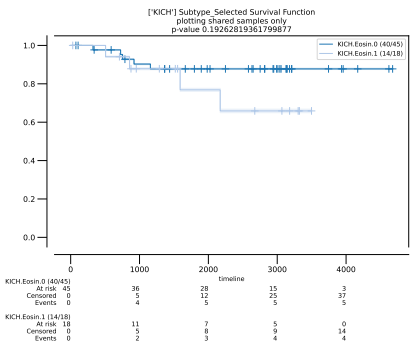
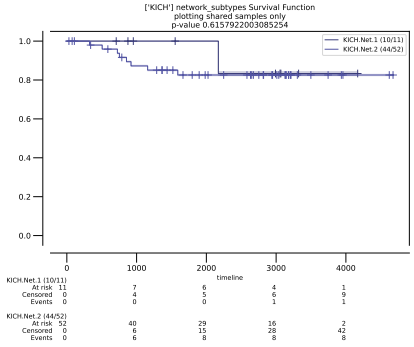


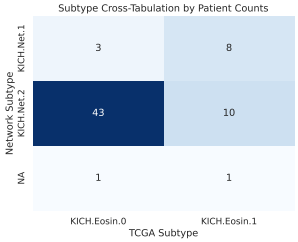

Figure S7: Exploration of network subtypes for Head and Neck squamous cell carcinoma (HNSC), looking at correlated clinical information, arm-level copy alterations, gene-level copy alterations, and gene-level single nucleotide variations.




(a) Known subtypes survival function




(b) Network subtypes survival function




(c) Network subtypes crosstabulation with known subtype

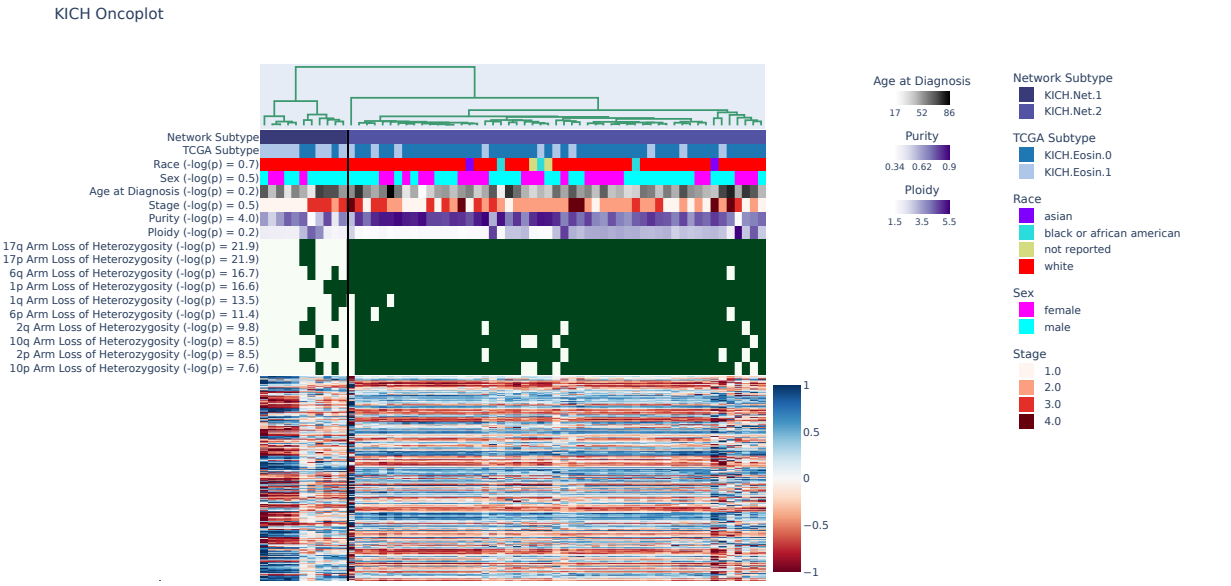
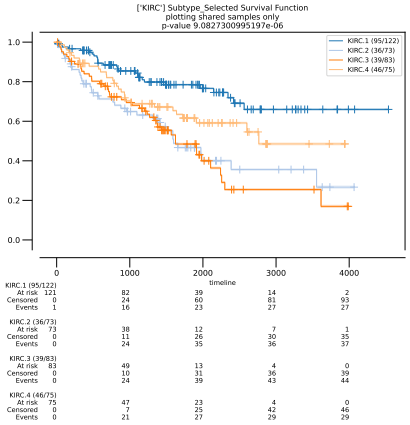
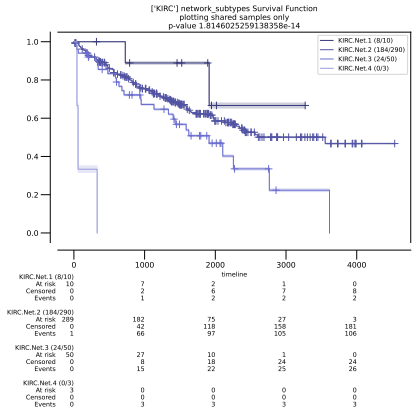


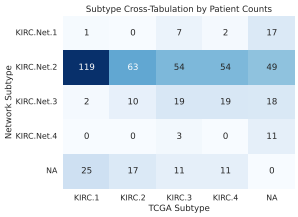

Figure S8: Exploration of network subtypes for Kidney Chromophobe (KICH), looking at correlated clinical information, arm-level copy alterations, gene-level copy alterations, and gene-level single nucleotide variations.




(a) Known subtypes survival function




(b) Network subtypes survival function




(c) Network subtypes crosstabulation with known subtype

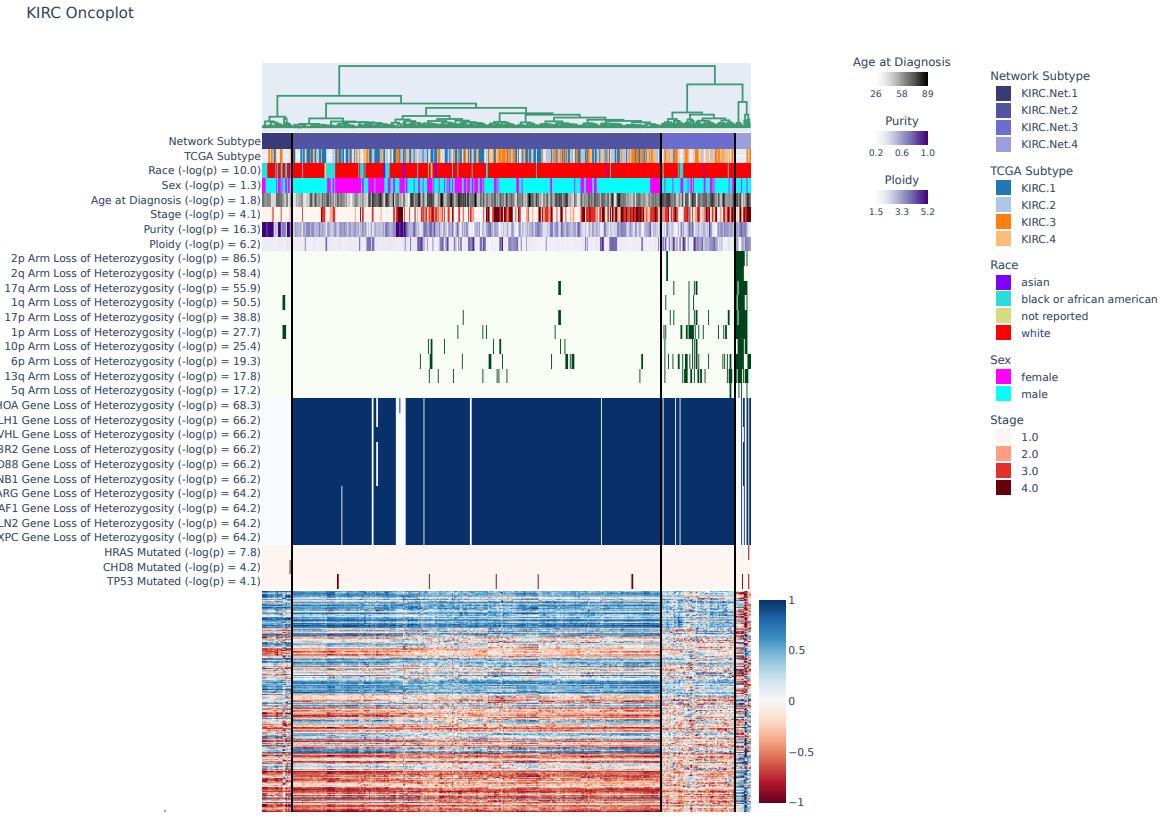
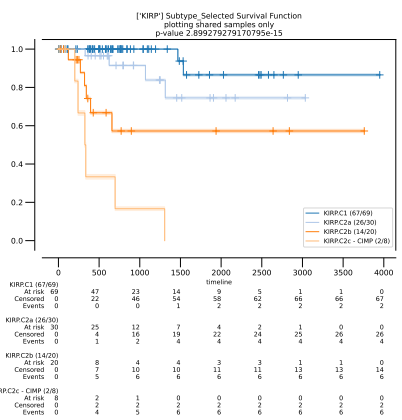
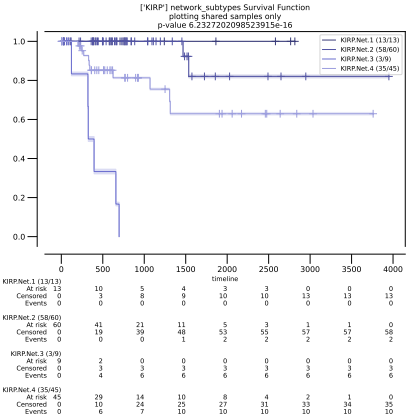


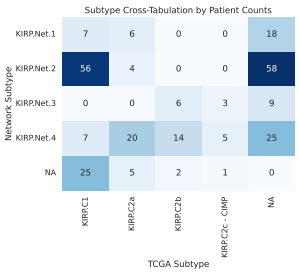

Figure S9: Exploration of network subtypes for Kidney renal clear cell carcinoma (KIRC), looking at correlated clinical information, arm-level copy alterations, gene-level copy alterations, and gene-level single nucleotide variations.




(a) Known subtypes survival function




(b) Network subtypes survival function




(c) Network subtypes crosstabulation with known subtype

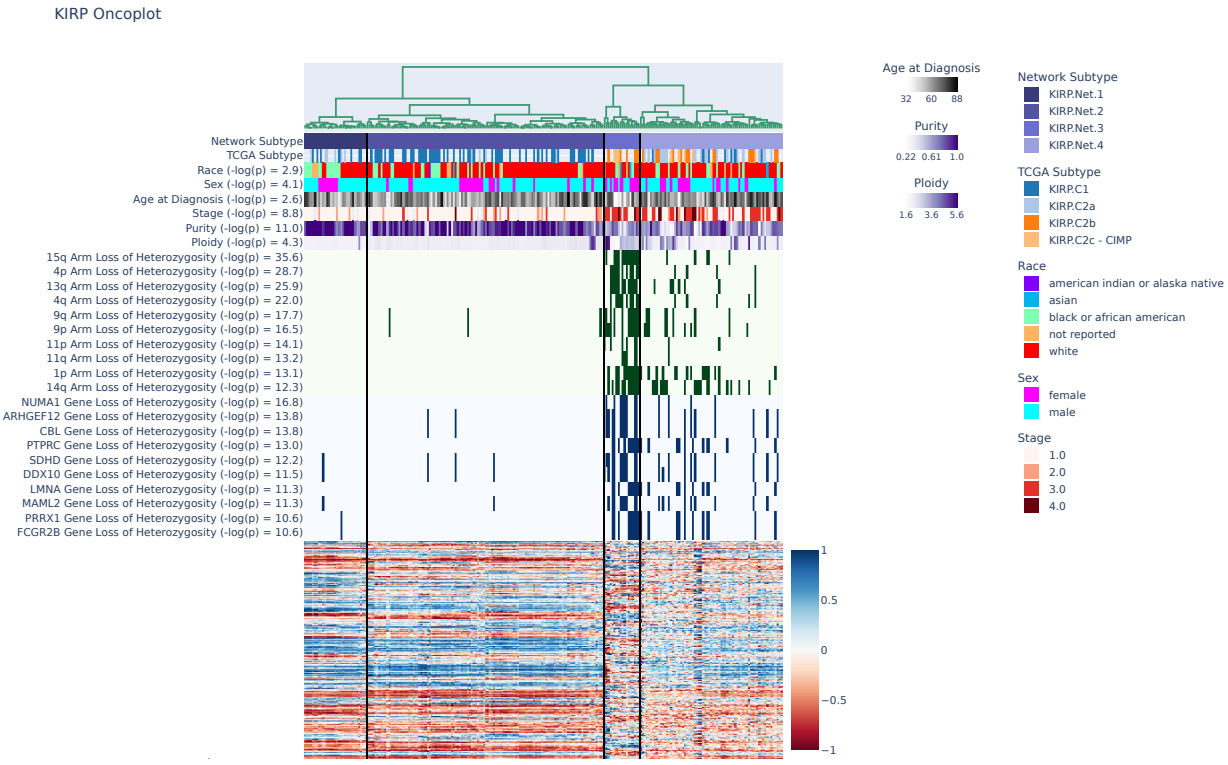
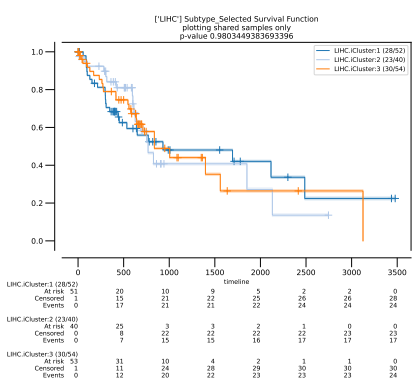
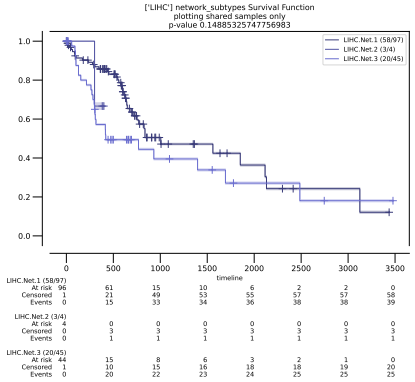


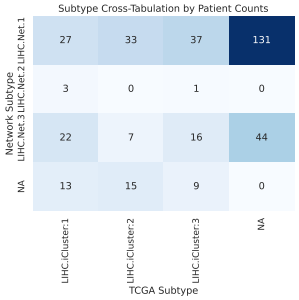

Figure S10: Exploration of network subtypes for Kidney renal papillary cell carcinoma (KIRP), looking at correlated clinical information, arm-level copy alterations, gene-level copy alterations, and gene-level single nucleotide variations.




(a) Known subtypes survival function




(b) Network subtypes survival function




(c) Network subtypes crosstabulation with known subtype

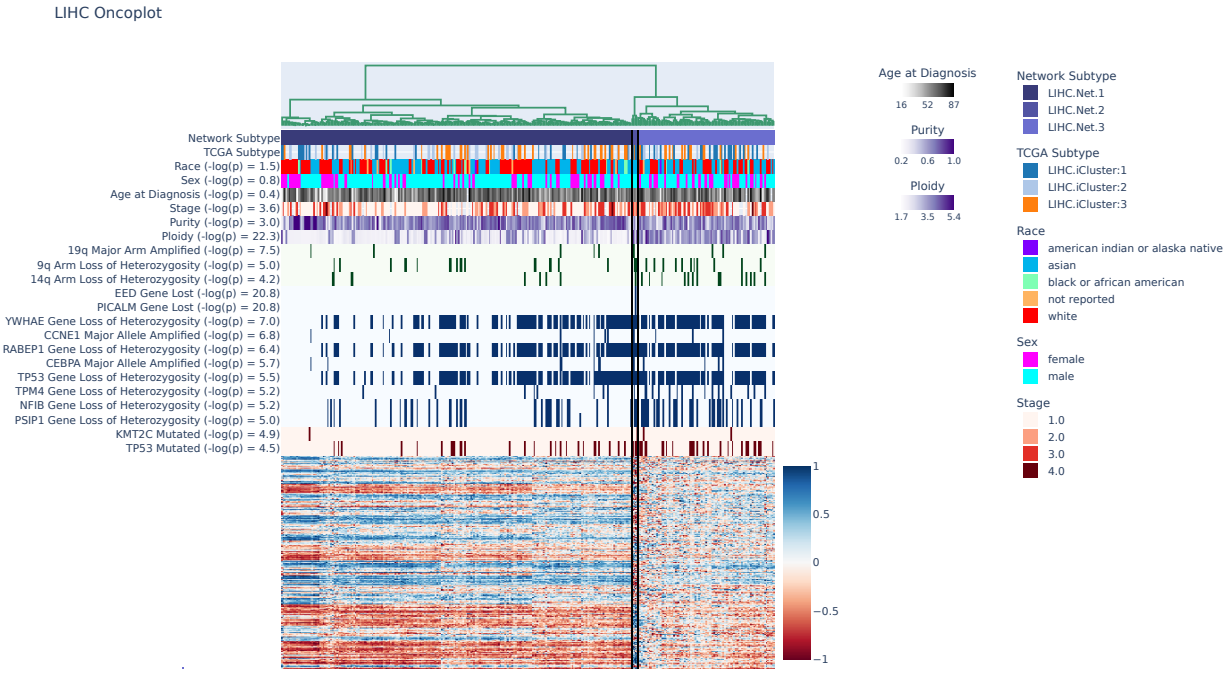
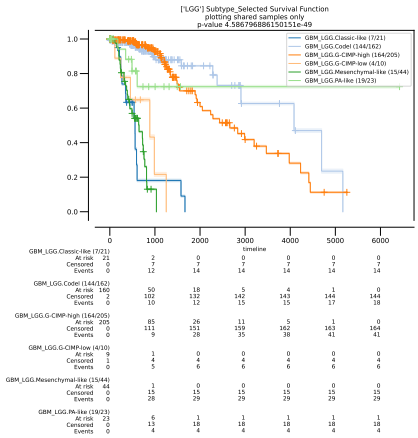
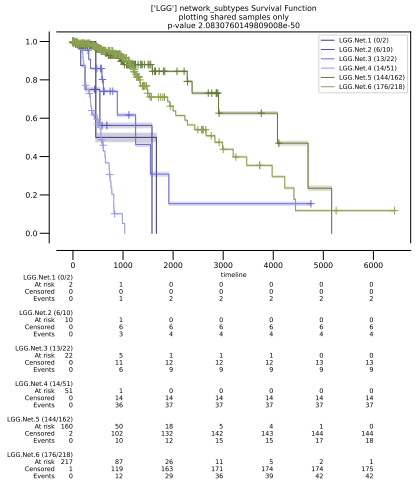


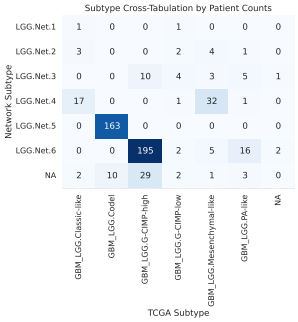

Figure S11: Exploration of network subtypes for Liver hepatocellular carcinoma (LIHC), looking at correlated clinical information, arm-level copy alterations, gene-level copy alterations, and gene-level single nucleotide variations.




(a) Known subtypes survival function




(b) Network subtypes survival function




(c) Network subtypes crosstabulation with known subtype

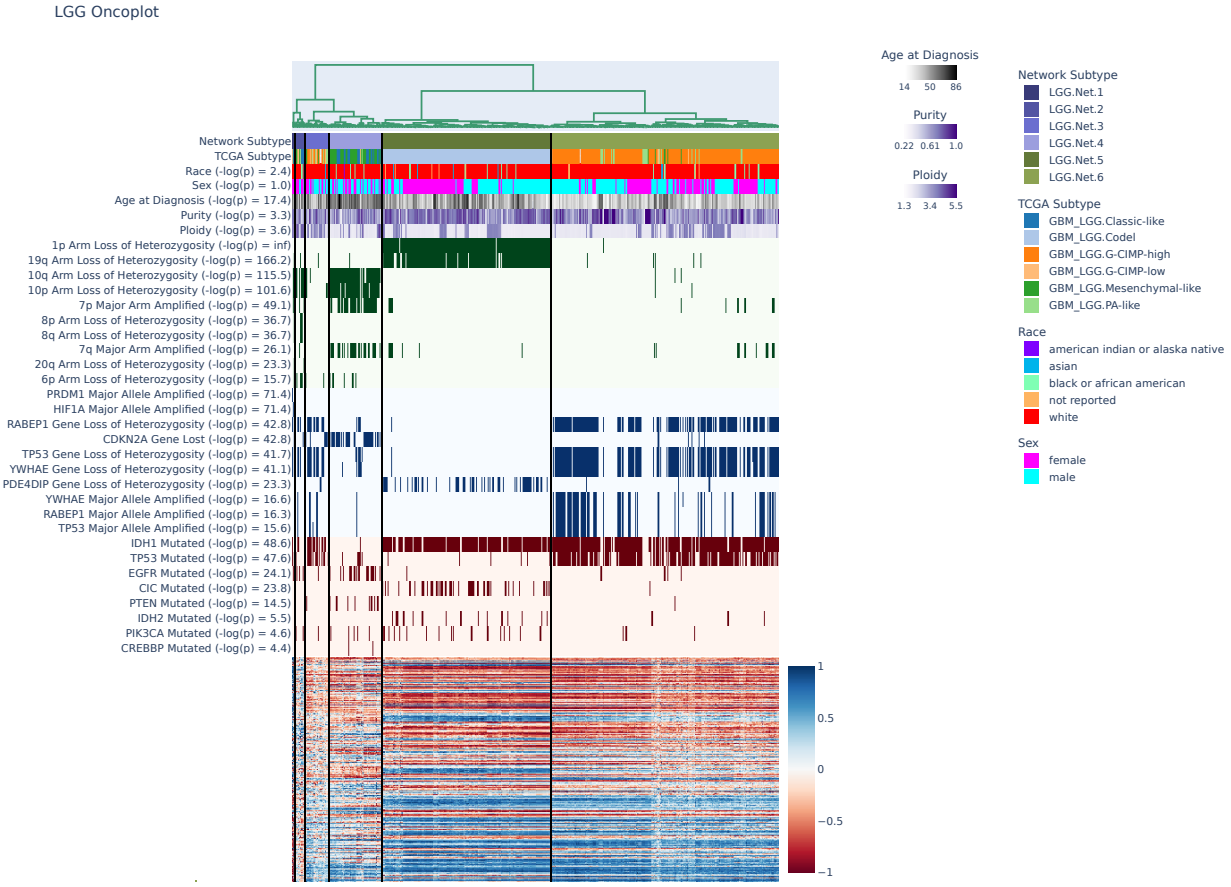
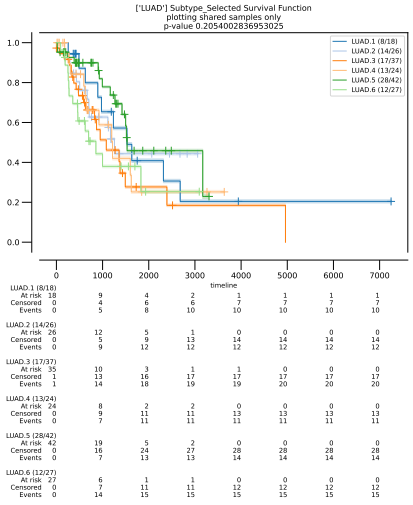
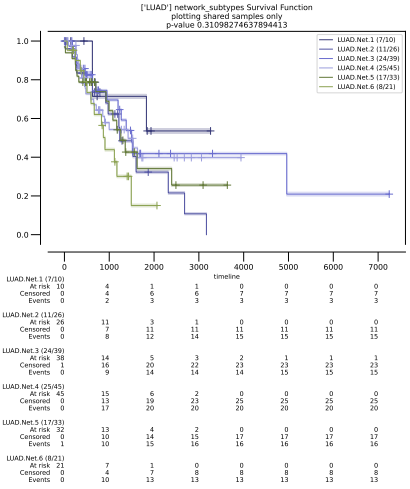


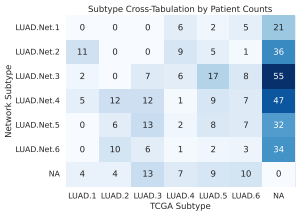

Figure S12: Exploration of network subtypes for Brain Lower Grade Glioma (LGG), looking at correlated clinical information, arm-level copy alterations, gene-level copy alterations, and gene-level single nucleotide variations.




(a) Known subtypes survival function




(b) Network subtypes survival function




(c) Network subtypes crosstabulation with known subtype

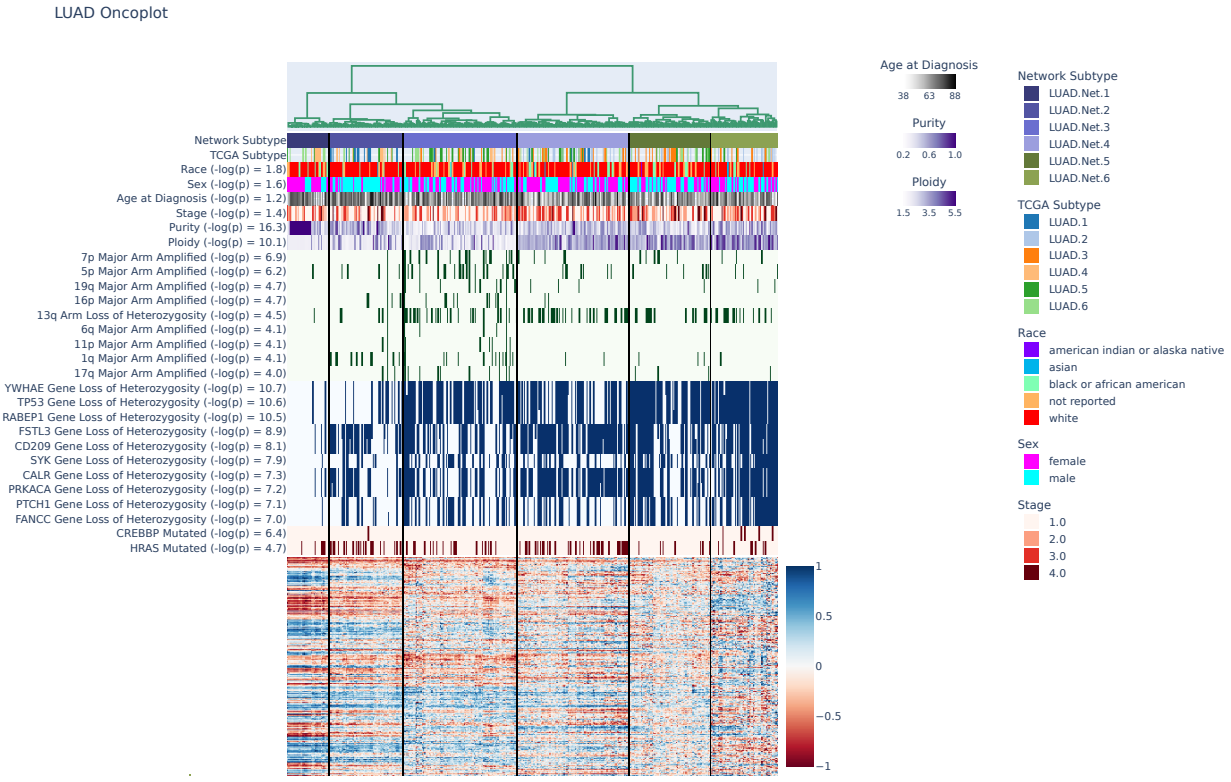
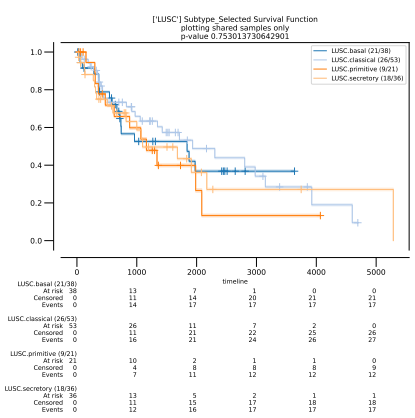
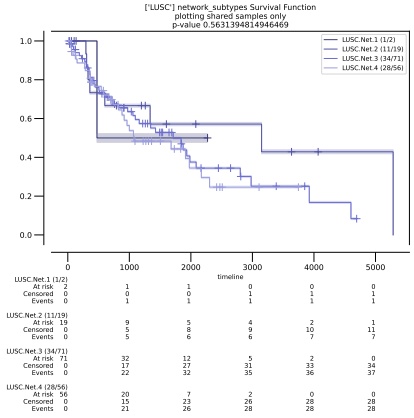


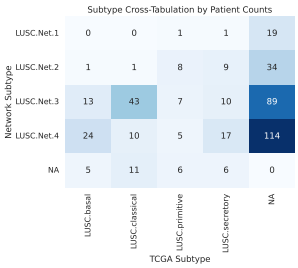

Figure S13: Exploration of network subtypes for Lung adenocarcinoma (LUAD), looking at correlated clinical information, arm-level copy alterations, gene-level copy alterations, and gene-level single nucleotide variations.




(a) Known subtypes survival function




(b) Network subtypes survival function




(c) Network subtypes crosstabulation with known subtype

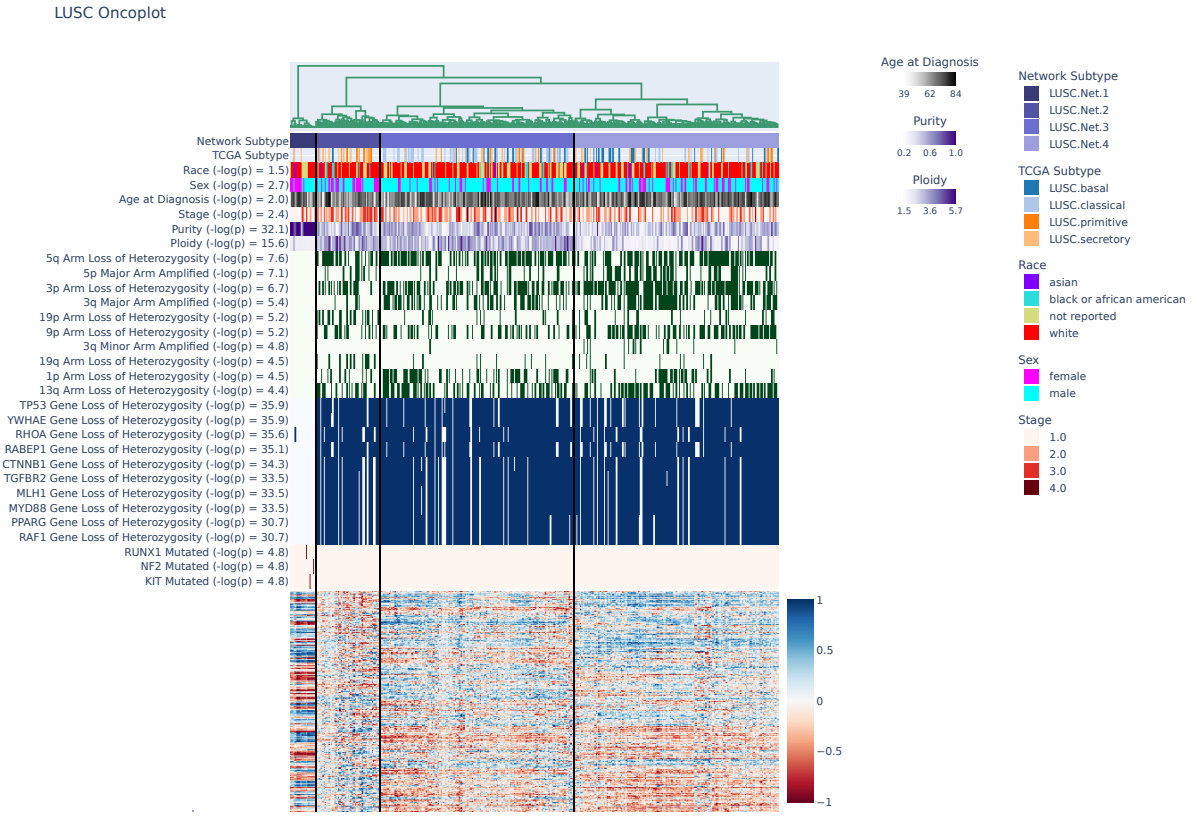
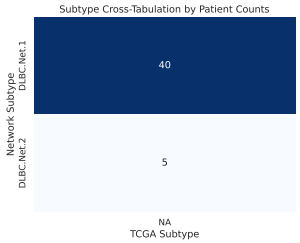



Figure S14: Exploration of network subtypes for Lung squamous cell carcinoma (LUSC), looking at correlated clinical information, arm-level copy alterations, gene-level copy alterations, and gene-level single nucleotide variations.




(a) Network subtypes crosstabulation with known subtype

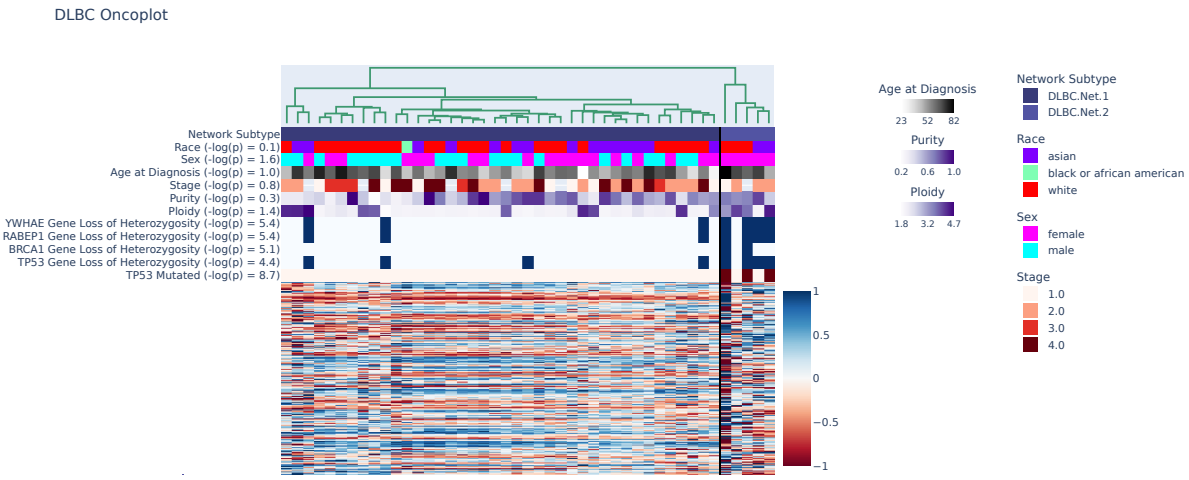
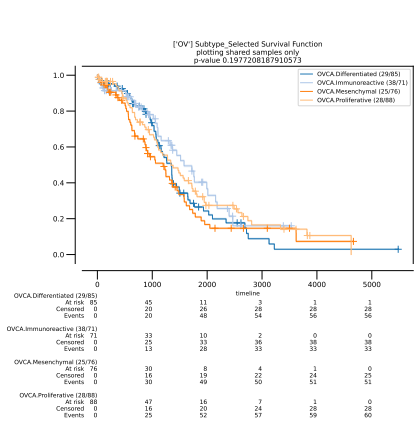
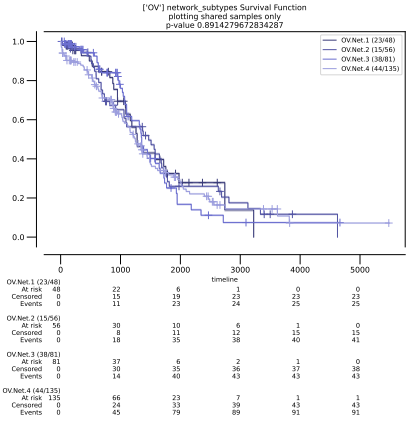


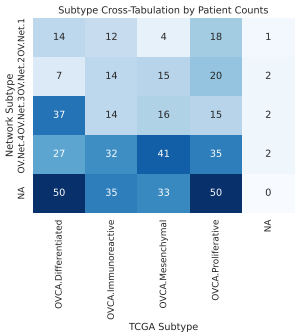

Figure S15: Exploration of network subtypes for Lymphoid Neoplasm Diffuse Large B-cell Lymphoma (DLBC), looking at correlated clinical information, arm-level copy alterations, gene-level copy alterations, and gene-level single nucleotide variations.




(a) Known subtypes survival function




(b) Network subtypes survival function




(c) Network subtypes crosstabulation with known subtype

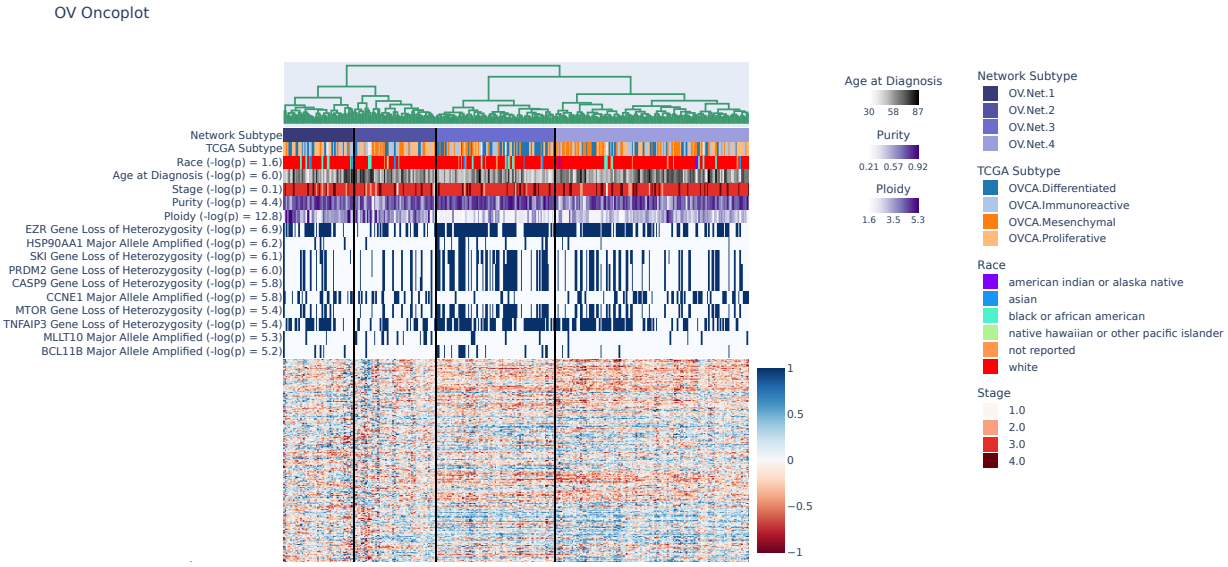
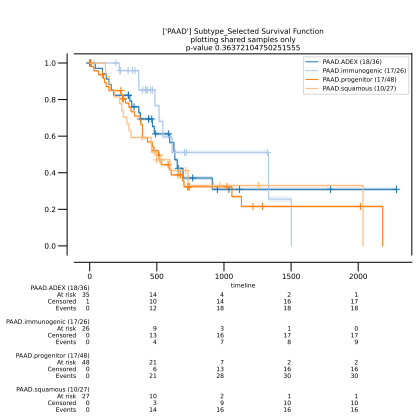
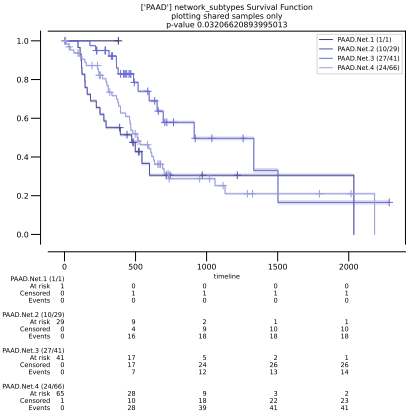


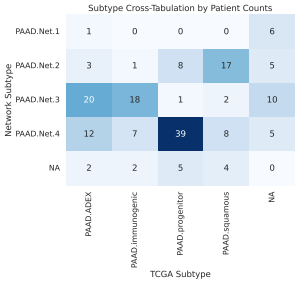

Figure S16: Exploration of network subtypes for Ovarian serous cystadenocarcinoma (OV), looking at correlated clinical information, arm-level copy alterations, gene-level copy alterations, and gene-level single nucleotide variations.




(a) Known subtypes survival function




(b) Network subtypes survival function




(c) Network subtypes crosstabulation with known subtype

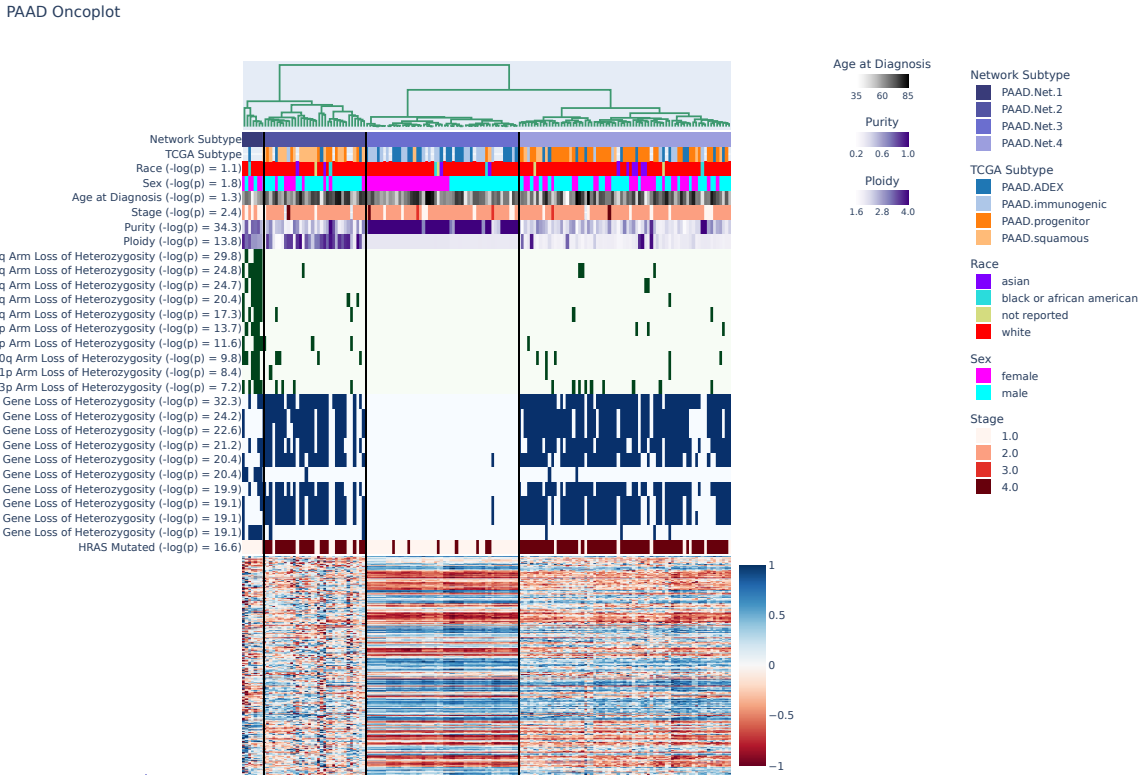
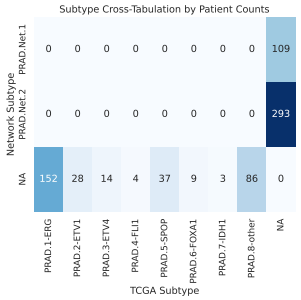



Figure S17: Exploration of network subtypes for Pancreatic adenocarcinoma (PAAD), looking at correlated clinical information, arm-level copy alterations, gene-level copy alterations, and gene-level single nucleotide variations.




(a) Network subtypes crosstabulation with known subtype

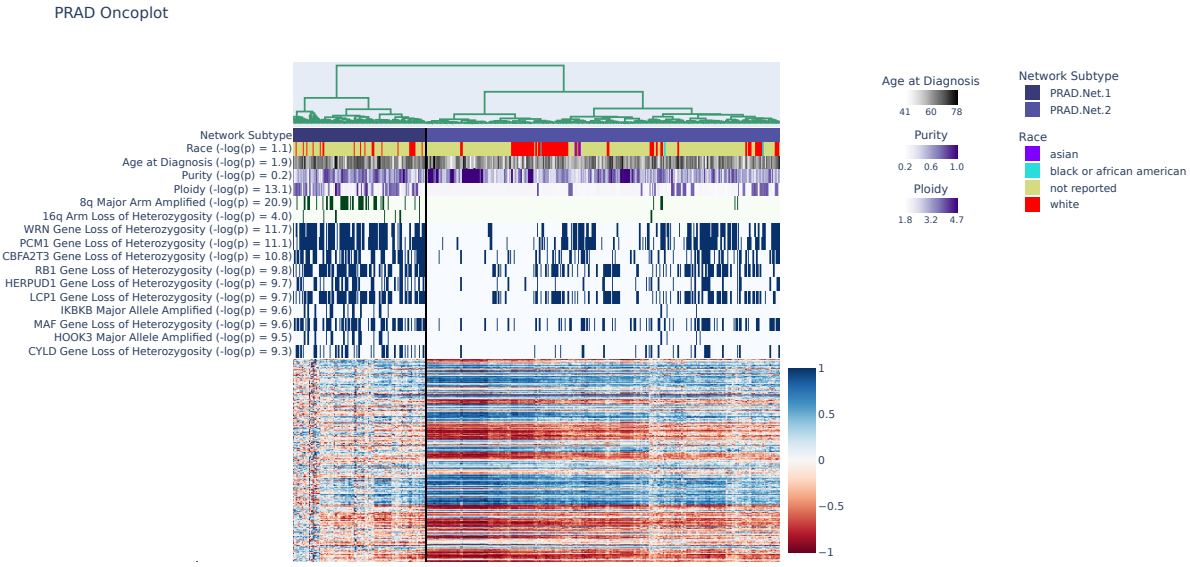
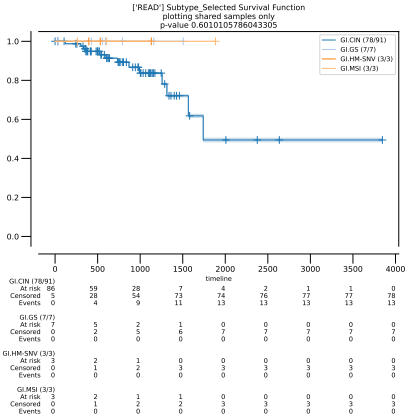
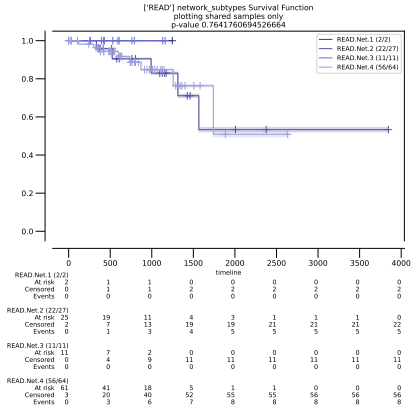


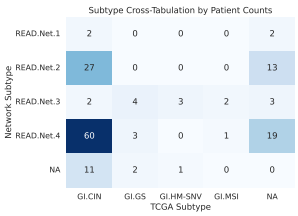

Figure S18: Exploration of network subtypes for Prostate adenocarcinoma (PRAD), looking at correlated clinical information, arm-level copy alterations, gene-level copy alterations, and gene-level single nucleotide variations.




(a) Known subtypes survival function




(b) Network subtypes survival function




(c) Network subtypes crosstabulation with known subtype

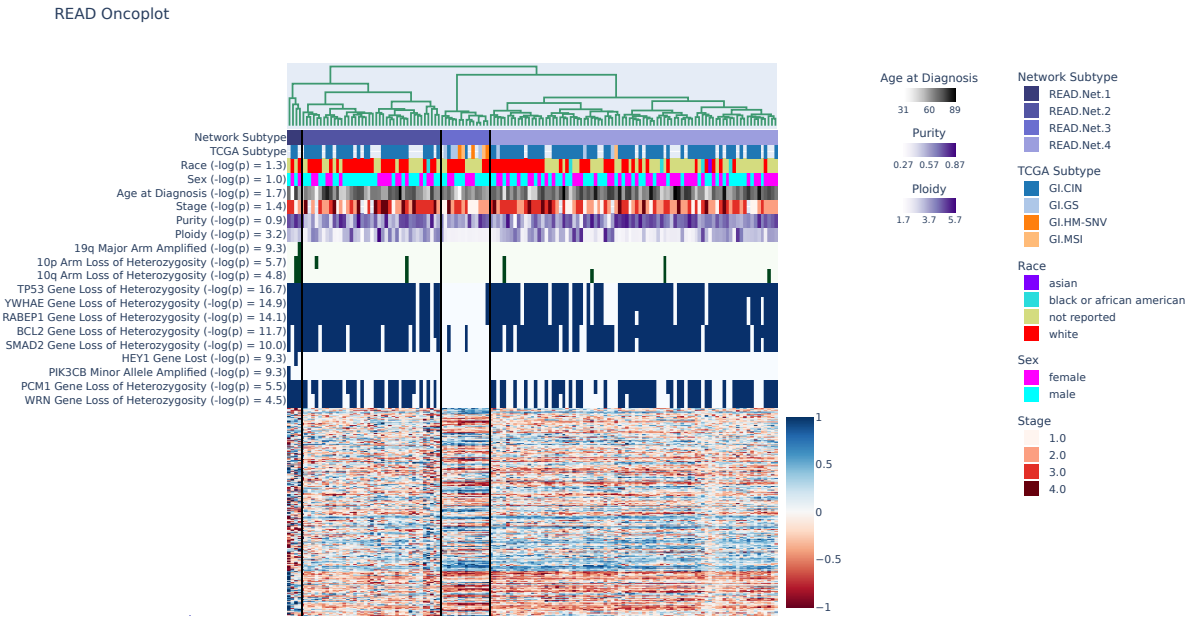
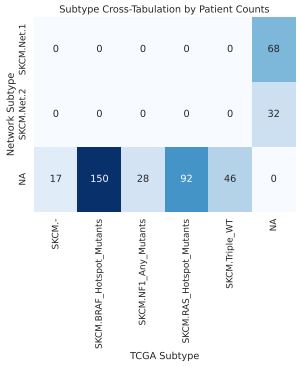



Figure S19: Exploration of network subtypes for Rectum adenocarcinoma (READ), looking at correlated clinical information, arm-level copy alterations, gene-level copy alterations, and gene-level single nucleotide variations.




(a) Network subtypes crosstabulation with known subtype

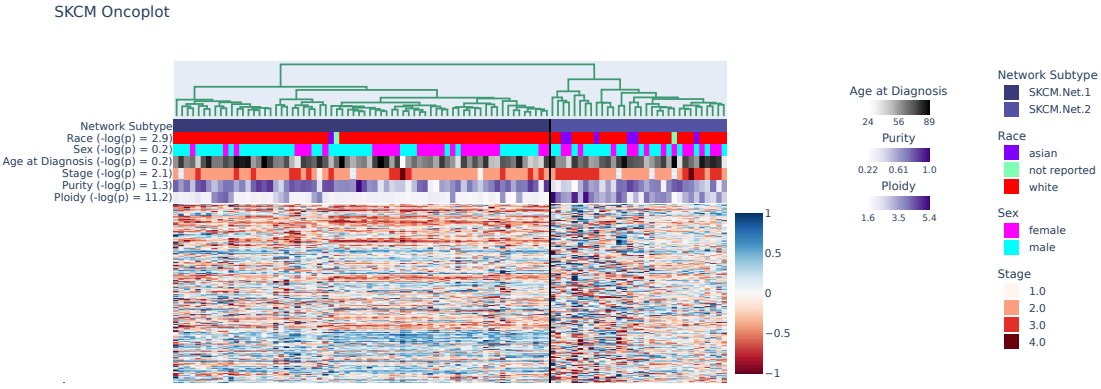
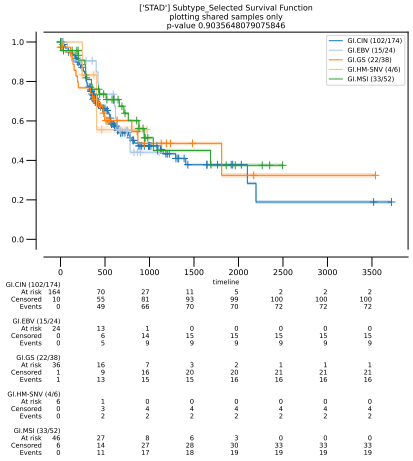
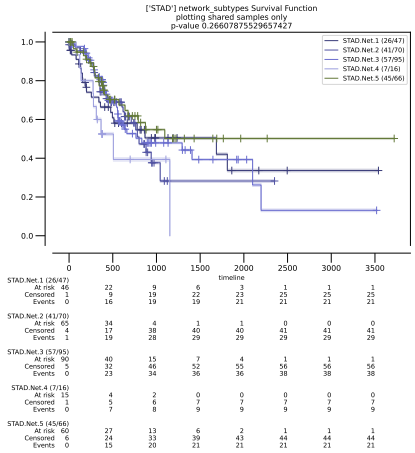


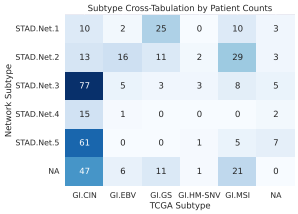

Figure S20: Exploration of network subtypes for Skin Cutaneous Melanoma (SKCM), looking at correlated clinical information, arm-level copy alterations, gene-level copy alterations, and gene-level single nucleotide variations.




(a) Known subtypes survival function




(b) Network subtypes survival function




(c) Network subtypes crosstabulation with known subtype

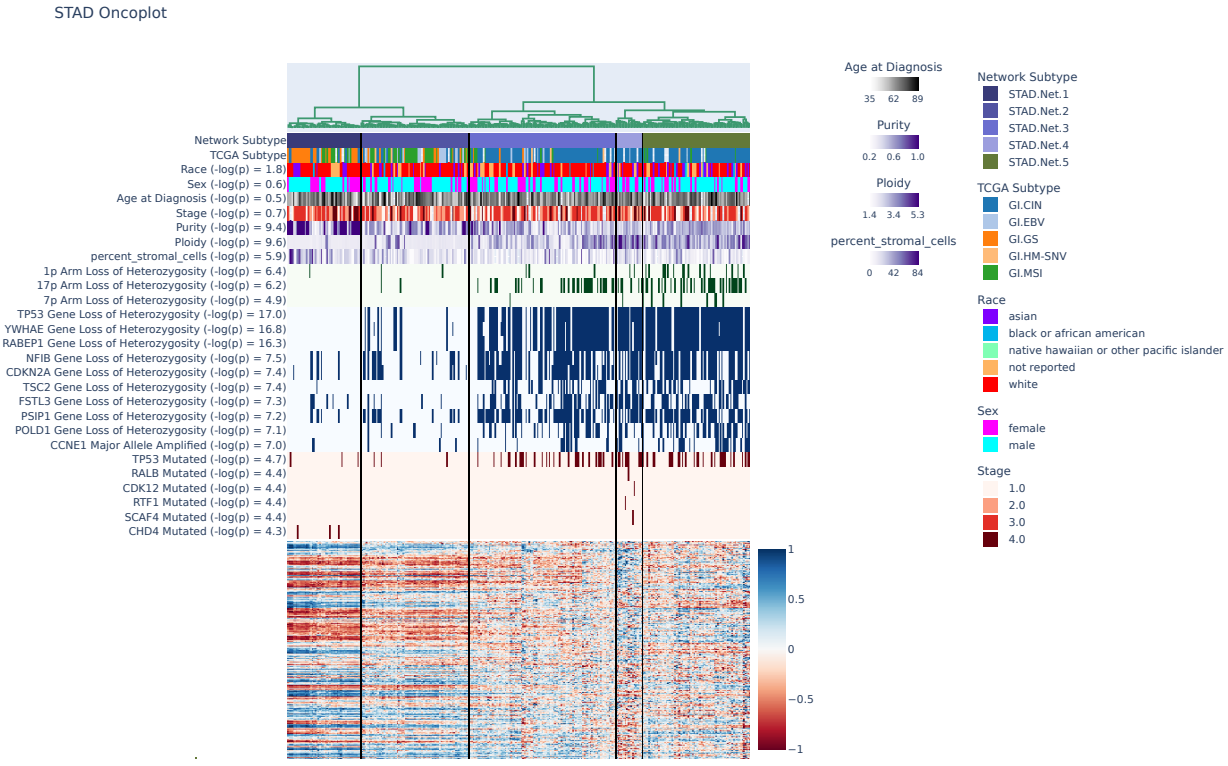
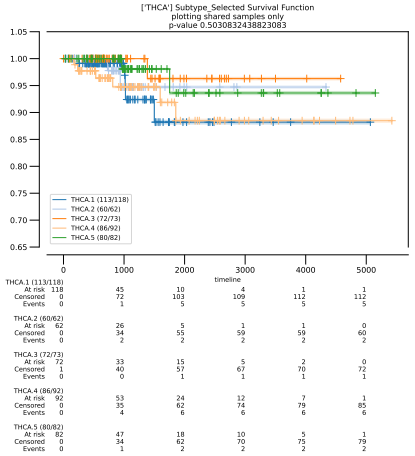
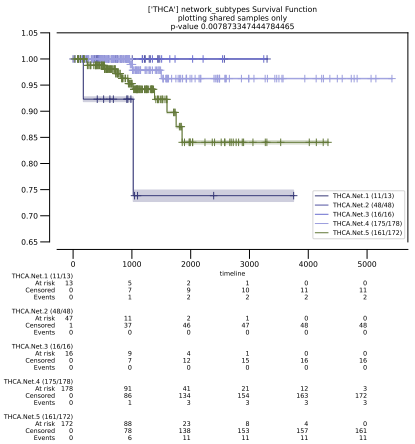


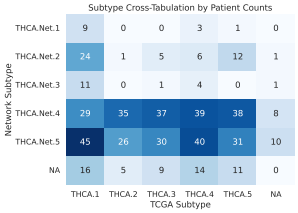

Figure S21: Exploration of network subtypes for Stomach adenocarcinoma (STAD), looking at correlated clinical information, arm-level copy alterations, gene-level copy alterations, and gene-level single nucleotide variations.




(a) Known subtypes survival function




(b) Network subtypes survival function




(c) Network subtypes crosstabulation with known subtype

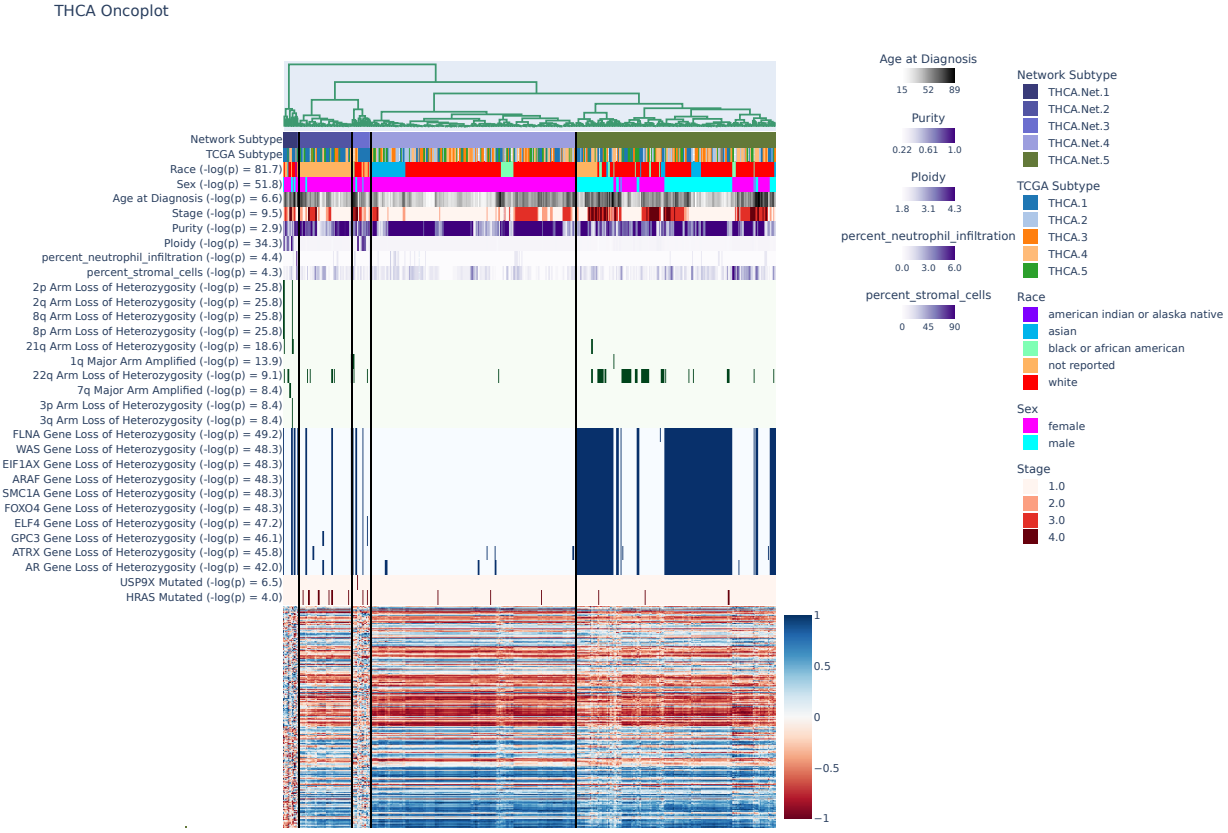
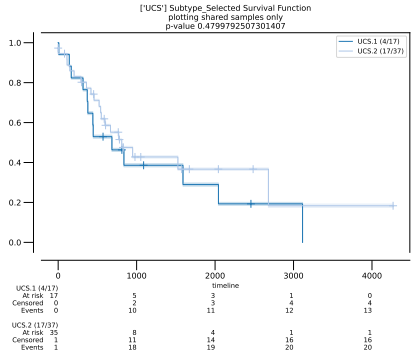
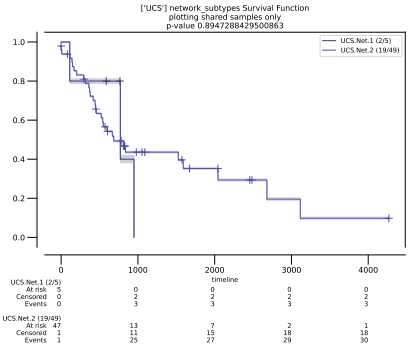


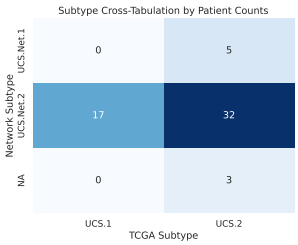

Figure S22: Exploration of network subtypes for Thyroid carcinoma (THCA), looking at correlated clinical information, arm-level copy alterations, gene-level copy alterations, and gene-level single nucleotide variations.




(a) Known subtypes survival function




(b) Network subtypes survival function




(c) Network subtypes crosstabulation with known subtype

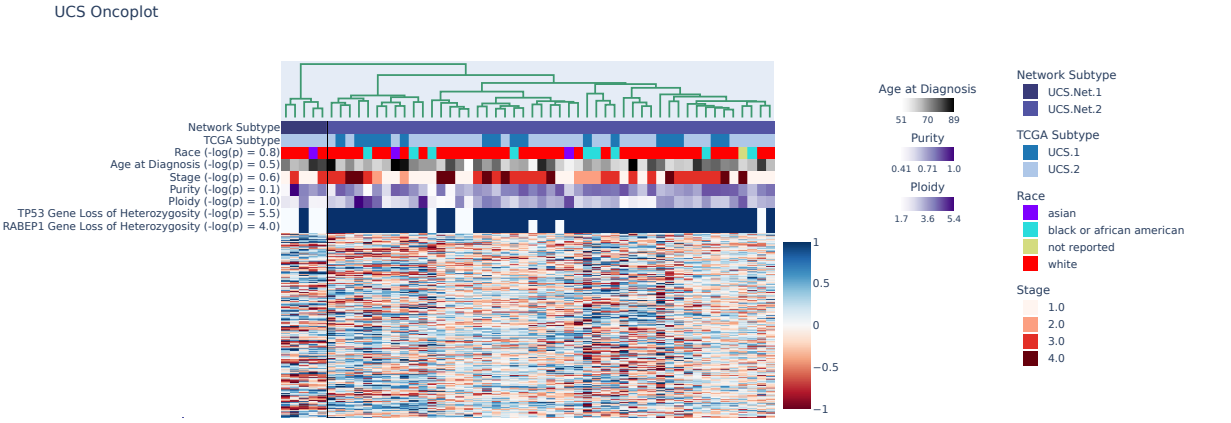
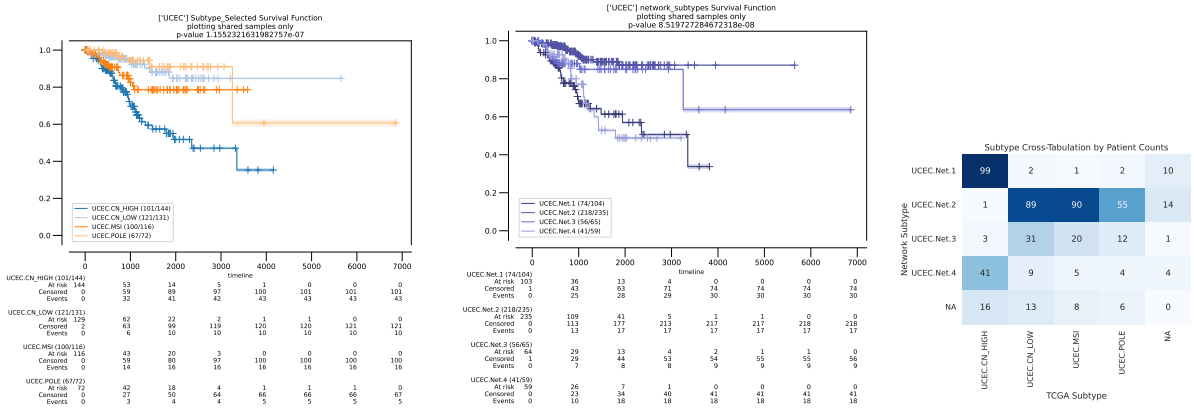



Figure S23: Exploration of network subtypes for Uterine Carcinosarcoma (UCS), looking at correlated clinical information, arm-level copy alterations, gene-level copy alterations, and gene-level single nucleotide variations.




(a) Known subtypes survival function

(b) Network subtypes survival function

(c) Network subtypes crosstabulation with known subtype

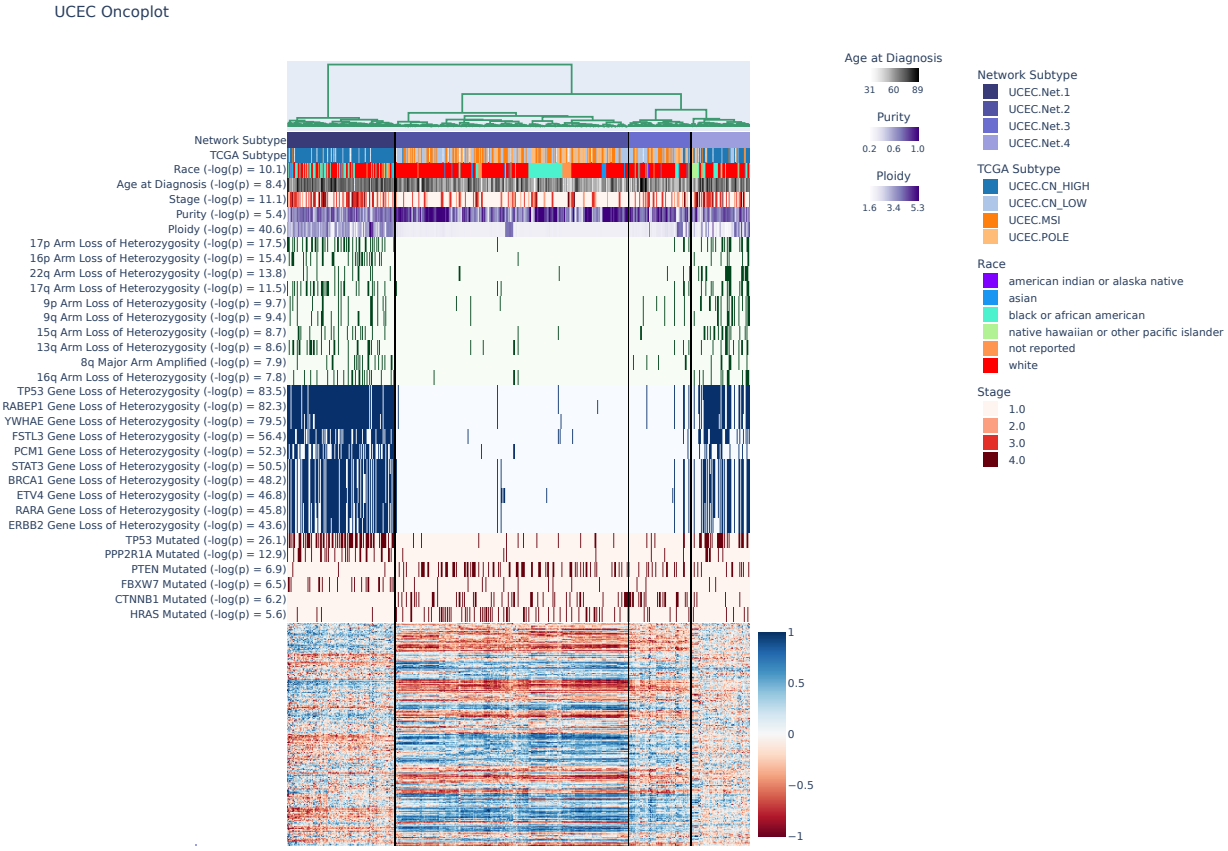
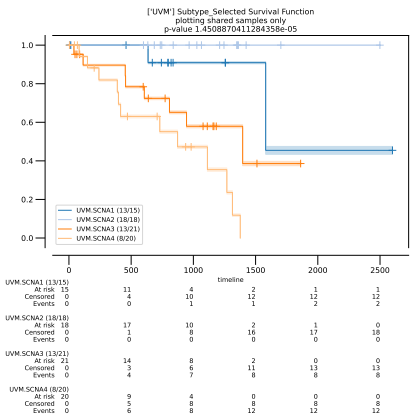
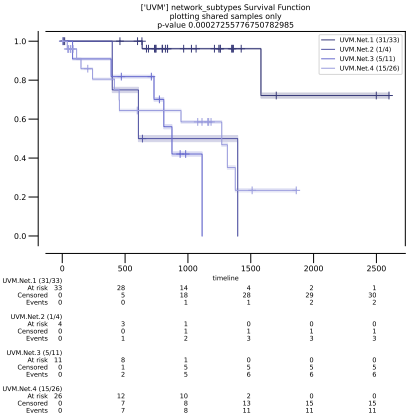


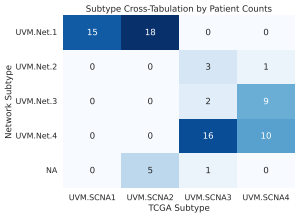

Figure S24: Exploration of network subtypes for Uterine Corpus Endometrial Carcinoma (UCEC), looking at correlated clinical information, arm-level copy alterations, gene-level copy alterations, and gene-level single nucleotide variations.




(a) Known subtypes survival function




(b) Network subtypes survival function




(c) Network subtypes crosstabulation with known subtype

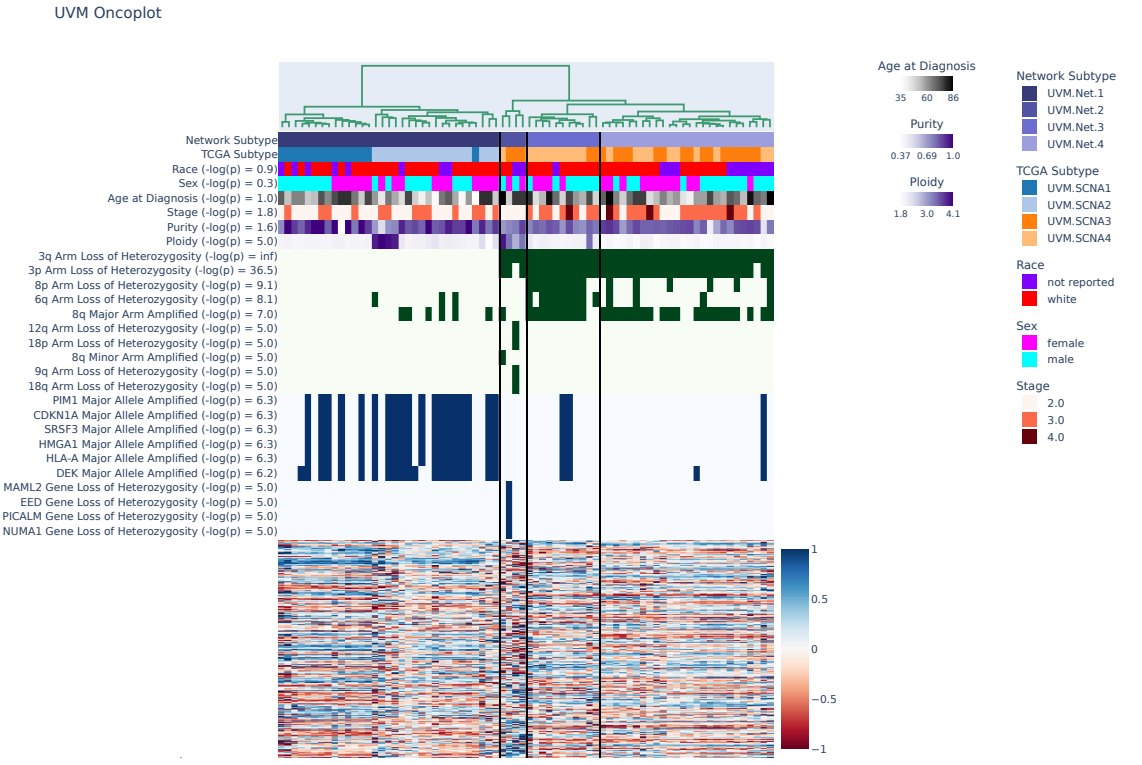



Figure S25: Exploration of network subtypes for Uveal Melanoma (UVM), looking at correlated clinical information, arm-level copy alterations, gene-level copy alterations, and gene-level single nucleotide variations.