De novo individualized disease modules reveal the synthetic penetrance of genes and inform personalized treatment regimens

Taylor Weiskittel, Choong Yong Ung, Cristina Correia, et al.

Genome Res. published online December 7, 2021
Access the most recent version at doi:10.1101/gr.275889.121

De novo individualized disease modules reveal the synthetic penetrance of genes and inform personalized treatment regimens

Authors
Taylor M. Weiskittel[1], Choong Y. Ung[1], Cristina Correia[1], Cheng Zhang[1], and Hu Li[1*]

Affiliations
1. Center for Individualized Medicine, Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic College of Medicine, Rochester, MN 55905, USA.
* Correspondence to: li.hu@mayo.edu

Running Title: Synthetic Penetrance of Individualized Modules

Abstract

Current understandings of individual disease etiology and therapeutics are limited despite great need. To fill the gap, we propose a novel computational pipeline which collects potent disease gene cooperative pathways to envision individualized disease etiology and therapies. Our algorithm constructs individualized disease modules *de novo* which enable us to elucidate the importance of mutated genes in specific patients and to understand the synthetic penetrance of these genes across patients. We reveal that importance of notorious cancer drivers *TP53* and *PIK3CA* fluctuate widely across breast cancers and peak in tumors with distinct numbers of mutations, and that rarely mutated genes such as *XPO1* and *PLEKHA1* have high disease module importance in specific individuals. Furthermore, individualized module disruption enables us to devise customized singular and combinatorial target therapies which were highly varied across patients demonstrating the need for precision therapeutics pipelines. As the first analysis of *de novo* individualized disease modules, we illustrate the power of individualized disease modules for precision medicine by providing deep novel insights on the activity of diseased genes in individuals.

Keywords: Individualized disease modules, Individualized medicine, Systems biology, Network biology, Precision oncology, Precision medicine, Synthetic penetrance

Background

Pooled -omic data from patient samples has enabled construction of cellular interaction modules that provide a systems level understanding of disease etiology. This new conceptualization of disease has led to discoveries of previously unknown mechanisms and has significantly expanded opportunities for therapeutic targeting (Iborra-Egea et al. 2017; Sharma et al. 2018). Specifically, systems and network science have pinpointed novel pharmacological targets and opportunities for drug repurposing or drug-drug synergies (Zhao and Iyengar 2012). Advancements due to systems biology have been even more pronounced in oncology. The reconstruction of complex cancer disease modules describes tumor biology at the systems level, which is particularly important for such a polygenic and dynamic disease (Lin et al. 2019; Zielinski et al. 2017).

Despite these recent advancements, a truly individualized systems approach has yet to be applied to individualized medicine. Oncology patients in particular experience highly variable disease phenotypes and drug responses. The need for precision approaches in oncology has therefore been well established with numerous scientists and clinicians calling for innovation (Relling and Evans 2015; Werner et al. 2017; Aronson and Rehm 2015; Carrasco-Ramiro et al. 2017). Patient derived xenograft (PDX) models and clinical studies have highlighted the heterogeneity of tumor mechanistic properties and therapeutic responses (Chiron et al. 2014; Xu et al. 2019; Dagogo-Jack and Shaw 2018). Some of this variability can be captured with patient stratification through disease subtype classification or biomarker testing, but the majority of inter-patient variability remains unexplained (Dagogo-Jack and Shaw 2018). The lack of broadly applicable biomarkers indicates that unique systems level interactions are at play within single patients. Systems and network biology are poised to capture these phenomena well, but new

theoretical frameworks and computational approaches must be implemented to make such precision network biology a reality.

Existing methodologies extract disease modules (or "disease networks") which are highly perturbed subnetworks of the larger cellular interactome where disease gene interactions occur (Menche et al. 2015). Previous approaches have attempted to detect and prioritize individualized cancer drivers, but these algorithms infer their individualized analyses from a cohort level disease modules (Bashashati et al. 2012; Cho et al. 2016; Reyna et al. 2018). For example, the LIONESS algorithm uses aggregate disease modules generated by existing approaches to linearly interpolate individual sample modules (Kuijjer et al. 2019). We hypothesize that although some individual patient disease activity is recapitulated in the cohort disease modules, there are additional unexplored interactions detectable only at the individual patient level which dictate patient-specific mechanisms, phenotypes, and therapeutic responses. We additionally suspect that at the gene level there are patient specific variations in pathogenicity. This is because patients possess highly varied basal cellular environments and mutations (Curtis, C. 2012; Dagogo-Jack and Shaw 2018). Given that current approaches rely heavily or exclusively on cohort inferred disease modules, we suspect that inter-patient variability and precision has been underrepresented.

While practical, using features inferred across the cohort fail to capture patient individualized features by disregarding rare unique factors within a patient. A new approach is needed to truly infer individualized disease modules that accurately recapitulate individualized disease. In this study, we examined on the collective actions of mutated genes to try to understand individualized disease at a deeper level. We hypothesized that cohort disease modules are poorly representative of individualized disease and that new insights precision medicine would reveal

themselves once we zoomed in on individual patients. Thus, we set out to first create a robust pipeline for individualized disease module construction and second to use these disease modules to characterize individualize disease pathobiology and therapeutics.

Results

Shortest path analysis of individual patient mutated genes encapsulates disease activity into individualized disease modules

Rather than occupying an average disease module, we predicted that individual disease modules likely occupy their own distinct foci within the larger protein-protein interactome (Fig 1A) (Menche et al. 2015). We thus set out to develop a pipeline for the construction of single patient disease modules. Disease modules have historically been inferred at the cohort level using correlative, random walk, or shortest path approaches (Cahan et al. 2014; Da Rocha et al. 2016; Jia and Zhao 2014; Codling et al. 2008; Chen and Zhang 2013; Cerami et al. 2010; Managbanag et al. 2008). Because disease modules are generally dense and scale-free collections of interacting disease genes, (Wuchty 2001) shortest path analysis has proven to be a robust tool for network analysis in biological settings (Managbanag et al. 2008; Yu et al. 2007; Chen et al. 2016). Shortest path analysis also offers technical advantages over the other approaches. Correlative approaches require numerous samples for each patient which is very rare, particularly in the multi-omic setting. On the other hand random walk analysis is extremely computationally intensive, requires several iterations to reach a consensus, and is only a locally optimized search strategy (Xia et al. 2019). Furthermore, shortest paths have been found to be the strongest and most likely interaction pathways in other applications of network topology (Katzav et al. 2015). We therefore began to build individual disease modules with a shortest path approach on a protein-protein interaction network (PPI) (Fig 1B). For our individualized

analysis, a shortest path approach allowed us to connect sparse unique sets of diseased genes. We utilized the well documented PPI iRefIndex containing numerous categories of well-validated protein-protein interactions to find these pathways (Razick et al. 2008). For this reason, we favored iRefIndex over experimental databases which typically only capture one mode of interaction (Luck et al. 2020) and are a fraction of the size. The other database we considered was the STRING database, which is larger than iRefIndex, but contains numerous interactions that are not limited to physical interactions and is less stringently validated.

We began with 90 breast cancer patients from the TCGA BRCA Project as a proof-of-concept cohort for our precision systems biology pipeline (Table S1) (Curtis, C. 2012). These patients were of diverse mutational burden and PAM50 subtype, thus capturing the biological heterogeneity of the breast cancer (Fig S1A). We designated mutated genes identified by TCGA mutation annotations as “diseased genes”. These included missense, nonsense, insertion, and deletion mutations which were detected using Sniper, MUTECT, VarScan, and Muse (Cibulskis et al. 2013; Koboldt et al. 2012; Fan et al. 2016; Larson et al. 2012). We then found all shortest paths between all pairs of diseased genes for each individual patient (Figs 1B and S1B). We then enriched for paths likely to contain disease activity with three parameters. First, we measured gene expression fold change of each gene within a path because proteins with pathogenic mutations are known to frequently cause transcriptomic changes in interacting partners (Zhong et al. 2009). Second, in order to capture multi-gene interactions, shortest paths that crossed a third diseased gene were prioritized. Finally, paths that contained established cancer drivers were also prioritized because minor or rarer disease actors have been found to encourage pathogenic activity through known cancer drivers (Sondka et al. 2018; Castro-Giner et al. 2015).

The probability density function of patient paths compared to 1,000 random paths of equivalent length indicated a highly significant subset of paths as measured by empirical p-value (see Methods). These significant paths were aggregated together to form disease modules for each patient (Figs 1B and S1B). To understand how differing p-value cutoffs affect the pipeline, we selected five patients of varying mutational burdens and recorded network parameters as we varied the p-value cutoff. As expected, the number of significant paths and network size increased with increasing p-values (Fig S2A-C). The increases seemed to be exponential indicating that stringent p-values are needed to keep modules small and enriched in the most important diseased genes. Given these results, we moved forward with a p-value cutoff of 0.01 which provided a good balance of network size and stringency. In our entire cohort, we saw a subset of highly enriched paths for each patient out of the numerous paths that were detected (Figs 1C, S2D-F). Approximately 5-15% of detected paths were ruled significant using the cutoff of 0.01 (Fig S2F). Using this process, we were able to generate individualized disease modules for each of the 90 patients without cohort inferred features.

The properties of patient disease modules (Table S2) were analyzed to understand the implications of the number of disease genes (mutational burden) on the resultant network. Almost all mutations were carried forward to the individualized disease modules (Fig 1D and Table S3). This indicates that even mutations which would normally be considered “passengers” can have impactful effects when cooperating with co-occurring “drivers”. This result is explained by the “mini-driver model” which postulates that, in cancer, disease genes lie on a spectrum from driver to passenger (Castro-Giner et al. 2015). Under this model, individualized patient context (mutated genes, underlying genetics, transcriptome) dictates where on this spectrum a disease gene lies. Our work supports this model because we observe that minor

mutations can be implicated in disease module activity through their interactions with more substantial cancer drivers. Patients with an increased mutational burden had larger individualized disease modules as measured by edge and node number (Figs 1E, S3A, and Table S3). With the addition of each mutation, more non-diseased genes are brought into the disease module.

We next compared individualized disease modules to a cohort-based module to confirm our hypothesis that the cohort module poorly recapitulates individualized disease (Fig 1A). A cohort disease module was generated using the GRNBoost algorithm and all patient transcriptomic profiles (Moerman et al. 2019). We found that for most patients only half of their individualized module nodes were represented in the cohort module (Fig S3B). Furthermore, we saw that only 60-20% of individual patient mutations were included the cohort module and that there were thousands of extra nodes in the cohort module (Fig S3C-D). To understand if there was a bias of cohort modules toward frequently mutated genes, we inspected the percent of rare versus frequent mutations carried over to the cohort module in each patient (Fig S3E). A discernable bias was not detected when looking at the percent of excluded rare and frequent mutations, but numerically many more rare mutations were left out of the cohort disease module per patient which is not unexpected because rare mutations are more frequent (Fig S3F). Cumulatively these results confirm that cohort disease modules are not representative of many features of individualized disease.

In silico disruption of individualized disease module identifies personalized mutated driver genes

With individualized disease modules established, we sought to understand the varying importance of disease genes for each patient. To quantify disease gene importance, we measured the changes in the network parameters hubs, bottlenecks, network components, and edges when a

disease gene and its cooperative paths are removed from the individual disease module (Figs 2A and Fig S4A-E). Here we hypothesize that changes in network parameters upon gene removal can give insights into how crucial a given gene is to the overall disease module and ultimately the overall disease phenotype. Thus, the network components were aggregated into a quantitative score of individualized disease module disruption that we termed individualized disease gene importance (IDGI) (Fig 2A).

Because this score was generated to assess disease genes in individual patients, we first validated IDGI by comparing real and randomized individualized patient disease modules. Randomizing trials were completed ten times in ten patients for a total of 100 separate trials. High scoring disease genes in real disease modules scored significantly above shuffled disease modules, thus confirming the IDGI methodology (Fig S5A-J). Furthermore, IDGI was able to recover genes listed by the Cancer Gene Census as implicated in cancer pathobiology with an AUC of 0.821 and 0.879 for all cancers and breast cancer, respectively (Fig S6A) (Sondka et al. 2018). Using the same five test patients and Cancer Gene Census (CGC) recovery from Figure S2A-C, we completed an ablation experiment to understand each network components' effect on the IDGI score. Edge number was the most important component as shown by the strongest decrease in AUC (Fig S6B-D). The other IDGI components did not markedly decrease AUC when removed, but together they were able to maintain a reasonable AUC when edges were removed indicating that these collectively contributed to an AUC increase (Fig S6B-D). With the IDGI score well validated, we returned to the path selection scheme to understand how each component contributes to individualized disease module construction. Again, using the five test patients and Cancer Gene Census recovery validation scheme, we found that disease information contributed the most to AUC followed by patient mutations (Fig S6E-F). This is not unexpected

because this validation scheme only tests the pipeline's ability to recover known disease drivers and thus it is an imperfect test of our pipeline because we seek to identify rare and frequent drivers in an individualized manner. We retained the individualized omic features of fold change and mutation status because of this in our final pipeline. Nevertheless, these validation *in silico* experiments together confirmed that our method provided individualized and disease relevant quantitation.

Looking at the IDGI scoring for each patient, several high scoring outlier disease genes were noted in most patients. Each patient's outliers were a mix of rarely (<5% in the TCGA BRCA cohort) and commonly (>5% in the TCGA BRCA cohort) mutated genes (Fig 2B) (Curtis, C. 2012). This finding affirms the need for an individualized approach as many of the rarely mutated genes would likely not be detected in statistical or cohort-based analyses. We next integrated gene scores with patient disease modules to visualize disease gene importance. In some patients, well known drivers such as PIK3CA carried high importance (Figs 3A and S7A-B), and other patient disease modules were instead driven by rare mutations. In one example, *PIK3CA* and *TP53* play a secondary role to the rarely mutated genes *XPO1* and *PLEKHA1* (Figs 3B and S8A-B). *XPO1* was of particular interest because an *XPO1* inhibitor induced complete remission in a limited number of patients during solid tumor clinical trials (Azizian et al. 2020). Furthermore, *PLEKHA1*, a largely uncharacterized membrane signaling protein, has not been described in breast cancer thus far (Dowler et al. 2000). Both genes are mutated in less than 1% of breast cancer cases (Curtis, C. 2012). To display all 90 patients' data, we created an interactive data portal (Table S4 and https://syspharm.shinyapps.io/PERMUTOR/). Each of these patients exemplify the unique potential of precision systems analysis which allows for the highlighting of rare important mutations that we predict drive patient-specific disease.

Synthetic penetrance describes the varying importance of disease genes across patients

We further appreciated the varying influence of known disease genes when we examined them across our entire cohort. Here we found that, irrespective of cancer subtype, scores varied drastically across patients. *TP53*, *PIK3CA*, *CDH1*, *GATA3*, and *MAPK* subunit genes scored the highest on average; *TP53* and *PIK3CA* demonstrated a broad range of importance across the cohort (Fig 4A-B). *TTN* mutations, which are common but largely inconsequential, were the third most frequent in our cohort (Oh et al. 2020). IDGI scores for *TTN* revealed low scoring across all patients, which indicates that our IDGI scoring scheme can detect "passenger" genes (Fig 4B).

We next examined the scores of *PIK3CA* and *TP53* to characterize how IDGI varied with tumor properties. Seven patients had mutations in both of these genes, and in five of the seven, both genes had a low IDGI indicating a diffusion of disease activity to lesser known mutated genes. *PIK3CA*, the second most frequently mutated gene in breast cancer, had the highest level of importance in patients with a lower mutational burden (<30 mutated genes) (Fig S9A). Conversely, the most commonly mutated gene, *TP53*, was more important in patients with a higher mutation burden (30 to 100 mutated genes) (Fig S9B). *TP53* IDGP demonstrated more variable activity in missense mutations when compared to splice site mutations (Fig S9C). Classifying missense mutations as in or outside of known mutational hotspots reveals that mutations within hotspots contribute more to the disease module on average (Fig S9D) (Baugh et al. 2018).

The presence of unique cancer drivers is often hypothesized, and our gene importance scoring reveals an additional aspect of this postulate. As we have shown, commonly mutated genes vary in importance in different patients. We term this variable impact of disease genes

across patients “synthetic penetrance” (Fig 5). Synthetic lethality was conceptualized following the observation that, in some model organisms and cancer cell lines, specific combinations of genetic perturbation result in cell death (O’Neil et al. 2017). As in synthetic lethality, here “synthetic” describes the requirement of specific settings and partners to enable disease gene activity, and “penetrance” encapsulates the variability in gene importance among individuals. Our examination of commonly mutated genes within breast cancer illustrates that synthetic penetrance is influenced by known biological contexts and is an important factor in precision medicine.

Simulation and prioritization of personalized therapeutic targets via disease module disruption

We next used disease module disruption to prioritize therapeutic targets. Target therapies are an emerging class of pharmaceuticals that are aimed at specific genes that cancer cells use for invasion or survival. We examined the disease module importance of FDA approved target therapy genes (Table S5) in individual patients using an adapted gene importance score termed Individualized Target Gene Importance (ITGI). Breast cancer target therapy genes showed significantly elevated importance over those used in other cancers (Fig 6A and Table S6). When examining each gene specifically, we found that breast cancer target therapy genes ranged in disease module disruption efficacy among individuals (Fig 6B). Surprisingly, some target therapy genes from other cancers had high disruptive potential for certain patients (Figs 6C and Fig S10). For example, *BTK* (zanubrutinib approved for mantle cell lymphoma) and *MAPK1* (trametinib approved for melanoma) were low scoring on average but scored with high disease module importance in one patient each despite not being clinically approved breast cancer targets (Fig 6C) (Wishart et al. 2008). Neither gene was mutated in these patients which confirms the

need for an individualized systems approach because such targets could not be identified with genomics alone.

As with IDGI scores (Fig 2B), each patient had high-scoring ITGI outliers, but the identity of top target therapy genes was highly variable from patient to patient (Fig S11A-D). Breast cancer target genes can be further classified by FDA approval for specific breast cancer subtypes. When the patient cohort was divided by breast cancer subtype, the target therapy genes scored higher on average for their approved subset of patients (Fig S11E-I). We added the target analysis to the data portal for further exploration of individual patients (Table S6 andhttps://syspharm.shinyapps.io/PERMUTOR/). Across these analyses, group behavior fell in line with conventional drug approvals, but individualized examination of ITGI scores revealed that there are important patient-specific variations that should be accounted for in precision medicine approaches.

Oncology drugs are often used in combination during cancer patient treatment (Zagidullin et al. 2019). Predicting efficacious combinations of approved drugs is a persistent clinical need, and devising patient-specific combinatorial regimens is an additional challenge that must be addressed for next generation therapeutics (DiNardo and Perl 2019). We reason that ITGI scoring could be applied to reveal potential combinatorial treatment regimens for single patients. All but one patient showed an increase in maximum ITGI score with an optimal target pair when compared to the highest-scoring single target (Fig 6D and Table S7). Combinatorial gene scores fell into three categories: less than additive (overlapping) single target scoring, equal to additive scoring, and higher than additive scoring (synergistic). We identified improved synergistic pairs for patients who had low single target scores (Figs 6E-F and S12A-B). In other cases, target therapy combinations showed far less than additive disease module disruption,

indicating redundancy in targeting (Fig S12C-F). Such redundancy in therapeutic coverage may be favorable depending on biological context. Our results demonstrate the success and feasibility of using disease module disruption for individualized therapeutic regimens and drug repurposing.

Discussion

Disease modules have become a vital tool for disease understanding cohort studies. Here we demonstrate that disease modules reveal even deeper meaning when applied at single patient resolution. Numerous calls for new approaches in precision medicine have been put forward by the oncology community, but marginal progress has been made (Zhu et al. 2019; Relling and Evans 2015; Werner et al. 2017; Yadav et al. 2020). This work adds to the toolset of clinicians and scientists in precision medicine by enabling the construction and functionalization of patient-specific disease modules. We anticipate that this approach will be generalizable to other cancers and even to other categories of polygenic disease such as metabolic and neurodegenerative disorders. Furthermore, our approach can easily accommodate updated whole cell networks in accordance with rapid advancements in the protein-protein interactome and domain-domain interactions. Yet, current interactomes are still incomplete but as suggested by Menche *et al.* despite this incompleteness highly valuable disease modules were able to be constructed (Menche et al. 2015). Interactomes are also subject to biases from experimental methods and the degree to which a gene has been studied, but aggregative databases with high standards for inclusion help mitigate these biases (Razick et al. 2008). Personalized protein-protein interactomes that documents changes due to SNPs or somatic mutations could additionally increase the precision of individualized disease modules (Bhattacharyya et al. 2021). Other groups have mapped individual genetic profiles to protein-protein interactomes when understanding complex disease (Loscalzo 2019), but a genetics only approach neglects to

examine the downstream consequences of mutations and compensatory changes in normal genes. By including patient transcriptomics, we were able to examine how non-mutated genes are pulled into the disease module, and which of these genes can be a potential therapeutic target. Individualized interactomes have been proposed, but a high throughput method for individualized interactome assessment has yet to emerge (Dehne and Green 2012). Such personalized advancements will further increase the accuracy and depth with which we are able to characterize individualized disease modules.

Through building this precision systems biology pipeline, we uncovered key insights about patient-to-patient variability in cancer disease genes. First, we found strong evidence for the mini-driver model, which postulates that there are subtle context-specific driver genes that enable individualized disease (Castro-Giner et al. 2015). Rare mutations have long been suspected to be implicated in tumor growth, but demonstrating this at the single patient level has previously been difficult. Clinically, extended genetic screens for tumors with atypical genetic alterations have addressed the obstacle of rare mutations (Zhu et al. 2019), but these assays neglect the context of a patient's baseline cellular environment and co-occurring mutations. To appreciate this broader context, we incorporated normal tissue transcriptomes and interacting mutation paths while building the individualized disease module. We quantify the sum cooperativity and importance of a disease gene by measuring several disease module parameters. Network parameters have been used successfully in cohort-based algorithms to prioritize drug targets and gene pathogenicity (Zhong et al. 2009; Dunn et al. 2005; Boezio et al. 2017). As ground truth for mini-drivers has yet to be established, the highest accuracy disease module disruption parameters remain unknown. Despite this, we were able to perform two-step *in silico* validation confirming that IDGI scores recover relevant putative mini-drivers for individual

patients. This is a crucial first step towards understanding pathogenic mechanisms in a patient and tumor-specific manner.

Similarly, we revealed the interacting unmutated genes that contributed to individualized disease. Disease module studies have found that increasing numbers of diseased genes bring in more seemingly normal actors into the disease module, but until now this has not been confirmed to apply to individual patients (Menche et al. 2015). A recent meta-analysis of patients treated with targeted or chemotherapies found that many cancers with high mutational burden had a decreased overall survival, and this effect could only be mitigated by immunotherapies (Valero et al. 2021). In light of our findings, the negative association between mutational burden and survival may reflect an increase in module size which allows for disease resilience against traditional pharmacotherapies in combination with other factors.

Through this work, we shed light on unique disease determinates and characterize the dynamic nature of more familiar pathogenic factors. Omic studies have found distinct driver genes for breast cancer subtypes and metastatic tumors, but deeper precision medicine contexts have been unexplored (Annunziato et al. 2019; Rajendran and Deng 2017; Bailey et al. 2018). With the unparalleled resolution shown here, we characterize the range of influence a gene can exert on an individual disease module. We describe this context specific variability as synthetic penetrance. Gene penetrance has traditionally been described as the percentage of individuals with a given gene variant that display a trait (Zlotogora 2003). Here, in synthetic penetrance, the "trait" is how influential a gene is in the disease module. We found that synthetic penetrance varied widely across patients, which is not surprising given the diversity of mutations and disease presentations that make up the analyzed cohort. For our two most prevalent disease genes, PIK3CA and TP53, we observed distinct peaks in IDGI scores at specific levels of

mutational burden. Precision systems analysis with our approach can further clarify how disease module dynamics act in individuals and how synthetic penetrance can evolve over time within a tumor.

Given the persistent treatment resistance and variable response rates of many cancers, we leveraged our individualized disease modules and gene importance pipelines to extend our analysis into precision pharmacology. Targeted therapies have provided durable response rates and survival benefits in many tumors, but for others, indications for use are unclear (Xie et al. 2020; Murdoch and Sager 2008; De Palma et al. 2017). Fully utilizing targeted therapies and other drugs for repurposing efforts requires a mechanistically informed predictive approach. We begin this effort in precision medicine by measuring the effect of targeted therapy on individualized disease module. Additionally, we performed a combinatorial screening to identify synergistic target therapy genes in individual patients. These results show that disease module disruption evaluation is a viable strategy for combinatorial therapy prioritization. Tuning of disease module parameter scoring from large scale *in vitro* drug screening on patient derived xenographs will be required to translate this pipeline to clinical settings. To date, such data has been limited, but the *in silico* work presented here shows that systems-informed precision oncology has immense promise.

In summary, we demonstrate a novel theoretical basis and corresponding computational pipeline for understanding individualized disease modules. Through this, we were able to demonstrate that there is an underlying phenomenon termed synthetic penetrance which is only appreciable when we examine a disease gene across several individually characterized breast cancer patients. Furthermore, these individualized disease modules can be functionalized to prioritize precision medicine targets. We anticipate that the principles and pipeline presented

here will be able to prioritize disease genes and therapeutic targets in other complex diseases such as metabolic syndrome or neurodegenerative conditions in future work. Even within monogenic diseases we believe that this framework can reveal new insights by understanding the pathogenic variant's interactions with background patient SNPs. Because of this, we anticipate that individualized disease modules have a strong future in the field of precision medicine and pharmacology.

Methods

Data acquisition and processing

Data processing and method creation were done in R (R Core Team 2021). Matched genomic and transcriptomic data correspond to 90 breast cancer (BC) patients were downloaded from the TCGA-BRCA project (https://portal.gdc.cancer.gov/projects/TCGA-BRCA) (Curtis, C. 2012). The selected patient profiles had RNA-seq data for both tumor and normal tissues and mutations annotations for tumor tissues. Patient IDs are in Supplemental Table S2. Tumor mutation annotations were available in four MAF files with each containing mutational identification analysis from Sniper, MUTECT, VarScan, and Muse, and this data was processed using the maftools package in R (Larson et al. 2012; Cibulskis et al. 2013; Fan et al. 2016; Koboldt et al. 2012; Mayakonda et al. 2018). Mutations annotations from these four MAF files were aggregated for each patient and used in downstream analyses. RNA-seq data for tumor and

normal BC samples was filtered for low variance genes and normalized. Differential gene expression analysis was performed using the package DESeq2 (Love et al. 2014).

Target Therapy Data Acquisition

Target therapy drugs were pulled from the NIH NCI Target Therapy Fact Sheet. The gene targets of these drugs were obtained from DrugBank (Wishart et al. 2008). Drugs whose targets were not available in DrugBank, immunotherapies, and broad targeting agents were excluded. Details on these targets are in Supplemental Table S1.

The PERMUTOR algorithm

Overall Design: We combined the described individualized disease module construction, gene importance scoring, and target therapy gene scoring into a complete pipeline called PERsonalized MUtation evaluaTOR (PERMUTOR). The PERMUTOR algorithm aims to prioritize the most important mutations and therapeutic targets within an individual patient's tumor according to their disruptive effects on the patient's individual disease module. Initially, an individualized disease module is created by extracting the most important shortest paths between pairs of mutations from the PPI. This is done to distill the entire interactome down to the cooperative interactions that contain disease activity. The impact of each disease gene on the individual disease module is then evaluated by measuring the change the removal of that gene and its cooperative paths cause in key network parameters. Target therapies were also examined

in this manner to find existing drugs that could be efficacious in a specific patient. The individual disease module themselves also offer biological insights.

PPI Preprocessing: Including only expressed proteins in the generalized network

The PPI network used in this work was generated from the irefindex database (http://irefindex.org/wiki/index.php?title=iRefIndex) (Razick et al. 2008). Connections (edges/nodes) were filtered as described in Lummertz da Rocha et al (Da Rocha et al. 2016). The resultant network contained 249,852 connections (edges) and 16,375 proteins (nodes). Before annotating the network with individualized -omics, genes that did not have significant expression in the patient's normal tissue sample or tumor sample were removed from the generalized network. The individual patient data from mutational profiles and RNA-seq differential expression was annotated on each patient's PPI for use in future scoring. Information from the Cancer Gene Census was also annotated on each patient's protein-protein interaction network (Sondka et al. 2018).

Stage 1: Computing the shortest protein-protein interaction paths that link two mutated genes in the PPI network

Since mutated genes exert their functional effects on cancer cells via interactions with their partners in the PPI network, we hypothesized that shortest paths between two mutated genes contain disease activity. The all_shortest_paths function in the igraph package was used to compute shortest paths for each pair of mutated genes found in a single patient.

Stage 2: Assessment of shortest protein-protein interaction paths that link two mutated genes in the PPI network

Not all mutations or cooperative paths were important in the disease etiology of a single patient, and thus a path scoring scheme was created to distill the most important interactions. The shortest paths connecting pairs of mutated genes were deemed potentially important in individualized disease etiology if one or more of the following criteria are met: (i) the shortest paths are enriched with differentially expressed genes; (ii) the paths are enriched with disease context genes; (iii) the paths contain one or more mutated genes from that patient. As such, the Path Score was devised to assess and identify shortest paths that are potentially important.

Each gene within the path was scored, and the path score was calculated to be the sum of the components averaged across the length of the path. This formula takes into account: i) fold change RNA-seq between tumor and tissue normal of that patient (r), ii) if the path contained non-terminal nodes that were mutated in that patient (m) and iii) a constant associated with the gene tier in the Cancer Gene Census (https://cancer.sanger.ac.uk/census). The Cancer Gene Census (CGC) is a database of genes found to be highly implicated in tumorigenesis. Tier 1 genes have very strong literature support for being impactful to cancer biology, and tier 2 genes have less evidence but are still suspected to be important. A constant was added for tier 1 and a smaller constant was added for tier 2. No constant was added if the gene was not in the CGC gene census.

Path Score: $p = \frac{\sum_{i=1}^{n} |r_i| + t_i + m_i}{n}$

r=Fold change RNA-seq

t= Cancer Gene Census Tier Score (tier 1= 10, tier 2= 2)

m= mutation score = 5

n= path length

To benchmark our method, paths scores were compared to the scores of 1,000 random paths of the same length to calculate an empirical p-value for a given real path. Randomized paths were scored using the same patients fold change and mutational data so each patient had unique sets of randomized paths scored with their omic data. The empirical p-value is calculated as the percentage of random paths that score higher than the real path. In this way, a significant path will have a lower empirical p-value and be higher than most randomized paths. All paths that have a p-value over a certain threshold (p-value < 0.01 in this work) were moved into a unique individualized disease module.

Stage 3: Construction personalized disease modules and evaluation of the importance of disease genes via IDGI scoring

To understand the impact of each mutated gene in the individual disease module, module parameters were measured pre and post removal of a gene and its cooperative relationships. All cooperative paths that began or ended with the gene under investigation were removed from the disease module. The separated disease module components, edge number, bottlenecks, hubs, and highly trafficked hubs were measured for the altered individual disease module. The difference in each parameter between the altered and the original individual disease module was recorded for each gene. These change in parameter values were used to calculate an Individualized Disease Gene Importance (IDGI) score that quantified the gene's importance to the individualized disease module.

<u>Parameter Definitions:</u>

Highly Trafficked Hubs: Highly trafficked hubs were defined as nodes that had edge connectivity and betweenness as measured by the betweenness function in igraph in the top 5th percentile of that disease module.

Hubs: Hubs were defined as nodes that met the edge connectivity cutoff of the 5th percentile, but not the betweenness connectivity.

Bottlenecks: Bottlenecks were defined as nodes that had edge connectivity in the bottom 5th percentile of that disease module.

Separated network components: Separated network components were defined as completely disconnected subgroups within the entire disease module architecture. The count_components function from igraph was used to find the number of disconnected submodules

$$\text{IDGI Score} = 10(\Delta\text{subnetworks}) + 3(\Delta\text{bottlenecks}) + 5(\Delta\text{hubs}) + 6(\Delta\text{highly trafficked hubs})$$

These topological parameters were weighted (10,3,5,6) to reflect their importance in network science connectivity and functionality based on previous work in the field (Al-aamri et al. 2019; Zhao and Liu 2019). These weights are also indicative of their frequency in biological networks. For example, disconnected sub-networks are rare occurrences that completely sever linkages between genes that once interacted, and because of this, disconnected subnetworks are weighted the largest because it is rare and has the most drastic effect on the network topology.

Stage 4: Individualized Drug Target Analysis

In the next phase, genes targeted by cancer target therapies were investigated. Approved targeted therapies were found via the NIH NCI Targeted Cancer Therapies website and DrugBank (Supplemental Table S1). All genes within the individualized disease module were compared to

this list of targets. All targets that were represented in the disease module were removed one at a time, and the disease module was rescored after each exclusion. Next, pairs of drug targets were removed, and the disease module was rescored to find potent combinations that displayed network disruption above single target disruption.

ITGI Score=10(Δsubnetworks)+3(Δbottlenecks)+5(Δhubs)+6(Δhighly trafficked hubs)

Assessment of the Performance of the PERMUTOR Algorithm

To validate and explore our pipeline, we constructed five testing schemes:

Test scheme 1: P-value Pruning Characterization

For five test patients of varying mutational burden, we constructed individualized disease modules using empirical p-value cutoffs of 0.005, 0.01, and 0.05. The resulting disease modules were analyzed for their number of nodes, edges, and percent of significant paths.

Test scheme 2: Generation of Cohort Disease Module

Here, a cohort disease module was constructed using the well-known algorithm GRNBoost2 (Moerman et al. 2019) from the Arboreto Python package. All patient transcriptomic profiles were used as input for GRNBoost2. This disease module was pruned to include edges weighted with values greater than 25. This cohort module was then compared all individualized disease modules.

Test scheme 3: Individualized Disease Module Randomization

For ten representative patients, their individualized disease module was tested by randomly shuffling node labels 10 times and rescoring the mutated genes. These randomly shuffled scores were compared to the original score for each gene in that patient.

Test scheme 4: Cancer Gene Census Receiver Operator Curve (ROC) Analysis and Ablation Studies

ROC analysis was completed by testing the recovery of CGC genes using all patients' IDGI scores. This was repeated for CGC genes that were indicated as impactful to breast cancer specifically.

For path and IDGI score ablation studies, the five test patients were used. For each run, components of the scores were removed and the ROC analysis was repeated.

Test scheme 5: Target Therapy Approved Indications Comparisons

Clinical information for each patient was extracted from the TCGA-BRCA project. Approved indications for each target therapy were found from DrugBank and the NIH NCI Target Therapy Fact Sheet. First, breast cancer approved targets were compared against non-breast cancer targets to demonstrate our ITGI score's ability to recapitulate cancer specific approvals. Second, the breast cancer approved targets were investigated for specific subtype approvals, and patients who fit the approval criteria were compared against those who did not meet the approval indications.

Figures, Analysis, and Website

Schematic figures were created in BioRender (https://biorender.com/). All other figures were created in R and combined in Adobe Illustrator. The PERMUTOR Data Portal was created using R Shiny.

Software availability

The software generated in this study have been submitted in the supplementary material and in the Li lab GitHub (Supplemental Code S1 and https://github.com/HuLiLab).

The data generated in this study have been submitted in the supplementary material and in the PERMUTOR Data Portal (https://syspharm.shinyapps.io/PERMUTOR/).

Competing Interest Statement

The authors declare no competing interest.

Acknowledgments

Funding: This work was supported by grants from National Institutes of Health (NIH) [R01CA208517, R01AG056318, R01AG61796, P50CA136393, R01CA240323, 5T32GM065841-1]; the Glenn Foundation for Medical Research, Mayo Clinic Center for Biomedical Discovery, Center for Individualized Medicine, Mayo Clinic Cancer Center (P30CA015083), and the David F. and Margaret T. Grohne Cancer Immunology and Immunotherapy Program. **Author's contributions:** T.M.W., C.Y.U., C.C., C.Z., and H.L. contributed to the conception and design of the study. T.M.W., C.Y.U., C.C., C.Z. and H.L. contributed to the acquisition of data, T.M.W., C.Y.U., C.C., C.Z., and H.L. contributed to the analysis and interpretation of data. T.M.W., C.Y.U., C.C., and H.L. drafted the manuscript. H.L. supervised the study. Competing interests: The authors declare that they have no competing interests. **Acknowledgements:** The authors would like to thank James J. Collins for his suggestions on this work.

References

Al-aamri A, Taha K, Al-hammadi Y, Maalouf M, Homouz D. 2019. Analyzing a co-occurrence gene-interaction network to identify disease-gene association. 1–15.

Annunziato S, de Ruiter JR, Henneman L, Brambillasca CS, Lutz C, Vaillant F, Ferrante F, Drenth AP, van der Burg E, Siteur B, et al. 2019. Comparative oncogenomics identifies combinations of driver genes and drug targets in BRCA1-mutated breast cancer. *Nat Commun* **10**. http://dx.doi.org/10.1038/s41467-019-08301-2.

Aronson SJ, Rehm HL. 2015. Building the foundation for genomics in precision medicine. *Nature* **526**: 336–342.

Azizian NG, Azizian NG, Li Y, Li Y. 2020. XPO1-dependent nuclear export as a target for cancer therapy. *J Hematol Oncol* **13**: 1–9.

Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, Colaprico A, Wendl MC, Kim J, Reardon B, et al. 2018. Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* **174**: 1034–1035.

Bashashati A, Haffari G, Ding J, Ha G, Lui K, Rosner J, Huntsman DG, Caldas C, Aparicio SA, Shah SP. 2012. DriverNet: uncovering the impact of somatic driver mutations on transcriptional networks in cancer. *Genome Biol* **13**: R124. http://genomebiology.com/2012/13/12/R124.

Baugh EH, Ke H, Levine AJ, Bonneau RA, Chan CS. 2018. Why are there hotspot mutations in the TP53 gene in human cancers? *Cell Death Differ* **25**: 154–160. http://dx.doi.org/10.1038/cdd.2017.180.

Bhattacharyya R, Ha MJ, Liu Q, Akbani R, Liang H. 2021. Personalized Network Modeling of the Pan-Cancer Patient and Cell Line Interactome. 399–411.

Boezio B, Audouze K, Ducrot P, Taboureau O. 2017. Network-based Approaches in Pharmacology. *Mol Inform* **36**: 1–10.

Cahan P, Li H, Morris SA, Lummertz Da Rocha E, Daley GQ, Collins JJ. 2014. CellNet: Network biology applied to stem cell engineering. *Cell* **158**: 903–915.

Carrasco-Ramiro F, Peiró-Pastor R, Aguado B. 2017. Human genomics projects and precision

medicine. *Gene Ther* **24**: 551–561.

Castro-Giner F, Ratcliffe P, Tomlinson I. 2015. The mini-driver model of polygenic cancer evolution. *Nat Rev Cancer* **15**: 680–685.

Cerami E, Demir E, Schultz N, Taylor BS, Sander C. 2010. Automated network analysis identifies core pathways in glioblastoma. *PLoS One* **5**.

Chen H, Zhang Z. 2013. Prediction of associations between OMIM diseases and MicroRNAs by random walk on OMIM disease similarity network. *Sci World J* 1–6.

Chen L, Xing Z, Huang T, Shu Y, Huang G, Li H. 2016. Application of the Shortest Path Algorithm for the Discovery of Breast Cancer-Related Genes. 51–58.

Chiron M, Bagley RG, Pollard J, Mankoo PK, Henry C, Vincent L, Geslin C, Baltes N, Bergstrom DA. 2014. Differential antitumor activity of aflibercept and bevacizumab in patient-derived xenograft models of colorectal cancer. *Mol Cancer Ther* **13**: 1636–1644.

Cho A, Shim JE, Kim E, Supek F, Lehner B, Lee I. 2016. MUFFINN: Cancer gene discovery via network analysis of somatic mutation data. *Genome Biol* **17**: 1–16. http://dx.doi.org/10.1186/s13059-016-0989-x.

Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M, Lander ES, Getz G. 2013. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* **31**: 213–219.

Codling EA, Plank MJ, Benhamou S. 2008. Random walk models in biology. *J R Soc Interface* **5**: 813–834.

Curtis, C. et al. 2012. Comprehensive molecular portraits of human breast tumors The Cancer Genome Atlas Network. *Nature* **490**: 61–70.

Da Rocha EL, Ung CY, Mcgehee CD, Correia C, Li H. 2016. NetDecoder: A network biology platform that decodes context-specific biological networks and gene activities. *Nucleic Acids Res* **44**: 1–13.

Dagogo-Jack I, Shaw AT. 2018. Tumour heterogeneity and resistance to cancer therapies. *Nat Rev Clin Oncol* **15**: 81–94. http://dx.doi.org/10.1038/nrclinonc.2017.166.

De Palma M, Biziato D, Petrova T V. 2017. Microenvironmental regulation of tumour angiogenesis. *Nat Rev Cancer* **17**: 457–474. http://dx.doi.org/10.1038/nrc.2017.51.

Dehne F, Green J. 2012. Personalized human protein interactomes. *Dimensions*. https://app.dimensions.ai/details/grant/grant.2889451.

DiNardo CD, Perl AE. 2019. Advances in patient care through increasingly individualized therapy. *Nat Rev Clin Oncol* **16**: 73–74.

Dowler S, Currie RA, Campbell DG, Deak M, Kular G, Downes CP, Alessi DR. 2000. Identification of pleckstrin-homology-domain-containing proteins with novel phosphoinositide-binding specificities. *Biochem J* **351**: 19–31.

Dunn R, Dudbridge F, Sanderson CM. 2005. The use of edge-betweenness clustering to investigate biological function in protein interaction networks. *BMC Bioinformatics* **6**: 1–14.

Fan Y, Xi L, Hughes DST, Zhang J, Zhang J, Futreal PA, Wheeler DA, Wang W. 2016. MuSE: accounting for tumor heterogeneity using a sample-specific error model improves

sensitivity and specificity in mutation calling from sequencing data. *Genome Biol* **17**: 178. http://dx.doi.org/10.1186/s13059-016-1029-6.

Iborra-Egea O, Gálvez-Montón C, Roura S, Perea-Gil I, Prat-Vidal C, Soler-Botija C, Bayes-Genis A. 2017. Mechanisms of action of sacubitril/valsartan on cardiac remodeling: a systems biology approach. *npj Syst Biol Appl* **3**: 1–8. http://dx.doi.org/10.1038/s41540-017-0013-4.

Jia P, Zhao Z. 2014. VarWalker: Personalized Mutation Network Analysis of Putative Cancer Genes from Next-Generation Sequencing Data. *PLoS Comput Biol* **10**.

Katzav E, Nitzan M, Ben-Avraham D, Krapivsky PL, Ross N, Biham O. 2015. Analytical results for the distribution of shortest path lengths in random networks. *Europhys Lett* **111**: 1–8.

Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. 2012. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* **22**: 568–576.

Kuijjer ML, Tung MG, Yuan GC, Quackenbush J, Glass K. 2019. Estimating Sample-Specific Regulatory Networks. *iScience* **14**: 226–240. https://doi.org/10.1016/j.isci.2019.03.021.

Larson DE, Harris CC, Chen K, Koboldt DC, Abbott TE, Dooling DJ, Ley TJ, Mardis ER, Wilson RK, Ding L. 2012. Somaticsniper: Identification of somatic point mutations in whole genome sequencing data. *Bioinformatics* **28**: 311–317.

Lin M, Ye M, Zhou J, Wang ZP, Zhu X. 2019. Recent Advances on the Molecular Mechanism of Cervical Carcinogenesis Based on Systems Biology Technologies. *Comput Struct Biotechnol J* **17**: 241–250. https://doi.org/10.1016/j.csbj.2019.02.001.

Loscalzo J. 2019. Precision Medicine: A New Paradigm for Diagnosis and Management of Hypertension? *Circ Res* **124**: 987–989.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 1–21.

Luck K, Kim D, Lambourne L, Spirohn K, Begg BE, Bian W, Brignall R, Cafarelli T, Campos-laborie FJ, Charloteaux B, et al. 2020. A reference map of the human binary protein interactome. *Nature* **580**. http://dx.doi.org/10.1038/s41586-020-2188-x.

Managbanag JR, Witten TM, Bonchev D, Fox LA, Tsuchiya M, Kennedy BK, Kaeberlein M. 2008. Shortest-path network analysis is a useful approach toward indentifying genetic determinants of longevity. *PLoS One* **3**.

Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. 2018. Maftools: Efficient and comprehensive analysis of somatic variants in cancer. *Genome Res* **28**: 1747–1756.

Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J, Barabási AL. 2015. Uncovering disease-disease relationships through the incomplete interactome. *Science (80-)* **347**: 841–843.

Moerman T, Santos SA, Gonza CB, Simm J, Moreau Y, Aerts J, Aerts S. 2019. Gene expression GRNBoost2 and Arboreto□: efficient and scalable inference of gene regulatory networks. *Bioinformatics* **35**: 2159–2161.

Murdoch D, Sager J. 2008. Will targeted therapy hold its promise? An evidence-based review. *Curr Opin Oncol* **20**: 104–111.

O'Neil NJ, Bailey ML, Hieter P. 2017. Synthetic lethality and cancer. *Nat Rev Genet* **18**: 613–

623. http://dx.doi.org/10.1038/nrg.2017.47.

Oh JH, Jang SJ, Kim J, Sohn I, Lee JY, Cho EJ, Chun SM, Sung CO. 2020. Spontaneous mutations in the single TTN gene represent high tumor mutation burden. *npj Genomic Med* **33**: 1–11. http://dx.doi.org/10.1038/s41525-019-0107-6.

Rajendran BK, Deng CX. 2017. Characterization of potential driver mutations involved in human breast cancer by computational approaches. *Oncotarget* **8**: 50252–50272.

Razick S, Magklaras G, Donaldson IM. 2008. iRefIndex: A consolidated protein interaction database with provenance. *BMC Bioinformatics* **9**: 1–19.

R Core Team. 2021. R: A Language and Environment for Statistical Computing. https://www.r-project.org.

Relling M V., Evans WE. 2015. Pharmacogenomics in the clinic. *Nature* **526**: 343–350.

Reyna MA, Leiserson MDM, Raphael BJ. 2018. Hierarchical HotNet: Identifying hierarchies of altered subnetworks. *Bioinformatics* **34**: i972–i980.

Sharma A, Halu A, Decano JL, Padi M, Liu YY, Prasad RB, Fadista J, Santolini M, Menche J, Weiss ST, et al. 2018. Controllability in an islet specific regulatory network identifies the transcriptional factor NFATC4, which regulates Type 2 Diabetes associated genes. *npj Syst Biol Appl* **4**. http://dx.doi.org/10.1038/s41540-018-0057-0.

Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA. 2018. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat Rev Cancer* **18**: 696–705. http://dx.doi.org/10.1038/s41568-018-0060-1.

Valero C, Lee M, Hoen D, Wang J, Nadeem Z, Patel N, Postow MA, Shoushtari AN, Plitas G, Balachandran VP, et al. 2021. The association between tumor mutational burden and prognosis is dependent on treatment context. *Nat Genet* **53**: 11–15. http://dx.doi.org/10.1038/s41588-020-00752-4.

Werner RJ, Kelly AD, Issa JPJ. 2017. Epigenetics and Precision Oncology. *Cancer J (United States)* **23**: 262–269.

Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, Gautam B, Hassanali M. 2008. DrugBank: A knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res* **36**: 901–906.

Wuchty S. 2001. Scale-free behavior in protein domain networks. *Mol Biol Evol* **18**: 1694–1702.

Xia F, Liu J, Nie H, Fu Y, Wan L, Kong X. 2019. Random Walks : A Review of Algorithms and Applications. *IEEE Trans Emerg Top Comput Intell* **0**: 1–13.

Xie YH, Chen YX, Fang JY. 2020. Comprehensive review of targeted therapy for colorectal cancer. *Signal Transduct Target Ther* **5**: 1–30. http://dx.doi.org/10.1038/s41392-020-0116-z.

Xu C, Li X, Liu P, Li M, Luo F. 2019. Patient-derived xenograft mouse models: A high fidelity tool for individualized medicine (review). *Oncol Lett* **17**: 3–10.

Yadav A, Vidal M, Luck K. 2020. Precision medicine — networks to the rescue. *Curr Opin Biotechnol* **63**: 177–189. https://doi.org/10.1016/j.copbio.2020.02.005.

Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M. 2007. The importance of bottlenecks in protein networks: Correlation with gene essentiality and expression dynamics. *PLoS*

Comput Biol **3**: 713–720.

Zagidullin B, Aldahdooh J, Zheng S, Wang W, Wang Y, Saad J, Malyutina A, Jafari M, Tanoli Z, Pessia A, et al. 2019. DrugComb: An integrative cancer drug combination data portal. *Nucleic Acids Res* **47**: W43–W51.

Zhao S, Iyengar R. 2012. Systems Pharmacology: Network Analysis to Identify Multiscale Mechanisms of Drug Action. *Annu Rev Pharmacol Toxicol* **52**: 505–512.

Zhao X, Liu Z. 2019. Analysis of Topological Parameters of Complex Disease Genes Reveals the Importance of Location in a Biomolecular Network.

Zhong Q, Simonis N, Li QR, Charloteaux B, Heuze F, Klitgord N, Tam S, Yu H, Venkatesan K, Mou D, et al. 2009. Edgetic perturbation models of human inherited disorders. *Mol Syst Biol* **5**: 1–10.

Zhu J, Tucker M, Marin D, Gupta RT, Healy P, Humeniuk M, Jarvis C, Zhang T, McNamara M, George DJ, et al. 2019. Clinical utility of FoundationOne tissue molecular profiling in men with metastatic prostate cancer. *Urol Oncol Semin Orig Investig* **37**: 813.e1-813.e9. https://doi.org/10.1016/j.urolonc.2019.06.015.

Zielinski DC, Jamshidi N, Corbett AJ, Bordbar A, Thomas A, Palsson BO. 2017. Systems biology analysis of drivers underlying hallmarks of cancer cell metabolism. *Sci Rep* **7**: 1–14.

Zlotogora J. 2003. Penetrance and expressivity in the molecular age. *Genet Med* **5**: 347–352.

Figure 1. Individualized disease module concept and construction pipeline.

A) *Schematic illustrating how individualized disease modules are related to cohort inferred disease modules.* ***B)*** *Our construction pipeline begins with annotation of a generic protein-protein interaction network (PPI) with disease context and individualized –omics data. Following annotation, all shortest paths between diseased genes are detected and evaluated for disease activity. These paths are compared to randomly generated pathways via empirical p-value. Pathways that are less than empirical p-values of 0.01 are added to the individualized disease modules.* ***C*** *and* ***D)*** *Scatter plot of the number of* ***C)*** *mutated genes and* ***D****) nodes in the individualized disease modules versus the number of mutated genes in the tumor.*

Figure 2. Disease gene importance scoring process and scoring results across patients.

A) *The Individualized Disease Gene Importance (IDGI) scoring pipeline begins with the accessing the baseline parameters of the individualized disease modules. Next the disease gene is scored, and the shortest paths associated with the disease genes are removed from the individualized disease module. The change in parameters are then quantified into the IDGI score.* ***B)*** *Boxplots showing the distribution of IDGI scores for each patient. High IDGI scoring genes (outliers) are colored to show rare (< 5% mutated in TCGA BC) and frequent mutations (> 5% mutated in TCGA BC).*

Figure 3. Individualized disease modules displaying varied gene importance in two representative patients.

A) *and* ***B****) Individualized disease modules for two representative patients. Edge color intensity is determined by the distance between the mutated genes. Node size corresponds to IDGI score and node color reflects mutation frequency within the entire TCGA BRCA cohort.*

Figure 4. Heterogeneity of disease gene importance across patients and disease genes.

A) *Heatmap showing the IDGI scores for the most commonly mutated genes. Side histograms show the percent of patients with that gene mutated in our cohort (green) and the TCGA-BRCA project as a whole (grey).* ***B)*** *Boxplots for IDGI scores by most commonly mutated genes.*

Figure 5. Synthetic penetrance of disease genes within individualized disease modules.

Schematic demonstrating the concept of synthetic penetrance in three individualized disease modules (red, blue, and yellow). Synthetic penetrance is shown for two representative genes (pink and green) by showing the varied importance of disease genes in different patients.

Figure 6. Individualized disease modules for formulating singular and combinatorial individualized therapies.

A) *Dot plot showing the target scoring for breast cancer versus non-breast targets.* ***B****) Heatmap of the highest scoring gene targets.* ***C****) Dot plots of targeting score for each target.* ***D****) Paired boxplots showing the maximum single target and maximum combinatorial target score.* ***E)*** *Drug synergy differential (synergistic- additive) for this patient's therapeutic combinations. Red bars: positively synergistic combination. Blue bars: negatively synergistic combination.* ***F****) A single patient's top ten combinatorial therapies scored simultaneously (synergistic) and additively removal.*