Subnetwork-based prognostic biomarkers exhibit performance and robustness superior to gene-based biomarkers in breast cancer

Michal R. Grzadkowski¹, Syed Haider^{1,2}, Dorota H. Sendorek¹ and Paul C. Boutros^{1,3,4,§}

¹ Ontario Institute for Cancer Research, Toronto, Ontario, Canada M5G 0A3

² Centre for Molecular Oncology, Barts Cancer Institute, London EC1M 6BQ, United Kingdom

³ Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada M5G 2M9

⁴ Department of Pharmacology and Toxicology, University of Toronto, Toronto, Ontario, Canada M5S 1A8

§Corresponding author: paul.boutros@oicr.on.ca; 661 University Avenue, Suite 510; Toronto, Ontario, Canada; M5G 0A3; 416-673-8564

Author Emails:

Michal R. Grzadkowski mgrzad@gmail.com

Syed Haider syed.haider@oicr.on.ca

Dorota H. Sendorek dorota.sendorek@oicr.on.ca

Paul C. Boutros paul.boutros@oicr.on.ca

Keywords: prognostic signatures, subnetwork biomarker, gene biomarker, breast cancer

Abstract

Background

Effective classification of cancer patients into groups with differential survival remains an important and unsolved challenge. Biomarkers have been developed based on mRNA abundance data, but their replicability and clinical utility is modest. Integrating functional information, such as pathway data, has been suggested to improve biomarker performance. To date, however, the advantages of subnetwork-based biomarkers have not been quantified.

Results

We deeply sampled the population of prognostic gene-based and subnetwork-based biomarkers in a breast cancer meta-dataset of 4,960 patients. Analysing the performance and robustness of 22,000,000 gene biomarkers and 6,250,000 subnetwork biomarkers across twenty different training:testing cohort partitions of the meta-dataset revealed that subnetwork biomarkers exhibit superior overall performance and higher concordance across partitions. We find evidence of an upper bound for optimal biomarker size of ~200 genes or ~100 subnetworks. Additionally, with both biomarker feature types, larger biomarkers tend to show less consistency in performance across partitions, suggestive of over-fitting. Finally, an evaluation of varying training cohort sizes quantifies the effects of training cohort size.

Conclusions

Many groups are developing techniques for exploiting network-based representations of biological pathways to characterize cancer and other diseases. By considering the distribution of gene- and subnetwork-based biomarkers, we show that pathway data improves

performance and replicability, and that smaller biomarkers are more robust across patient cohorts. These insights may facilitate development of clinically useful biomarkers.

Background

Breast cancer continues to be the leading cause of neoplastic incidence and mortality in women worldwide [1]. A major obstacle to the effective diagnosis and treatment of breast tumours is their molecular heterogeneity, which contributes to widely divergent outcomes [2–4]. Techniques for accurate classification of breast cancer patients by risk profile or therapeutic response are therefore urgently needed as they enable more efficient allocation of clinical resources. A wide variety of models utilizing the mRNA abundance scores of a select set of genes, commonly referred to as “prognostic biomarkers”, have been developed for this purpose in breast cancer [5–11] as well as other tumour types and diseases [12–16].

Although several prognostic biomarkers have proven sufficiently effective to be successfully commercialized [5,6], there is no consensus on the best procedure for identifying useful and robust biomarkers. Furthermore, the persistent difficulty in replicating the reported performance of published prognostic biomarkers has led some to question the value of this approach [17–20]. A proposed method for improving the efficacy and robustness of biomarkers has been to leverage regulatory pathway information. Traditionally, a prognostic biomarker comprises a set of genes whose individual expression levels serve as input for a scoring algorithm that computes the individual risk levels for a patient sample cohort. However, more recently introduced biomarkers consist of *gene subnetworks*, with each

subnetwork being a set of genes identified *a priori* as sharing a common functionality via a separate dataset [21–26].

Subnetworks are most commonly derived from protein-protein interaction (PPI) networks since protein network data are relatively easy to obtain and contain information on gene-product interactions that it is believed cannot be inferred from mRNA abundance levels alone [27]. A *subnetwork biomarker* thus differs from a *gene biomarker* in that the expression levels of the genes it uses are first aggregated into a set of subnetwork expression scores before being fed into the risk scoring algorithm. We recently introduced a novel scoring system for subnetwork biomarkers constructed using PPI data [28]. This “SIMMS node-only” model (hereafter referred to simply as *SIMMS*) first calculates the dependence of patient survival on the mRNA abundance of each gene to get a gene risk factor. It then aggregates the gene-level mRNA abundance values within a subnetwork for each patient by taking their sum weighed by the gene risk factors. Several other subnetwork-based schemes have been proposed, and research in this area is active [29–31].

Several studies have cast doubt on the value of using multi-gene models compared to simpler univariate models by citing the performance of subnetwork biomarkers relative to gene biomarkers [32, 33]. This raises the question of whether time and resources increasingly designated for the development of novel feature selection algorithms as well as the implementation of increasingly composite data for biomarker discovery would be best reserved for research using a more traditional approach of simply identifying patient outcome-related gene sets (*i.e.* gene biomarkers). If building increasingly complex models using advanced machine learning methods and interaction network data yields only marginal

improvements in biomarker performance, then identifying informative groups of genes is clearly a more efficient approach to generating clinically relevant prognostic tools. To resolve this issue, we examined whether using subnetworks as the building blocks of prognostic biomarkers confers any advantage over orthodox gene-based approaches. The SIMMS model was generalized to score gene biomarkers as well as subnetwork biomarkers and then used to evaluate and compare the performance of 22,000,000 gene biomarkers and 6,250,000 subnetwork biomarkers. This experiment was performed with the same set of biomarkers on twenty different permutations of training and testing patient cohorts to test biomarker robustness. Our systematic evaluation of the behaviour of subnetwork biomarkers and gene biomarkers is an important contribution to determining the worth of network-based approaches in biomarker discovery.

Results and discussion

Using SIMMS to measure the performance of gene and subnetwork biomarkers

The prognostic performance of subnetwork biomarkers was evaluated on a 4,960 breast cancer patient meta-dataset using the SIMMS model as fully detailed in Methods [28]. Briefly, a univariate Cox proportional hazards model is used to estimate the association between mRNA abundance levels and survival information of the training cohort for each gene separately. Next, for each patient, the per-gene coefficients from the model are aggregated into subnetwork risk scores. Then, for each subnetwork biomarker, a multivariate Cox model is fit to the patient risk scores of the subnetworks comprising the biomarker.

Finally, the subnetwork coefficients of this model are used to generate patient risk scores for an independent testing cohort by combining them with the mRNA abundance values. A univariate Cox model is fit to the testing cohort (dichotomized by the training cohort's median risk score) and the resulting hazard ratio is used to assess the biomarker's prognostic capability.

In order to evaluate the prognostic capability of a given gene biomarker in a manner that is directly comparable to subnetwork biomarker performance, we modified the SIMMS model so that it handles individual genes as subnetworks (*i.e.* each gene is treated as a subnetwork of size one). We refer to this approach as *geneSIMMS*.

Jackknifing reveals patterns in gene and subnetwork biomarker performance

To study the characteristics of the population of gene biomarkers and subnetwork biomarkers, we randomly sampled 22,000,000 gene biomarkers and 6,250,000 subnetwork biomarkers by utilizing *jackknifing*, a resampling method in which samples are drawn from a population without replacement [34]. For this purpose, we only used genes for which mRNA abundance scores were available in all eighteen datasets comprising our breast cancer meta-dataset (**Table 1**). To ensure that our set of gene biomarkers contained the same underlying mRNA abundance data as our set of subnetwork biomarkers, all gene biomarkers were drawn from a pool of the 1,500 genes which were included in at least one of the 500 subnetworks from which the subnetwork biomarkers were created. Since gene biomarkers were drawn from a set of 1,500 features while subnetwork biomarkers were drawn from a set of 500 features, the population of all possible gene biomarkers is substantially larger than that of all possible

subnetwork biomarkers. This necessitated a larger test set of gene biomarkers to fully capture the null distribution of gene biomarker performance. The prognostic capability of each gene and subnetwork biomarker was estimated using SIMMS on twenty different partitions of the breast cancer meta-dataset, each of which was composed of varying training and testing cohort sizes (**Tables 2-3**). For full details on how the meta-dataset was partitioned, see Methods.

Considering only a single meta-dataset partition of equally-sized training and testing cohorts (the 'default' partition in **Table 2**), we found that the performance of gene and subnetwork biomarkers is highly sensitive to biomarker size (**Figure 1**). Adding more features to gene biomarkers results in substantial improvements in performance up to ~100 genes, at which point improvement in performance starts levelling off before peaking at ~175 genes. After this, larger gene biomarkers show a sharp and consistent decline in prognostic capability, particularly after biomarker size exceeds ~200 features. On the other hand, the performance of subnetwork biomarkers improves much more gradually with increasing biomarker size, plateauing at around 100 subnetworks.

Comparison of gene and subnetwork biomarkers across gene counts

To better contrast the performance of gene and subnetwork biomarkers, we first normalized the sizes of our subnetwork biomarkers to make them more directly comparable to our gene biomarkers. Since a subnetwork biomarker is essentially a set of gene lists, the most straightforward way to perform this normalization is to simply sum the sizes of all these gene lists to get the total number of genes in the biomarker. However, there is gene content overlap

in our set of 500 subnetworks so a single subnetwork biomarker may include multiple subnetworks that contain the same gene. Therefore, we additionally considered the number of unique genes in each biomarker. At each gene count, we compared the 95th to 99th percentiles of subnetwork biomarker performance against the 95th to 99th percentiles of gene biomarker performance (**Figure 2**).

The difference in performance between the two gene count normalization techniques is negligible as both gene and subnetwork biomarkers exhibit similar peak performance. Although gene biomarkers require fewer genes to reach optimal prognostic efficiency, they also show a higher propensity for over-fitting at higher gene counts.

Randomly sampling training:testing meta-dataset partitions reveals effect of training cohort size on biomarker performance and stability

To test the robustness of our finding that biomarker size affects performance, we re-evaluated our biomarker sets on a partition of the meta-dataset with a substantially smaller training cohort at only 36.5% of the total patient number. As seen in **Additional file 1: Supplementary Figure 1**, the general characteristics of biomarker performance with regard to biomarker size remain the same even though the performance of both gene and subnetwork biomarkers decreases slightly overall in this setting. However, subnetwork biomarkers do reach higher peak performance than gene biomarkers, suggesting that they are less sensitive to a suboptimal training environment. To verify these observations, we went on to create eighteen more meta-dataset partitions using random sampling for a total of ten partitions with larger training cohorts and ten partitions with smaller training cohorts.

As first evidenced by our initial finding, both gene and subnetwork biomarkers perform routinely better when trained on the larger patient cohorts (**Figure 3**). In only one dataset partition containing a smaller training cohort do biomarkers consistently outperform those trained on the larger cohorts. For most of the additional eighteen dataset partitions, the performance patterns resemble our initial two partitions. This is particularly true for gene biomarker performance which peaks rapidly at smaller biomarker sizes then drops with increasing biomarker size. Subnetwork biomarker performance peaks slowly, as seen previously, and then either plateaus or undergoes a more gradual decline.

To quantify biomarker performance stability across partitions we used Lin's concordance correlation coefficient (CCC) [35, 36]. The output of this metric ranges from -1, indicating complete discordance in biomarker performance across all partitions, to 1, indicating identical biomarker performance across all partitions. The stability of biomarker performance at individual biomarker sizes is only slightly greater in partitions with larger training cohorts compared to partitions with smaller training cohorts in gene and subnetwork biomarkers. However, when stability is measured using all biomarkers simultaneously this effect gets amplified (**Additional file 2: Supplementary Figure 2**). More significantly, both classes of biomarkers show the highest performance stability at smaller biomarker sizes, with steep declines in stability as more features are included. This demonstrates that although larger biomarkers may exhibit better prognostic performance, the superiority of individual large biomarkers is much harder to replicate. Therefore, partitioning a dataset to have fewer patients in the training cohort than in the testing cohort has a clear negative effect on both biomarker performance and stability. However, when only small biomarkers are considered,

the difference in stability is quite trivial. Since the stability of small gene and subnetwork biomarkers is very high, these biomarkers can be relied upon to have consistent performance when trained on patient cohorts of varying sizes.

Comparing subnetwork biomarkers to gene biomarkers reveals advantages of subnetworks as features

Finding a link between biomarker performance and feature size prompted us to explore the performance of gene and subnetwork biomarkers in more detail. A clearly defined optimal size was observed for gene biomarkers in all twenty meta-dataset permutations and we used this size as the standard against which subnetwork biomarkers would be judged. Subnetwork biomarker performance was normalized against peak gene biomarker performance in all twenty dataset permutations (**Figure 4**). This reveals that optimal subnetwork biomarker performance tends to be greater than optimal gene biomarker performance, an effect that is particularly pronounced in partitions with small training cohorts. Furthermore, in the majority of partitions where the performance of subnetwork biomarkers exceeds that of optimally-sized gene biomarkers, it does so at a wide range of feature and gene sizes (**Figure 4** and **Additional file 3: Supplementary Figure 3**). In particular, when considering only these partitions, the approximate biomarker sizes at which subnetwork biomarkers start to outperform optimal gene biomarkers is never greater than 50 features, 300 unique genes or 400 total genes.

Although the stability of gene biomarker performance is significantly higher than the stability of subnetwork biomarker performance when considering biomarkers of all feature sizes, at

comparable gene counts this relationship is inverted (**Figure 5**). The instability of subnetwork biomarkers that contain many large subnetworks obscures the fact that subnetwork biomarkers consisting of a small or moderate number of modules perform relatively consistently across different training:testing cohort partitions. As discussed above, subnetwork biomarkers tend to outperform even optimal gene biomarkers at fairly small sizes. Subnetwork biomarkers without an exorbitant number of features are, therefore, not only more stable than gene biomarkers but also have greater prognostic capabilities.

Conclusions

In this study, we applied jackknifing to a 4,960 breast cancer patient meta-dataset and identified two key advantages that subnetwork biomarkers have over gene biomarkers. In the majority of our randomly sampled partitions of training and testing patient cohorts, subnetwork biomarkers of various sizes outperform optimally-sized gene biomarkers. Additionally, subnetwork biomarkers show more consistency in their performance across different meta-dataset partitions than gene biomarkers of equal biomarker size (*i.e.* biomarkers containing the same number of features). In other words, a randomly selected subnetwork biomarker can be expected to do a better job of predicting high-risk patients repeatedly across numerous patient cohorts than a randomly selected gene biomarker of comparable size, provided that the number of genes comprising the biomarker does not exceed a relatively high threshold. The breadth of our analysis also allows us to make several other important observations about prognostic biomarker performance and robustness. Small training cohort sizes have a

greater effect on overall biomarker performance than on the consistency of said performance. Furthermore, the performance of large biomarkers is substantially more unstable than that of small biomarkers. Therefore, to optimize biomarker development, it is critical to maximize training cohort size and to minimize biomarker size.

Nevertheless, we recognize there are several caveats to be made regarding our findings that subnetwork biomarkers are superior to gene biomarkers. Because generally gene biomarkers will contain fewer genes than subnetwork biomarkers, and biomarkers with fewer genes are more robust, it is easier to find gene biomarkers with consistently high performance especially since gene biomarker performance peaks at a smaller number of features. Moreover, our study considers only the null distribution of biomarker performance. It is thus possible, albeit highly unlikely, that among the “rare” biomarkers, whose performance falls well above that of those that are randomly sampled, gene biomarkers are preferable to subnetwork biomarkers. Finally, there is a potential limitation in considering only 500 subnetworks and, in particular, only 1500 genes as these may not necessarily be representative of the actual human genome and the approximately 20,000 genes it contains. As information on PPI and gene pathways continues to accumulate, this matter can be thoroughly investigated in the future with access to a more complete and biologically accurate set of data.

Finally, deeply sampling the total biomarker space allows for a comprehensive evaluation of biomarker characteristics that is not attainable through analyses on a smaller number of biomarkers acquired using unsubstantiated feature selection strategies. Since experiment replicability has proven to be a great challenge in the development of clinically useful biomarkers, it is vital to study the performance of biomarkers in a variety of settings,

especially using different training and testing cohorts. Jackknifing and meta-dataset permutation are, therefore, two key tools for prognostic biomarker research and can be utilized to great effect in studies hindered by limited sample sizes and data access.

Methods

Preparation of mRNA abundance and pathway data

Datasets and subnetworks were prepared as described by Haider *et al* [28]. Raw breast cancer mRNA abundance datasets (**Table 1**) were normalized using the RMA algorithm [37] (R package affy v1.28.0) and probes were mapped to Entrez genes using custom CDFs (R packages hgu133ahsentrezgcdf v12.1.0, hgu133bhsentrezgcdf v12.1.0, hgu133plus2hsentrezgcdf v12.1.0, hthgu133ahsentrezgcdf v12.1.0 and hgu95av2hsentrezgcdf v12.1.0). Subnetworks were drawn from the NCI-Nature and BioCarta/Reactome databases and screened such that there was a maximum overlap of 80% between any two subnetworks and each subnetwork consisted of at least three genes. Genes without mRNA abundance values available in all eighteen datasets were removed along with genes that did not appear in any of the subnetworks. This resulted in a pool of 500 subnetworks and 1500 genes.

SIMMS and geneSIMMS

The original SIMMS node-only model trains a set of subnetworks on a cohort of patients to obtain a “module dysregulation score” (MDS) for each subnetwork. For each gene g , the

training cohort is median dichotomized by the mRNA abundance value of g . A univariate Cox proportional hazards model is then fit to the dichotomization by gene g using training cohort survival information and a hazard ratio HR_g is obtained (R package survival v2.36.14). A risk score for each combination of patient P and subnetwork S is computed using the formula

$$risk_{S,P} = \sum_{g \in S} \log_2(HR_g) \times mRNA_{g,P} \quad (1)$$

where $mRNA_{g,P}$ is the mRNA abundance in patient P of gene g contained within subnetwork S . Next, for each subnetwork biomarker B , a multivariate Cox proportional hazards model is fit to its subnetworks' training cohort's risk scores, as calculated in (1), and the training cohort survival data to get a Cox beta β_S for each subnetwork S in the biomarker. For each patient P , we thus calculate a risk score as measured by B using the formula

$$\sum_{S \in B} \beta_S \times risk_{S,P} \quad (2)$$

The median risk score is calculated for the training cohort and is then used as the break-point at which to dichotomize the testing cohort. A univariate Cox model is fit to this testing cohort and the risk classification is assessed by the resulting hazard ratio.

The SIMMS model was extended to gene biomarkers by treating each gene as a subnetwork of size one. We refer to this as the geneSIMMS model. The risk score for patient P and gene g is thus given by

$$risk_{g,P} = \log_2(HR_g) \times mRNA_{g,P} \quad (3)$$

with HR_g and $mRNA_{g,P}$ defined as in (1). For a gene biomarker B , a Cox beta β_g is calculated using a multivariate model for each gene g in the biomarker (as is done with the subnetworks in a subnetwork biomarker). For each patient P , we thus calculate a risk score as measured

by B using the formula

$$\sum_{g \in B} \beta_g \times risk_{g,P} \quad (4)$$

This is followed by patient dichotomization of the testing cohort and then biomarker performance evaluation, same as described above for SIMMS.

Biomarker jackknifing

Jackknifing is a resampling method in which samples are drawn from a population without replacement [34]. This technique was applied to estimate the characteristics of the gene and subnetwork biomarker populations. Twenty-five subnetwork biomarker sizes were chosen (all multiples of five between 5 and 100 inclusive and all multiples of ten between 110 and 150 inclusive) along with forty-four gene biomarker sizes (all multiples of five between 5 and 150 inclusive, all multiples of twenty-five between 175 and 400 inclusive and all multiples of one hundred between 500 and 800 inclusive). 250,000 biomarkers were jackknifed at each subnetwork biomarker size for a total of 6,250,000 biomarkers and 500,000 biomarkers were jackknifed at each gene biomarker size for a total of 22,000,000 biomarkers.

The jackknifing was performed to sample more gene biomarkers than subnetwork biomarkers since the pool of potential gene biomarkers is much larger due to there being a greater number of gene features ($n = 1500$) than subnetwork features ($n = 500$). Because of the large computational expense of using signatures with a high feature number, gene biomarkers comprising 500 or more genes were only tested on the two initial (default) meta-dataset partitions.

Normalizing subnetwork biomarker size

A subnetwork biomarker S of size n consists of a set of modules $M_1, M_2, \dots, M_n$, each of which is a list of genes. The total number of genes in S is therefore

$$\sum_{i=1}^n |M_i| \quad (5)$$

i.e. the sum of the sizes of the modules comprising S . The number of unique genes in S is

$$\left| \bigcup_{i=1}^n M_i \right| \quad (6)$$

i.e. the size of the set of genes appearing in at least one of the modules comprising S . A gene biomarker G of size n consists of a set of n unique genes, so the total number of genes and the number of unique genes in G both equal n .

Meta-dataset partitioning and concordance correlation calculations

As seen in **Table 1**, our breast cancer meta-dataset consists of eighteen different datasets. To create a partition of this meta-dataset into training and testing patient cohorts, we randomly chose a subset of the eighteen datasets to use for training such that the total number of patients in this subset satisfied some bound and the remaining datasets served as the testing cohort. For partitions with large training cohorts, this bound was at least 47% of the patients in the meta-dataset (**Table 2**). For partitions with small training cohorts this bound was between 33% and 40% of the patients in the meta-dataset (**Table 3**). To ensure random sampling over all partitions satisfying a given bound, we first exhaustively identified all possible partitions and then sampled from that set.

The concordance correlation coefficient for multiple classes was used as first defined and

later amended by Lin [35, 36]. Each dataset partition was treated as a class (or observer) in our analysis. Therefore, all gene biomarkers containing 400 or fewer genes and all subnetwork biomarkers had twenty matched observations of their performance as measured by the raw hazard ratio returned by SIMMS or geneSIMMS.

Comparing subnetwork biomarker performance to peak gene biomarker performance

To determine peak gene biomarker performance on a given meta-dataset partition, we considered the 99th percentile of biomarker performance at each biomarker size as measured by the log-transformed Cox model hazard ratio. The biomarker size which had the highest such percentile of performance defined the optimal size and the corresponding 99th percentile was set as the peak gene biomarker performance for the partition. To compare subnetwork biomarker performance at a given size to this peak, we again took the 99th percentile of performance as measured by the log-transformed hazard ratio and calculated the difference of subnetwork biomarker performance and peak gene biomarker performance to obtain a “relative” performance.

Visualizations

All figures were created in the R statistical environment (v2.15.3) using the lattice (v0.20-16), latticeExtra (v0.6-24), RColorBrewer (v1.0-5), BPG (v5.6.8) [38] and cluster (v1.14.3) packages.

List of abbreviations used

CCC: Lin's concordance correlation coefficient

PPI: protein-protein interaction

SIMMS: subnetwork integration for multi-modal signatures

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets supporting the conclusions of this article are included within the article and its additional files, as well as obtained from the ArrayExpress, Gene Expression Omnibus (GEO) and European Genome-phenome Archive (EGA) repositories:

- Bild: <https://www.ebi.ac.uk/arrayexpress/experiments/E-GEOD-3143>
- Chin: <https://www.ebi.ac.uk/arrayexpress/experiments/E-TABM-158>
- Desmedt-1: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE7390>
- Desmedt-2: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE16446>
- Hatzis-1: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE25055>
- Hatzis-2: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE25065>
- Ivshina: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE4922>
- Miller: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE3494>
- Pawitan: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE1456>

- Sabatier: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE21653>
- Schmidt: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE11121>
- Sotiriou: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE6532>
- Symmans-JBI: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE17700>
- Symmans-MDA: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE17705>
- Wang: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE2034>
- Zhang: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12093>
- Metabric-Training: <https://www.ebi.ac.uk/ega/datasets/EGAD00010000210>
- Metabric-Validation: <https://www.ebi.ac.uk/ega/datasets/EGAD00010000211>

Competing interests

All authors declare that they have no competing interests.

Funding

This study was conducted with the support of the Ontario Institute for Cancer Research to PCB through funding provided by the government of Ontario. PCB was supported by a TFRI New Investigator Award and a CIHR New Investigator Award.

Authors' contributions

Initiated the project: MRG, PCB

Developed the SIMMS model: SH

Extended the SIMMS model to gene biomarkers: MRG

Performed biomarker jackknifing and meta-dataset permutation analysis: MRG

Wrote the first draft of the manuscript: MRG, DHS

Supervised research: PCB

Acknowledgements

The authors thank all members of the Boutros Lab for their helpful suggestions.

References

1. Stewart BW, Wild CP. World Cancer Report 2014. Lyon: 2014.
2. Perou CM, Sørli T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lønning PE, Børresen-Dale AL, Brown PO, Botstein D. Molecular portraits of human breast tumours. *Nature*. 2000;406(6797):747–52.
3. Garcia-Closas M, Hall P, Nevanlinna H, Pooley K, Morrison J, Richesson DA, Bojesen SE, Nordestgaard BG, Axelsson CK, Arias JI, Milne RL, Ribas G, González-Neira A, Benítez J, Zamora P, Brauch H, Justenhoven C, Hamann U, Ko YD, Bruening T, Haas S, Dörk T, Schürmann P, Hillemanns P, Bogdanova N, Bremer M, Karstens JH, Fagerholm R, Aaltonen K, Aittomäki K, von Smitten K, Blomqvist C, Mannermaa A, Uusitupa M, Eskelinen M, Tengström M, Kosma VM, Kataja V, Chenevix-Trench G, Spurdle AB, Beesley J, Chen X; Australian Ovarian Cancer Management Group; Kathleen Cuninghame Foundation Consortium for Research into Familial Breast Cancer, Devilee P, van Asperen CJ, Jacobi CE, Tollenaar RA, Huijts PE, Klijn JG, Chang-Claude J, Kropp S, Slanger T, Flesch-Janys D, Mutschelknauss E, Salazar R, Wang-

- Gohrke S, Couch F, Goode EL, Olson JE, Vachon C, Fredericksen ZS, Giles GG, Baglietto L, Severi G, Hopper JL, English DR, Southey MC, Haiman CA, Henderson BE, Kolonel LN, Le Marchand L, Stram DO, Hunter DJ, Hankinson SE, Cox DG, Tamimi R, Kraft P, Sherman ME, Chanock SJ, Lissowska J, Brinton LA, Peplonska B, Klijn JG, Hoening MJ, Meijers-Heijboer H, Collee JM, van den Ouweland A, Uitterlinden AG, Liu J, Lin LY, Yuqing L, Humphreys K, Czene K, Cox A, Balasubramanian SP, Cross SS, Reed MW, Blows F, Driver K, Dunning A, Tyrer J, Ponder BA, Sangrajrang S, Brennan P, McKay J, Odefrey F, Gabrieau V, Sigurdson A, Doody M, Struwing JP, Alexander B, Easton DF, Pharoah PD. Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. *PLoS Genet.* 2008;4(4):e1000054.
4. Voduc KD, Cheang MC, Tyldesley S, Gelmon K, Nielsen TO, Kennecke H. Breast cancer subtypes and the risk of local and regional relapse. *J Clin Oncol.* 2010;28(10):1684–91.
5. van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH. Gene expression profiling predicts clinical outcome of breast cancer. *Nature.* 2002;15:530–6.
6. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Park T, Hiller W, Fisher ER, Wickerham DL, Bryant J, Wolmark N. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med.*

- 2004;351(27):2817–26.
7. Ma Y, Qian Y, Wei L, Abraham J, Shi X, Castranova V, Harner EJ, Flynn DC, Guo L. Population-based molecular prognosis of breast cancer by transcriptional profiling. Clin Cancer Res. 2007;13(7):2014–22.
8. Naderi A, Teschendorff A, Barbosa-Morais NL, Pinder SE, Green AR, Powe DG, Robertson JF, Aparicio S, Ellis IO, Brenton JD, Caldas C. A gene-expression signature to predict survival in breast cancer across independent data sets. Oncogene. 2007;26(10):1507–16.
9. Wan Y-W, Qian Y, Rathnagiriswaran S, Castranova V, Guo NL. A breast cancer prognostic signature predicts clinical outcomes in multiple tumor types. Oncol Rep. 2010;24(2):489–94.
10. Sotiriou C, Pusztai L. Gene-expression signatures in breast cancer. N Engl J Med. 2009;360(8):790–800.
11. Duan Q, Kou Y, Clark NR, Gordonov S, Ma'ayan A. Metasignatures identify two major subtypes of breast cancer. CPT Pharmacometrics Syst Pharmacol. 2013;2:e35.
12. Sun Z, Wigle DA, Yang P. Non-overlapping and non-cell-type-specific gene expression signatures predict lung cancer survival. J Clin Oncol. 2008;26(6):877–83.
13. Boutros PC, Lau SK, Pintilie M, Liu N, Shepherd FA, Der SD, Tsao M-S, Penn LZ, Jurisica I. Prognostic gene signatures for non-small-cell lung cancer. Proc Natl Acad Sci U S A. 2009;106(8):2824–8.

14. Sabatier R, Finetti P, Bonense J, Jacquemier J, Adelaide J, Lambaudie E, Viens P, Birnbaum D, Bertucci F. A seven-gene prognostic model for platinum-treated ovarian carcinomas. *Br J Cancer*. 2011;105(2):304–11.
15. Hoshida Y, Villanueva A, Sangiovanni A, Sole M, Hur C, Andersson KL, Chung RT, Gould J, Kojima K, Gupta S, Taylor B, Crenshaw A, Gabriel S, Minguez B, Iavarone M, Friedman SL, Colombo M, Llovet JM, Golub TR. Prognostic gene expression signature for patients with hepatitis C-related early-stage cirrhosis. *Gastroenterology*. 2013;144(5):1024–30.
16. Glatt SJ, Stone WS, Nossova N, Liew C-C, Seidman LJ, Tsuang MT. Similarities and differences in peripheral blood gene expression signatures of individuals with schizophrenia and their first-degree biological relatives. *Am J Med Genet B Neuropsychiatr Genet*. 2011;156B(8):869–87.
17. Buyse M, Sargent DJ, Grothey A, Matheson A, de Gramont A. Biomarkers and surrogate end points--the challenge of statistical validation. *Nat Rev Clin Oncol*. 2010;7(6):309–17.
18. Subramanian J, Simon R. Gene expression-based prognostic signatures in lung cancer: ready for clinical use? *J Natl Cancer Inst*. 2010;102(7):464–74.
19. Ein-Dor L, Zuk O, Domany E. Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. *Proc Natl Acad Sci U S A*. 2006;103(15):5923–8.

20. Chibon F. Cancer gene expression signatures - the rise and fall? *Eur J Cancer*. 2013;49(8):2000–9.
21. Dao P, Colak R, Salari R, Moser F, Davicioni E, Schönhuth A, Ester M. Inferring cancer subnetwork markers using density-constrained biclustering. *Bioinformatics*. 2010;26(18):i625–31.
22. Chuang H-Y, Lee E, Liu Y-T, Lee D, Ideker T. Network-based classification of breast cancer metastasis. *Mol Syst Biol*. 2007;3:140.
23. Abraham G, Kowalczyk A, Loi S, Haviv I, Zobel J. Prediction of breast cancer prognosis using gene set statistics provides signature stability and biological context. *BMC Bioinformatics*. 2010;11:277.
24. Vaske CJ, Benz SC, Sanborn JZ, Earl D, Szeto C, Zhu J, Haussler D, Stuart JM. Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. *Bioinformatics*. 2010;26(12):i237–45.
25. Zhang W, Ota T, Shridhar V, Chien J, Wu B, Kuang R. Network-based survival analysis reveals subnetwork signatures for predicting outcomes of ovarian cancer treatment. *PLoS Comput Biol*. 2013;9(3):e1002975.
26. Liu K-Q, Liu Z-P, Hao J-K, Chen L, Zhao X-M. Identifying dysregulated pathways in cancers from pathway interaction networks. *BMC Bioinformatics*. 2012;13:126.
27. Chowdhury SA, Nibbe RK, Chance MR, Koyutürk M. Subnetwork state functions define dysregulated subnetworks in cancer. *J Comput Biol*. 2011;18(3):263–81.

28. Haider S, Yao C, Sabine V, Grzadkowski M, Stimper V, Starmans MH, Wang J, Nguyen F, Moon NC, Lin X, Drake C, Crozier CA, Brookes C, van de Velde CJH, Hasenburg A, Kieback D, Markopoulos CJ, Dirix LY, Seynaeve C, Rea DW, Kasprzyk A, Lio P, Lambin P, Bartlett JMS, Boutros PC. Network-based biomarkers enable cross-disease biomarker discovery. *BioRxiv* 2018; doi: <https://doi.org/10.1101/289934>.
29. Chuang HY, Lee E, Liu YT, Lee D, Ideker T. Network-based classification of breast cancer metastasis. *Mol Syst Biol.* 2007;3:140.
30. Wu G, Stein L. A network module-based method for identifying cancer prognostic signatures. *Genome Biol.* 2012;13(12):R112.
31. Li W, Bai X, Hu E, Huang H, Li Y, He Y, Lv J, Chen L, He W. Identification of breast cancer prognostic modules based on weighted protein-protein interaction networks. *Oncol Lett.* 2017;13(5):3935–41.
32. Haibe-Kains B, Desmedt C, Sotiriou C, Bontempi G. A comparative study of survival models for breast cancer prognostication based on microarray data: does a single gene beat them all? *Bioinformatics.* 2008;24(1):2200–8.
33. Staiger C, Cadot S, Kooter R, Dittrich MT, Müller T, Klau GW, Wessels LF. A critical evaluation of network and pathway-based classifiers for outcome prediction in breast cancer. *PLoS One.* 2012;7(4):e34796.
34. Miller RG. The jackknife – a review. *Biometrika.* 1974;61(1):1–15.
35. Lin LI-K. A concordance correlation coefficient to evaluate reproducibility. *Biometrics.*

1989;45(1):255–68.

36. Lin LI-K. A note on the concordance correlation coefficient. *Biometrics*. 2000;56:324–5.

37. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*. 2003;4(2):249–64.

38. P'ng C, Green J, Chong LC, Waggott D, Prokopec SD, Shamsi M, Nguyen F, Mak DYF, Lam F, Albuquerque MA, Wu Y, Jung EH, Starmans MHW, Chan-Seng-Yue MA, Yao CQ, Liang B, Lalonde E, Haider S, Simone NA, Sendorek D, Chu KC, Moon NC, Fox NS, Grzadkowski MR, Harding NJ, Fung C, Murdoch AR, Houlahan KE, Wang J, Garcia DR, de Borja R, Sun RX, Lin X, Chen GM, Lu A, Shiah Y-J, Zia A, Kearns R, Boutros P. BPG: Seamless, automated and interactive visualization of scientific data. *bioRxiv* 2017; doi: <https://doi.org/10.1101/156067>.

Figure Legends

Figure 1: Biomarker performance percentiles on a 50:50 training:testing meta-dataset partition

250,000 subnetwork biomarkers were sampled at 25 biomarker sizes ranging from 5 to 150 for a total of 6,250,000 subnetwork biomarkers. 500,000 gene biomarkers were sampled using the same pool of genes at 44 biomarker sizes ranging from 5 to 800 for a total of

22,000,000 gene biomarkers. Each biomarker was trained on an initial cohort of 2,490 patients and then applied to dichotomize an independent testing cohort of 2,470 patients using the SIMMS algorithm. The various percentiles of biomarker performance are displayed by biomarker size for **(A)** gene biomarkers and **(B)** subnetwork biomarkers. Performance is defined by the hazard ratio of the Cox model fit to the biomarker dichotomized testing cohort survival data.

Figure 2: Gene versus subnetwork biomarker performance on 50:50 meta-dataset partition using two biomarker size normalization techniques

For each subnetwork biomarker, we obtained the total number of genes from across all individual subnetworks as well as the number of unique genes. The range between the 95th and 99th percentiles of biomarker performance was re-calculated by grouping biomarkers according to **(A)** the total number of genes and **(B)** the number of unique genes. Note that because the features of a gene biomarker are by definition always a unique set of genes, the total number of genes and the number of unique genes are the same in this case.

Figure 3: Gene biomarker versus subnetwork biomarker performance across all twenty meta-dataset partitions

Twenty unique partitions of the breast cancer meta-dataset into training and testing cohorts were created such that ten partitions would have at least 47% of the total number of patients to train on (*large training set*) and the remaining ten partitions would have at most 40% of the patients to train on (*small training set*). The performance of 20 million gene biomarkers (the original set of 22 million biomarkers minus those of over 400 genes in size) and the original

set of 6,250,000 subnetwork biomarkers was re-evaluated on each of these training:testing cohort partitions. The 99th percentile of **(A)** gene biomarker performance and **(B)** subnetwork biomarker performance on each of the meta-dataset partitions by biomarker size, along with the 99th percentile of subnetwork biomarker performance by **(C)** the total number of genes and by **(D)** the number of unique genes.

Figure 4: Subnetwork biomarker performance relative to peak gene biomarker performance

On each meta-dataset partition, the peak gene biomarker performance was calculated by taking the maximum of the 99th percentile of jackknifed biomarker performance across all biomarker sizes. The 99th percentile of jackknifed subnetwork biomarker performance at each biomarker size was normalized to this peak on each partition by taking the difference between their Cox model betas. Relative subnetwork biomarker performance on **(A,B)** partitions with large training sets (in blue) and on **(C,D)** partitions with small training sets (in red) by **(A,C)** biomarker size (*i.e.* the number of features) and by **(B,D)** the number of unique biomarker genes. For relative subnetwork performance by total number of biomarker genes, see **Additional file 3: Supplementary Figure 3.**

Figure 5: Subnetwork biomarker concordance versus gene biomarker concordance

Lin's concordance correlation coefficient (CCC) was calculated for gene and subnetwork biomarkers using the twenty meta-dataset partitions as independent observers of biomarker performance. This was done at **(A)** each biomarker size (*i.e.* the number of features) and at **(B)** each total gene count. Dotted lines indicate the CCC calculated for gene and subnetwork

biomarkers of all sizes combined.

Tables

Table 1: Breast cancer dataset sizes

Source	Number of Samples
Bild	158
Chin	129
Desmedt-1	198
Desmedt-2	107
Hatzis-1	310
Hatzis-2	198
Ivshina	249
Metabric-Training	996
Metabric-Validation	992
Miller	236
Pawitan	159
Sabatier	252
Schmidt	200
Sotiriou	94
Symmans-JBI	65
Symmans-MDA	195
Wang	286
Zhang	136

Table 2: Meta-dataset partitions with larger training cohorts

Partition	Training datasets	Testing datasets	% of cohort used for training	Training Size	Testing Size
Default (1)	Chin, Hatzis-1, Ivshina, Metabric-Training, Miller, Pawitan, Sabatier, Sotiriou, Symmans-	Bild, Desmedt-1, Desmedt-2, Hatzis-2, Metabric-Validation, Schmidt, Symmans-MDA,	50.20	2490	2470

	JB1	Wang, Zhang			
2	Bild, Desmedt-2, Hatzis-1, Hatzis-2, Metabric-Validation, Miller, Pawitan, Sabatier, Zhang	Chin, Desmedt-1, Ivshina, Metabric-Training, Schmidt, Sotiriou, Symmans-JB1, Symmans-MDA, Wang	51.37	2548	2412
3	Bild, Chin, Desmedt-1, Hatzis-1, Metabric-Training, Pawitan, Wang, Zhang	Desmedt-2, Hatzis-2, Ivshina, Metabric-Validation, Miller, Sabatier, Schmidt, Sotiriou, Symmans-JB1, Symmans-MDA	47.82	2372	2588
4	Bild, Chin, Desmedt-1, Hatzis-1, Ivshina, Metabric-Validation, Miller, Schmidt, Symmans-JB1	Desmedt-2, Hatzis-2, Metabric-Training, Pawitan, Sabatier, Sotiriou, Symmans-MDA, Wang, Zhang	51.15	2537	2423
5	Chin, Hatzis-1, Ivshina, Metabric-Validation, Schmidt, Symmans-JB1, Wang, Zhang	Bild, Desmedt-1, Desmedt-2, Hatzis-2, Metabric-Training, Miller, Pawitan, Sabatier, Sotiriou, Symmans-MDA	47.72	2367	2593
6	Bild, Hatzis-1, Hatzis-2, Metabric-Validation, Pawitan, Sabatier, Schmidt, Symmans-JB1	Chin, Desmedt-1, Desmedt-2, Ivshina, Metabric-Training, Miller, Sotiriou, Symmans-MDA, Wang, Zhang	47.06	2334	2626
7	Desmedt-1, Hatzis-2, Ivshina, Metabric-Validation, Sotiriou, Symmans-MDA, Wang, Zhang	Bild, Chin, Desmedt-2, Hatzis-1, Metabric-Training, Miller, Pawitan, Sabatier, Schmidt, Symmans-JB1	47.34	2348	2612
8	Bild, Hatzis-1, Ivshina, Metabric-Training, Miller, Pawitan, Schmidt,	Chin, Desmedt-1, Desmedt-2, Hatzis-2, Metabric-Validation, Sabatier,	49.74	2467	2493

9	Sotiriou, Symmans-JBI Chin, Desmedt-2, Hatzis-2, Ivshina, Metabric-Validation, Sabatier, Sotiriou, Symmans-JBI, Wang	Symmans-MDA, Wang, Zhang Bild, Desmedt-1, Hatzis-1, Metabric-Training, Miller, Pawitan, Schmidt, Symmans-MDA, Zhang	47.82	2372	2588
10	Bild, Chin, Desmedt-2, Hatzis-1, Hatzis-2, Ivshina, Metabric-Training, Pawitan, Schmidt, Symmans-MDA, Wang	Desmedt-1, Metabric-Validation, Miller, Sabatier, Sotiriou, Symmans-JBI, Zhang	60.22	2987	1973

Table 3: Meta-dataset partitions with smaller training cohorts

Partition	Training datasets	Testing datasets	% of cohort used for training	Training Size	Testing Size
Default (1)	Bild, Chin, Desmedt-2, Hatzis-1, Hatzis-2, Miller, Pawitan, Sabatier, Symmans-JBI, Symmans-MDA	Desmedt-1, Ivshina, Metabric-Training, Metabric-Validation, Schmidt, Sotiriou, Wang, Zhang	36.47	1809	3151
2	Desmedt-1, Desmedt-2, Hatzis-1, Hatzis-2, Ivshina, Miller, Schmidt, Symmans-MDA, Zhang	Bild, Chin, Metabric-Training, Metabric-Validation, Pawitan, Sabatier, Sotiriou, Symmans-JBI, Wang	36.88	1829	3131
3	Bild, Desmedt-1, Hatzis-2, Ivshina, Metabric-Validation	Chin, Desmedt-2, Hatzis-1, Metabric-Training, Miller, Pawitan, Sabatier, Schmidt, Sotiriou, Symmans-JBI, Symmans-MDA,	36.19	1795	3165

4	Bild, Desmedt-1, Metabric-Training, Miller, Pawitan, Symmans-JBI	Wang, Zhang Chin, Desmedt-2, Hatzis-1, Hatzis-2, Ivshina, Metabric-Validation, Sabatier, Schmidt, Sotiriou, Symmans-MDA, Wang, Zhang	36.53	1812	3148
5	Bild, Desmedt-2, Hatzis-1, Hatzis-2, Ivshina, Schmidt, Sotiriou, Symmans-JBI, Wang, Zhang	Chin, Desmedt-1, Metabric-Training, Metabric-Validation, Miller, Pawitan, Sabatier, Symmans-MDA	36.35	1803	3157
6	Bild, Desmedt-1, Desmedt-2, Hatzis-2, Sabatier, Schmidt, Sotiriou, Symmans-MDA, Wang, Zhang	Chin, Hatzis-1, Ivshina, Metabric-Training, Metabric-Validation, Miller, Pawitan, Symmans-JBI	36.77	1824	3136
7	Bild, Chin, Desmedt-1, Desmedt-2, Hatzis-1, Ivshina, Pawitan, Sabatier, Symmans-MDA	Hatzis-2, Metabric-Training, Metabric-Validation, Miller, Schmidt, Sotiriou, Symmans-JBI, Wang, Zhang	35.42	1757	3203
8	Desmedt-2, Metabric-Training, Miller, Sotiriou, Symmans-JBI, Symmans-MDA	Bild, Chin, Desmedt-1, Hatzis-1, Hatzis-2, Ivshina, Metabric-Validation, Pawitan, Sabatier, Schmidt, Wang, Zhang	34.13	1693	3267
9	Bild, Desmedt-1, Desmedt-2, Ivshina, Pawitan, Sabatier, Schmidt, Sotiriou, Wang	Chin, Hatzis-1, Hatzis-2, Metabric-Training, Metabric-Validation, Miller, Symmans-JBI, Symmans-MDA, Zhang	34.33	1703	3257
10	Bild, Chin, Ivshina,	Desmedt-1,	34.72	1722	3238

	Pawitan, Sabatier, Schmidt, Sotiriou, Symmans-MDA, Wang	Desmedt-2, Hatzis- 1, Hatzis-2, Metabric-Training, Metabric-Validation, Miller, Symmans- JBI, Zhang			
--	--	--	--	--	--

Additional files

Additional file 1: Supplementary Figure 1.

TIFF 4.5 Mb

Gene biomarker versus subnetwork biomarker performance using a 37/63 meta-dataset partition. The performances of the 22,000,000 gene biomarkers and 6,250,000 subnetwork biomarkers sampled were re-evaluated using a smaller training set. From the total 4,960 patient meta-dataset, 1,809 patients were randomly drawn for training and the remaining 3,151 patients were used to test dichotomization efficacy. The various percentiles of biomarker performance using this smaller training cohort by biomarker size for **(A)** gene biomarkers and **(B)** subnetwork biomarkers, along with the range of biomarker performance between the 95th and 99th percentiles by **(C)** the total number of feature genes and by **(D)** the number of unique feature genes for both biomarker types.

Additional file 2: Supplementary Figure 2.

TIFF 4.8 Mb

Biomarker performance concordance across small training cohorts and large training cohorts. Lin's concordance correlation coefficient (CCC) was calculated treating meta-

dataset partitions as independent observers of biomarker performance. Partitions were grouped together using the two training cohort size classes ('small' and 'large') and the CCC for each partition class was calculated for **(A)** gene biomarkers and **(B)** subnetwork biomarkers at each biomarker size and also for subnetwork biomarkers at each **(C)** total gene count and **(D)** unique gene count. Dotted lines indicate the CCC calculated for each partition class using all biomarkers at once.

Additional file 3: Supplementary Figure 3.

TIFF 6.2 Mb

Subnetwork biomarker performance relative to gene biomarker performance by the total number of genes. On each meta-dataset partition, the peak gene biomarker performance was calculated by taking the maximum of the 99th percentile of jackknifed biomarker performance across all biomarker sizes. The 99th percentile of jackknifed subnetwork biomarker performance at each biomarker size was normalized to this peak on each partition by taking the difference between the Cox model betas. Relative subnetwork biomarker performance on **(A)** partitions with large training sets and on **(B)** partitions with small training sets by the total number of biomarker genes.