

RESEARCH

Open Access

Network-based stratification analysis of 13 major cancer types using mutations in panels of cancer genes

Xue Zhong^{1,2}, Hushan Yang³, Shuyang Zhao⁴, Yu Shyr^{1,2,5*}, Bingshan Li^{2,6*}

From The International Conference on Intelligent Biology and Medicine (ICIBM) 2014
San Antonio, TX, USA. 04-06 December 2014

Abstract

Background: Cancers are complex diseases with heterogeneous genetic causes and clinical outcomes. It is critical to classify patients into subtypes and associate the subtypes with clinical outcomes for better prognosis and treatment. Large-scale studies have comprehensively identified somatic mutations across multiple tumor types, providing rich datasets for classifying patients based on genomic mutations. One challenge associated with this task is that mutations are rarely shared across patients. Network-based stratification (NBS) approaches have been proposed to overcome this challenge and used to classify tumors based on exome-level mutations. In routine research and clinical applications, however, usually only a small panel of pre-selected genes is screened for mutations. It is unknown whether such small panels are effective in classifying patients into clinically meaningful subtypes.

Results: In this study, we applied NBS to 13 major cancer types with exome-level mutation data and compared the classification based on the full exome data with those focusing only on small sets of genes. Specifically, we investigated three panels, FoundationOne (240 genes), PanCan (127 genes) and TruSeq (48 genes). We showed that small panels not only are effective in clustering tumors but also often outperform full exome data for most cancer types. We further associated subtypes with clinical data and identified 5 tumor types (CRC-Colorectal carcinoma, HNSC-Head and neck squamous cell carcinoma, KIRC-Kidney renal clear cell carcinoma, LUAD-Lung adenocarcinoma and UCEC-Uterine corpus endometrial carcinoma) whose subtypes are significantly associated with overall survival, all based on small panels.

Conclusion: Our analyses indicate that effective patient subtyping can be carried out using mutations detected in smaller gene panels, probably due to the enrichment of clinically important genes in such panels.

Background

Cancers are complex diseases with highly heterogeneous causes and clinical outcomes. At the molecular level histologically and clinically similar patients often exhibit drastically distinct genomic aberrations. Large-scale studies such as The Cancer Genome Atlas (TCGA) have comprehensively cataloged multi-layered genomic aberrations in multiple cancers [1,2]. Identification of clinically meaningful subtypes of tumors based on molecular patterns is

critical to provide insights into the biological mechanisms of tumor progression and to guide better treatment and prognosis. Most previous studies of tumor classification utilized messenger RNA (mRNA) expression [3-5], which often yielded subtypes that are not highly predictive of clinical outcomes [1,6]. On the other hand, somatic mutations, which often disrupt the function of mutated genes, provide insights not only to the mechanisms of tumorigenesis and progression but also the candidates for targeted therapy [7-9]. Therefore classification of patients based on somatic mutations may provide more effective clinical guidance. However, mutations are rarely shared across patients [10,11] so that the similarity between

* Correspondence: yu.shyr@vanderbilt.edu; bingshan.li@vanderbilt.edu

¹Department of Biostatistics, Vanderbilt University, Nashville, TN, USA

²Center for Quantitative Sciences, Vanderbilt University, Nashville, TN, USA

Full list of author information is available at the end of the article

tumors cannot be directly measured based on mutated genes. Network-based stratification (NBS) was recently proposed to overcome this challenge by leveraging information provided in protein-protein interaction networks (PPI) [12]. Briefly, NBS uses label propagation on PPI to assign higher values to non-mutated genes that are closer to genes (in PPI) that harbor mutations. This guilt-by-association principle governed by genetic networks has many applications for biological discovery utilizing prior knowledge [13,14]. For somatic mutations in genes, this principle fits well with the underlying biology: driver genes are often interacting directly or indirectly in common pathways and mutations in different genes in the same pathway are likely to cause genetically similar tumors [11,15-18]. NBS has been applied on several cancers using exome-level mutation data and showed improved association of subtypes with clinical outcomes than using mRNA data [12]. In general NBS provides a unified framework to further investigate tumor subtyping by integrating somatic mutations with biological networks.

Exome-level mutations were used in previous NBS analyses. In routine research and clinical application, instead of exome sequencing, a viable cost-effective alternative is to screen mutations in a panel of pre-selected cancer genes [19,20]. It is unknown whether such small panels are effective in classifying patients into clinically meaningful subtypes. Although exome sequencing provides comprehensive characterization of coding mutations, it is likely that a large portion of mutations are passengers, as it was estimated that few mutations in a patient are drivers (e.g. ranging from 2 to 8) [16,21]. Such passengers, if included in the analysis, may obscure clinically and biologically important mutations. On the other hand, gene panels are usually designed to include known driver genes or genes involved in important pathways, resulting in highly enriched signal vs. noise. We reasoned that a panel of cancer genes, although likely that some important genes were not included, may provide adequate mutation information for effective subtyping. It is an important issue to investigate since ultimately it is only practical to routinely screen a panel of genes in clinical settings.

In this study, we set out to evaluate the effectiveness of various gene panels on classifying tumors into clinically meaningful subtypes. Specifically, we collected three panels for evaluation, representing different numbers of genes: FoundationOne, PanCan and TruSeq. Briefly, FoundationOne panel includes 240 genes; PanCan was derived from 12 cancer types analyzed by TCGA and includes 127 genes [16]; TruSeq was developed by Illumina and includes 48 genes. We applied the NBS approach to 13 major cancer types with a total of ~4000 solid tumor samples profiled by exome-sequencing, using the full exome-level mutation data (termed “Full” dataset) as well as the three cancer panels.

Materials & methods

Cancer gene panels and mutation data

Three gene panels were investigated: FoundationOne (June 2014 version) (<http://foundationone.com/genelist1.php>), PanCan 127-gene panel [16] and TruSeq (http://www.illumina.com/products/truseq_amplicon_cancer_panel.ilmn). The relationship of the gene lists in the three panels is illustrated in Figure 1.

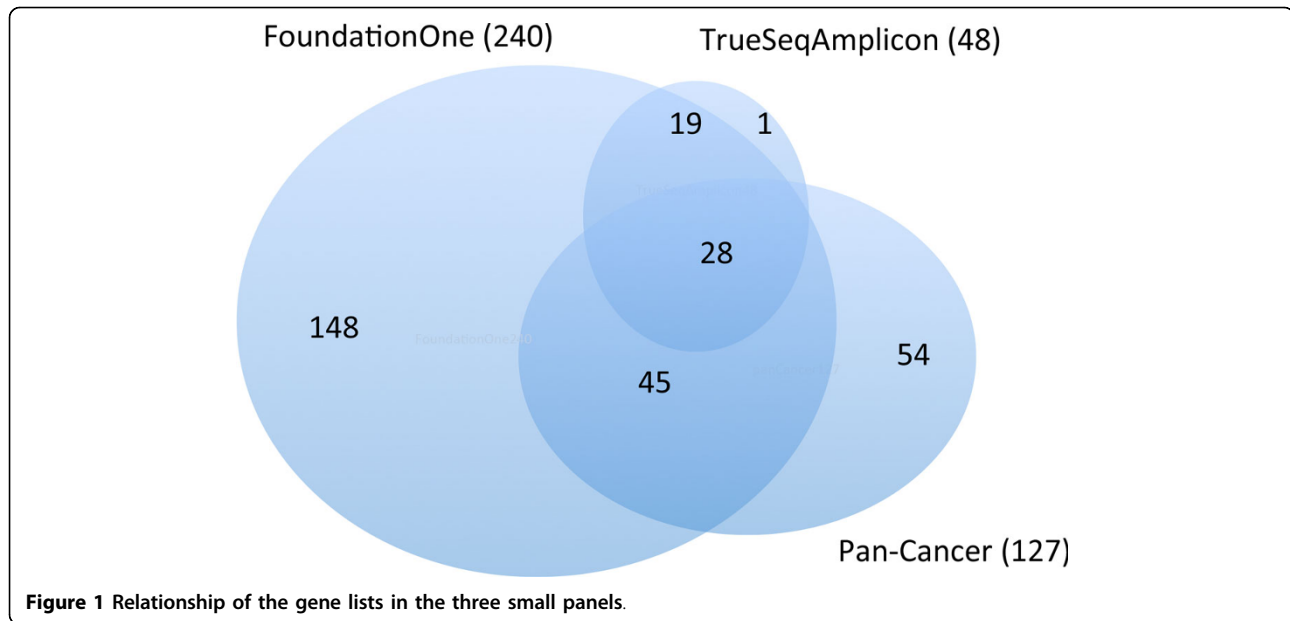
We initially collected mutation data on ~ 4700 solid tumors of 21 major cancer types that were profiled by exome-sequencing and investigated previously [10] (<http://cancergenome.broadinstitute.org>). In this study, we focus on mutations that are non-synonymous, occurred on splice sites or stop codons (termed “functional” mutations). In contrast, the non-functional mutations refer to the synonymous mutations and mutations in intronic or intergenic regions. Samples with too few mutations were treated as outliers and excluded from analysis. Specifically, samples with no mutations in gene panels or fewer than 6 functional mutations in exomes were discarded. Several cancer types, such as CARC, CLL, DLBCL and LAML that are low in mutation rates have only dozens of samples after filtering. These cancer types were excluded from further analyses. Due to the extreme molecular similarity between colon and rectal cancers [22], we combined these two into a single CRC dataset. Finally, we are left with 13 cancer types with a total of ~4000 samples (Table 1) (See Table S2 in Additional file for all 21 cancer types). For each sample we coded a gene as 1 if it has at least one functional mutation and zero otherwise. The mutation profile per patient is represented by a vector of genes marked with 1’s (mutated) or 0’s (not mutated).

Gene interaction network

HumanNet [23] v.1 (<http://www.functionalnet.org/humannet/download.html>) with the top 10 percent most confident edges were used as the gene-to-gene interaction network. Mutations were mapped to genes on this network for label propagation (see below). This gene interaction network was also transformed into a k -nearest-neighbors graph by connecting vertices (i.e. genes) v_i and v_j if v_i is among the k -nearest neighbors of v_j or if v_j is among the k -nearest neighbors of v_i . The result is a symmetric connectivity matrix, which was used to derive the graph Laplacian in network-regularized NMF (see below).

Network propagation

Let n be the number of patients, m be the number of genes, F_0 be the initial gene \times patient matrix ($m \times n$ matrix), and A be the symmetric adjacency matrix representing gene-to-gene interaction network ($m \times m$



matrix). The network propagation process is carried out by the following iterative algorithm [12]:

$$F_{t+1} = a \cdot A \cdot F_t + (1 - a) \cdot F_0 \quad (1)$$

We set $a = 0.7$ as previously described [12]. The propagation function was run iteratively until F_t converges ($|F_{t+1} - F_t| < 0.001$). Following the propagation, quantile normalization was applied to F_t to ensure each patient follows the same distribution for the smoothed mutation

profile. We use F to denote the final normalized and smoothed mutation matrix.

Network-regularized NMF

Non-negative matrix factorization (NMF) aims to decompose a matrix into two lower rank non-negative matrices whose product can well approximate the original matrix [24]. We applied network-regularized NMF to constrain NMF to respect the structure of the underlying gene

Table 1. Sample sizes of 13 tumor types before and after filtering to ensure sufficient mutations per sample and total samples

Cancer ^{a,b}	Sample size	Full (# mutations ≥ 6)	FoundationOne (# mutations ≥ 1)	PanCan (# mutations ≥ 1)	TrueSeq (# mutations ≥ 1)
BLCA	99	99 (100%)	97(98%)	95 (96%)	84 (85%)
BRCA	887	849 (96%)	661 (75%)	647 (73%)	528 (60%)
CRC	233	233 (100%)	227 (97%)	226 (97%)	219 (94%)
ESO	140	140 (100%)	133 (95%)	125 (89%)	112 (80%)
GBM	291	288 (99%)	247 (85%)	237 (81%)	199 (68%)
HNSC	384	372 (97%)	357 (93%)	347 (90%)	303 (79%)
KIRC	417	414 (99%)	328 (79%)	310 (74%)	220 (53%)
LUAD	398	391 (98%)	372 (93%)	359 (90%)	322 (81%)
LUSC	176	176 (100%)	176 (100%)	175 (99%)	158 (90%)
MEL	118	118 (100%)	117 (99%)	113 (96%)	112 (95%)
MM	204	200 (98%)	157 (77%)	146 (72%)	121(59%)
OV	316	313 (99%)	276 (87%)	281 (89%)	238 (75%)
UCEC	247	247 (100%)	245 (99%)	242 (98%)	229 (93%)
Total	4143	4008			

^aBLCA-Bladder urothelial carcinoma; BRCA-Breast invasive carcinoma; CRC-Colorectal carcinoma; ESO-Esophageal adenocarcinoma; GBM-Glioblastoma multiforme; HNSC-Head and neck squamous cell carcinoma; KIRC-Kidney renal clear cell carcinoma; LUAD-Lung adenocarcinoma; LUSC-Lung squamous cell carcinoma; MEL-Melanoma; MM-Multiple myeloma; OV-Ovarian serous cystadenocarcinoma; UCEC-Uterine corpus endometrial carcinoma

^bCancer types with survival data: BLCA, BRCA, CRC, ESO, GBM, HNSC, KIRC, LUAD, LUSC, OV,UCEC.

interaction network as previously described [25]. The objective is to minimize the following function:

$$\min_{W,H>0} \|F - WH\|_F^2 + \lambda \text{trace}(W^tLW) \quad (2)$$

where $\|\cdot\|_F$ denotes the matrix Frobenius norm, W is an m by K matrix and H is a K by n matrix, with entries in both W and H non-negative. W is a collection of basis vectors or “metagenes”, and H contains the loading of the basis. The value K controls the dimension reduction, and we used $K = 3,4,5,6$ in this study. L is the graph Laplacian of a k -nearest-neighbor network. We chose $k = 11$ as previously described [12]. λ is the regularization parameter and the value was set as 200, which is on the same scale as previously described [25]. The iterative algorithm proposed in Cai et al [25] was used to find solutions W and H . The iteration was run until the objective function converges ($|F_{t+1} - F_t| < 0.1$).

Consensus clustering

In order to achieve robust clustering, we used consensus clustering [26] to generate final clustering of patients. Specifically, we ran network-regulated NMF using a random sample without replacement of 80% patients to construct a clustering, and repeated this process 500 times. The collection of 500 clustering results was used to construct the similarity matrix, which records the frequency with which each pair of patients was observed to share the same membership among all replicates. Hierarchical clustering with average linkage was generated based on the similarity matrix using the R “NMF” package. We used cophenetic correlation coefficient (ccc) to assess the dispersion of the consensus matrix as previously described [27] also using the R “NMF” package.

Survival analysis

Survival analysis was performed using the R “survival” package. Kaplan-Meier survival curves were plotted for each NBS subtype and log-rank tests were performed to test the association of subtypes with survival. Fisher’s exact tests were used to test the association of subtypes with tumor grade or stage.

Results

Clustering patterns

We examined the clustering patterns of 13 cancer types for each of the gene sets (the three panels plus the Full dataset) by applying the NBS approach. The clustering outcomes ($K = 3, 4, 5, 6$) for each of the 13 cancer types were displayed in Supplemental Fig. S1-S13 (Additional file). Overall, we observed different clustering patterns across four sets of genes for a given cancer type and a rank K . For most cases, the smallest panel (TrueSeq) produced the clearest stratification while for the Full data the

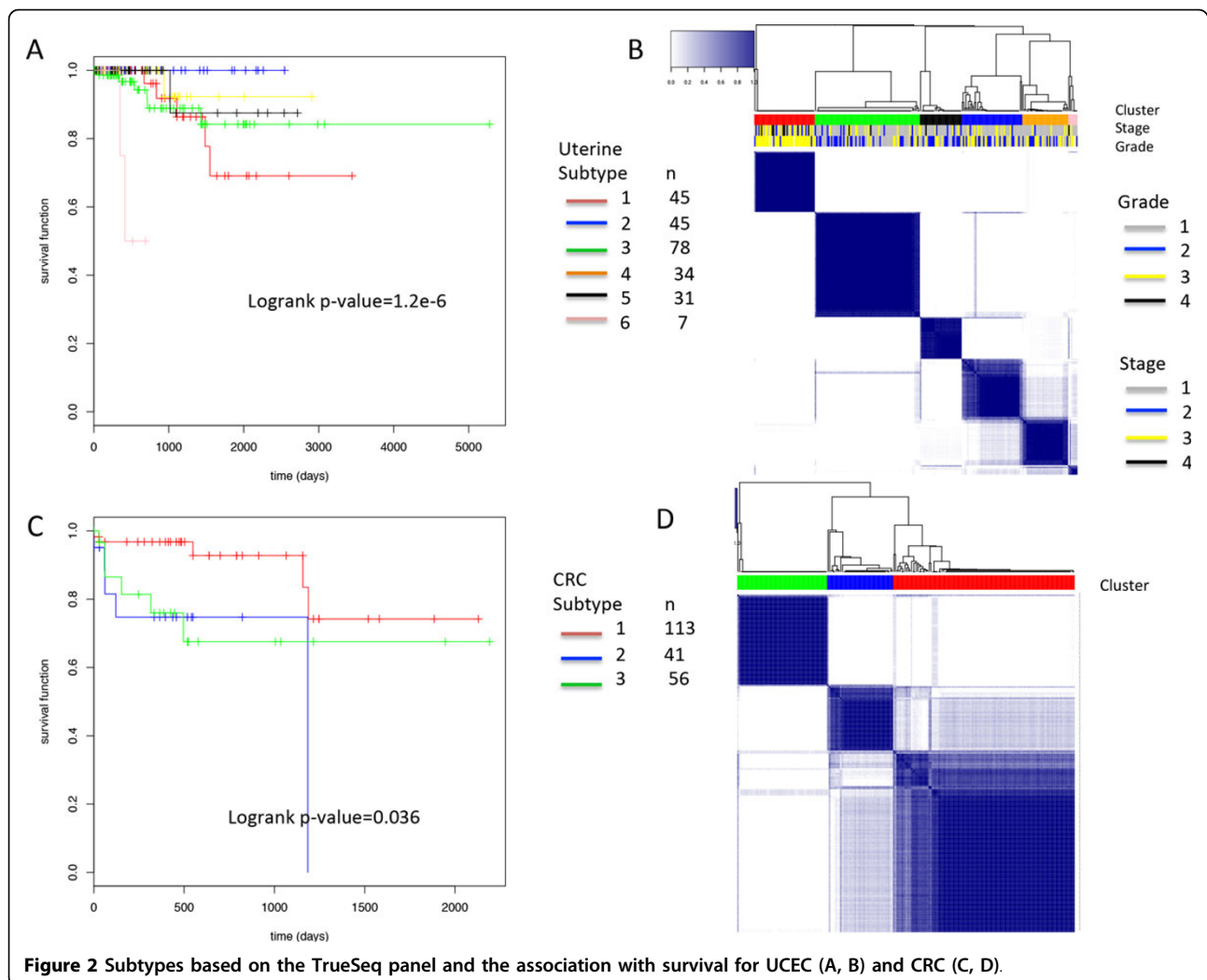
stratifications are less clear (e.g., ESO, GBM, KIRC, MM). In a few cases similar patterns were observed for different gene panels (e.g., CRC, HNSC). FoundationOne and PanCan panels often produced clusters of unbalanced sizes. Among the three panels, FoundationOne and PanCan tend to produce similar patterns (e.g., KIRC, HNSC, GBM, LUAD, MM). For a few cancers (e.g., LUSC, OV) all three panels generated similar patterns. For a fixed panel, the clustering outcome appears insensitive to K (e.g., BLCA, KIRC, HNSC, LUSC, MEL and MM). The goodness of the cluster separation was assessed using the cophenetic correlation coefficient (ccc) and for clusters that exhibit clear patterns the ccc values are over 0.99. Consistent with the visual impression, the TrueSeq panel in general gave the clearest separation (median of ccc = 0.988), followed by PanCan (median ccc = 0.982), FoundationOne (median ccc = 0.972) and the Full set (median cccc = 0.627).

NBS subtypes associated with clinical data

We first investigated the association between clusters and survival and found that not all subtypes are associated with survival. Of the 11 cancer types with survival information available, 5 (CRC, HNSC, KIRC, LUAD and UCEC) showed significant association between NBS subtypes and survival (log-rank test p-value < 0.05) (Table 2). Such significant associations were observed for certain gene panels and cluster numbers (Table 2). For example, when the UCEC samples were classified into 6 clusters using the TrueSeq panel, the clusters were well-separated, differed in survival (log-rank test p-value = 1.2e-6) and were balanced in cluster size (see Figure 2A-2B). Also with the TrueSeq panel, the CRC samples were classified into 3 clusters with significant association with survival (Figure 2C-D). Among the cancer types with survival data, only OV showed a significant association of subtypes with survival using the Full dataset (log-rank p-value 0.02) with $K = 3$. This is in contrast to the gene panels as listed in Table 2, under which significant (or marginally significant) associations between subtypes and survival were observed across multiple tumors and multiple K 's (Figure S15-S17 in Additional file).

Table 2. Significant associations between NBS subtypes (clusters) and survival in 5 tumor types using 3 gene panels

Tumor	Panel	K	p-value
CRC	TruSeq	3	0.036
HNSC	TruSeq	4	0.02
KIRC	PanCan	3	0.02
LUAD	PanCan	3	1.1*10 ⁻⁴
UCEC	TruSeq	6	1.2*10 ⁻⁶



We then investigated the association between clusters and tumor grade/stage. Of the cancer types (HNSC, KIRC, OV, UCEC) with grade or stage information available, we found no association between NBS subtypes and tumor grade/stage except for UCEC. As shown in Figure 2, the subtypes of UCEC tumors with relatively poor survival (red, pink) were also enriched for high grade ($p = 0.03$, Fisher exact test).

For some tumor types, e.g. HNSC, KIRC and CRC, different K values generated nested subtypes (Fig. S15-S17 in Additional file). For example, for HNSC the yellow cluster for $K = 3$ was further divided into yellow and black clusters for $K = 4$, and the black cluster for $K = 4$ was further divided into black and pink clusters for $K = 5$, while all other clusters remain the same (Fig. S15 in Additional file). For larger K values, NBS classified patients into finer subgroups with distinct survival curves, and as a result the significance of association of survival with subtypes is consistent across different K 's (Fig. S15 in Additional file).

Similar nested patterns were observed in KIRC and CRC (Fig. S16,S17 in Additional file).

We further examined the mutation patterns (before network smoothing) of UCEC and CRC that showed significant associations with survival using the TruSeq panel. In UCEC 6 subtypes were discovered. The cluster 2 (blue) has the best survival, and cluster 1 (red) and 6 (pink) are the worst (Figure 2A). The mutation patterns showed that cluster 2 harbors more mutations than other clusters (Figure 3A). It was observed that *POLE* is associated with hypermutations in UCEC [2] but this gene is absent in the TruSeq panel. We calculated the mutation frequency of *POLE* in these clusters and found that 32% of patients in cluster 2 have *POLE* mutations compared to 0-8% in other clusters, indicating that important mutation signatures can be captured by a panel of genes via NBS although the signature gene was not included in the panel. Cluster 1 (red) is the second worst survival cluster and a distinct signature is the high frequency of *TP53* mutations compared to

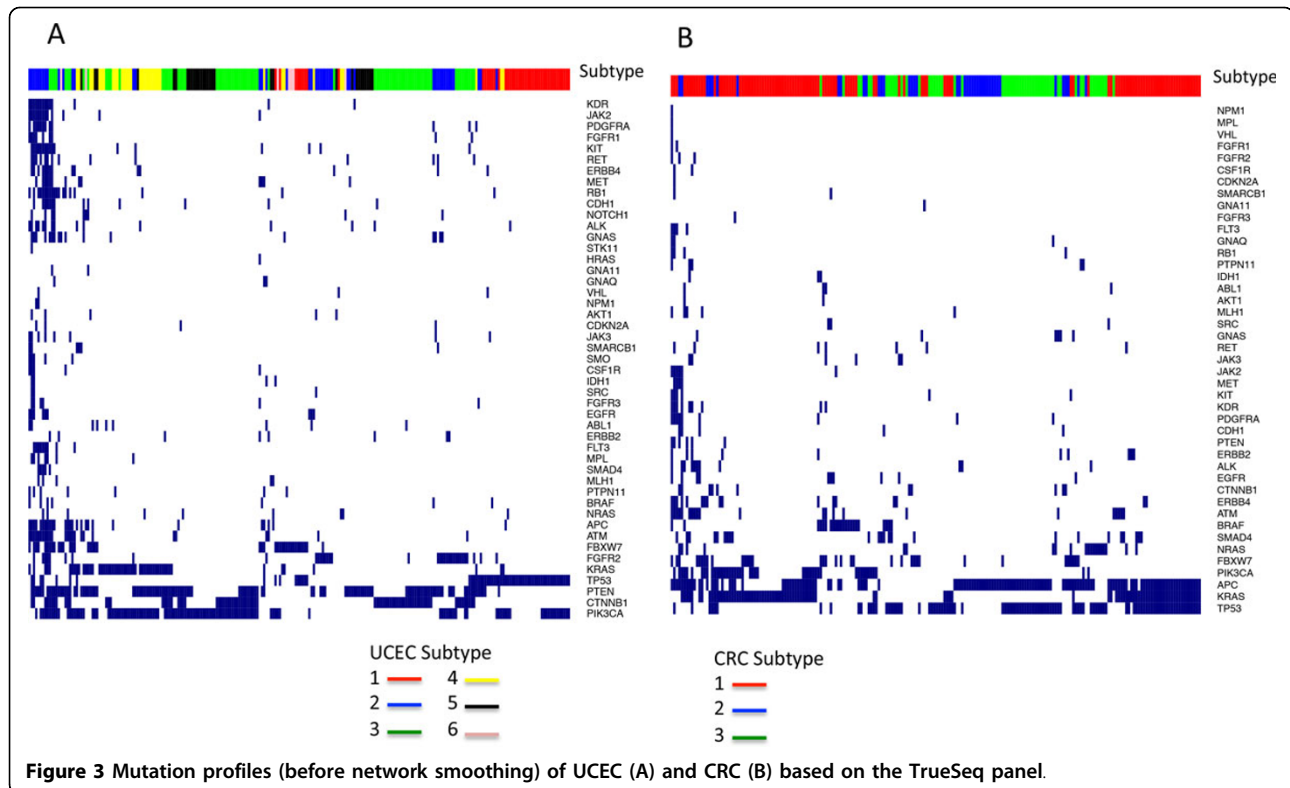


Figure 3 Mutation profiles (before network smoothing) of UCEC (A) and CRC (B) based on the TrueSeq panel.

others (Figure 3A). Only 7 patients are in the worst cluster 6 (pink) and it is not clear what signatures are associated with this subtype due to the small size of the cluster. Overall, the clustering based on this small panel is highly consistent with that obtained previously, which was based on integrative analyses of multiple platforms [2]. For CRC, there are no clear mutation patterns in the gene panel (Figure 3B), and the separation of the better survival cluster (cluster 1, red, Figure 2C,D) from others may be due to complex combination of gene mutations linked through the biological network.

To investigate the effectiveness of traditional clustering approaches on mutation data, we applied hierarchical clustering on the mutation data without network smoothing of the 5 tumors that showed significant associations with survival. All of the tumors except UCEC failed to produce clear clustering, and essentially all samples were clustering into a single group (data not shown). Although UCEC showed separable subtypes, such subtypes are not associated with survival, indicating that clustering without NBS failed to capture clinically relevant pathways that are encoded in protein interaction networks.

Our results were consistent with the previous work in which NBS approach was applied to three cancers: OV, UCEC and LUAD [12]. Using the full exome-level functional mutation data, we observed classification of UCEC

samples into 3 clusters with no significant difference in survival (log-rank test p-value = 0.63) (Fig. S14 in Additional file), similar to the result reported previously [12]. We also observed significant association between OV subtypes and survival (log-rank p = 0.02) (Fig. S14 in Additional file), although the significance levels are not the same, probably due to different samples used in the two studies (e.g. ~1/3 samples are different due to differences in downloaded data).

Discussion

Exome and whole genome sequencing have identified somatic mutations across multiple tumors; however most mutations are deemed passengers. Comprehensive analyses have pinpointed important genes and pathways that drive tumor progression. This enables the classification of patients into genetically distinct subtypes so that biologically driven prognostication and therapy selection become feasible in clinic. Network-based approaches integrate mutations and protein interaction networks to achieve knowledge-guided subtyping and have shown promises in linking subtypes with clinical data. In this study we showed that using a panel of important genes can achieve superior classification than using the full exome-level mutations, and often has better predictive performance of clinical outcome as well. Due to extensive tumor heterogeneity, however, it is infeasible to

design a single panel to fit all cancers. The performance of a gene panel will likely depend on specific tumor types. We investigated three panels with different contents and numbers of genes in an effort to span a range of scenarios. For optimal performance it may warrant a focused investigation if a custom gene panel is to be designed for a specific tumor type.

NBS approaches leverage prior knowledge of protein interaction networks to overcome the challenge of the sparsity of somatic mutations. It is unclear, however, how to construct an optimal network used in NBS. It is expected that biological networks in different organs or tissues are different, and some genes in the network may not even be expressed in certain tissues. Other complicating factors in the NBS procedure involve various tuning parameters, such as the α in the network propagation step, the γ in the graph-regularization of the NMF. Together with the influence of the genes in panels, it is unclear how to obtain robust and consistent subtyping. For example, some tumor types have clear clustering of patients but the subtypes are not associated with clinical outcomes. For such cases it is also possible that clinically relevant tumor subtypes, if present, are driven by other mechanisms, such as methylation or copy number aberration, instead of somatic mutations. It is generally unknown *a priori* what strategy is optimal in tumor classification for clinic use, and more integrative analyses across large cohorts may provide further guidance towards this goal.

The motivation behind NBS is to deal with the sparsity of mutation data. For mutations that are frequently observed in multiple patients, it may be tempting to apply clustering methods on the sparse mutation data without network smoothing. Our results showed that even in subtypes that are enriched for specific mutations, e.g. UCEC, it is crucial to leverage prior knowledge encoded in gene networks to classify patients into clinically relevant subtypes. Gene interaction networks used in the NBS approach are not merely to deal with the sparsity of mutations but to provide meaningful biological knowledge to have effective subtyping. It may be true in general that NBS is beneficial for subtyping for other somatic mutation data and effectiveness of NBS in other scenarios needs further exploration.

Conclusion

We applied NBS to 13 major cancer types using both exome-level mutation data and panels of genes for classification and showed that small panels are often more effective in clustering tumors than full exome data. In addition, subtypes discovered by panels of genes in 5 tumor types (CRC, HNSC, KIRC, LUAD and UCEC) are significantly associated with patient survival, while subtypes based on full exome mutations are less predictive

of clinical data. Screening of panels of genes combined with the NBS analysis strategy has potential for clinical use and the performance may be further improved by selection of clinically important genes and proper gene interaction networks.

Additional material

Additional file 1: This file contains all supplementary Tables and Figures. The file name is "Additional file" in PDF format.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

XZ and BL designed the study and drafted the manuscript. XZ performed all the data analyses. SZ helped the preparation of figures and tables. HY and YS contributed to the writing of the manuscript. All authors read and approved the final manuscript.

Declarations

The publication costs for this article were funded by the last author. This article has been published as part of *BMC Genomics* Volume 16 Supplement 7, 2015: Selected articles from The International Conference on Intelligent Biology and Medicine (ICIBM) 2014: Genomics. The full contents of the supplement are available online at <http://www.biomedcentral.com/bmcgenomics/supplements/16/S7>.

Authors' details

¹Department of Biostatistics, Vanderbilt University, Nashville, TN, USA. ²Center for Quantitative Sciences, Vanderbilt University, Nashville, TN, USA. ³Department of Medical Oncology, Thomas Jefferson University, Philadelphia, PA, USA. ⁴Department of Statistics, Middle Tennessee State University, Murfreesboro, TN, USA. ⁵Cancer Biology, Vanderbilt University, Nashville, TN, USA. ⁶Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN, USA.

Published: 11 June 2015

References

1. Cancer Genome Atlas Research Network: **Integrated genomic analyses of ovarian carcinoma.** *Nature* 2011, **474**(7353):609-615.
2. Cancer Genome Atlas Research Network, Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, et al: **Integrated genomic characterization of endometrial carcinoma.** *Nature* 2013, **497**(7447):67-73.
3. Konstantinopoulos PA, Spentzos D, Cannistra SA: **Gene-expression profiling in epithelial ovarian cancer.** *Nat Clin Pract Oncol* 2008, **5**(10):577-587.
4. Konstantinopoulos PA, Spentzos D, Karlan BY, Taniguchi T, Fountzilas E, et al: **Gene expression profile of BRCAness that correlates with responsiveness to chemotherapy and with outcome in patients with epithelial ovarian cancer.** *J Clin Oncol* 2010, **28**(22):3555-3561.
5. Reis-Filho JS, Pusztai L: **Gene expression profiling in breast cancer: classification, prognostication, and prediction.** *Lancet* 2011, **378**(9805):1812-1823.
6. Cancer Genome Atlas Network: **Comprehensive molecular characterization of human colon and rectal cancer.** *Nature* 2012, **487**(7407):330-337.
7. Chmielecki J, Pietanza MC, Aftab D, Shen R, Zhao Z, et al: **EGFR-mutant lung adenocarcinomas treated first-line with the novel EGFR inhibitor, XL647, can subsequently retain moderate sensitivity to erlotinib.** *J Thorac Oncol* 2012, **7**(2):434-442.
8. Olivier M, Taniere P: **Somatic mutations in cancer prognosis and prediction: lessons from TP53 and EGFR genes.** *Curr Opin Oncol* 2011, **23**(1):88-92.
9. Pirazzoli V, Nebhan C, Song X, Wurtz A, Walther Z, Cai G, et al: **Acquired resistance of EGFR-mutant lung adenocarcinomas to afatinib plus**

- cetuximab is associated with activation of mTORC1. *Cell Rep* 2014, **7**(4):999-1008.
10. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, *et al*: **Discovery and saturation analysis of cancer genes across 21 tumour types.** *Nature* 2014, **505**(7484):495-501.
 11. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko P, *et al*: **Mutational heterogeneity in cancer and the search for new cancer-associated genes.** *Nature* 2013, **499**(7457):214-218.
 12. Hofree M, Shen JP, Carter H, Gross A, Ideker T: **Network-based stratification of tumor mutations.** *Nat Methods* 2013, **10**(11):1108-1115.
 13. Chuang HY, Lee E, Liu YT, Lee D, Ideker T: **Network-based classification of breast cancer metastasis.** *Mol Syst Biol* 2007, **3**:140.
 14. Ideker T, Sharan R: **Protein networks in disease.** *Genome Res* 2008, **18**(4):644-652.
 15. Ciriello G, Cerami E, Sander C, Schultz N: **Mutual exclusivity analysis identifies oncogenic network modules.** *Genome Res* 2012, **22**(2):398-406.
 16. Kandath C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, *et al*: **Mutational landscape and significance across 12 major cancer types.** *Nature* 2013, **502**(7471):333-339.
 17. Hanahan D, Weinberg RA: **Hallmarks of cancer: the next generation.** *Cell* 2011, **144**(5):646-674.
 18. Kreeger PK, Lauffenburger DA: **Cancer systems biology: a network modeling perspective.** *Carcinogenesis* 2010, **31**(1):2-8.
 19. Balko JM, Giltinan JM, Wang K, Schwarz LJ, Young CD, Cook RS, *et al*: **Molecular profiling of the residual disease of triple-negative breast cancers after neoadjuvant chemotherapy identifies actionable therapeutic targets.** *Cancer Discov* 2014, **4**(2):232-245.
 20. Frampton GM, Fichtenholtz A, Otto GA, Wang K, Downing SR, He J, *et al*: **Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing.** *Nat Biotechnol* 2013, **31**(11):1023-1031.
 21. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW, *et al*: **Cancer genome landscapes.** *Science* 2013, **339**(6127):1546-1558.
 22. Hoadley KA, Yau C, Wolf DM, Cherniack AD, Tamborero D, Ng S, *et al*: **Multiplatform Analysis of 12 Cancer Types Reveals Molecular Classification within and across Tissues of Origin.** *Cell* 2014, **158**(4):929-944.
 23. Lee I, Blom UM, Wang PI, Shim JE, Marcotte EM: **Prioritizing candidate disease genes by network-based boosting of genome-wide association data.** *Genome Res* 2011, **21**(7):1109-1121.
 24. Lee DD, Seung HS: **Learning the parts of objects by non-negative matrix factorization.** *Nature* 1999, **401**(6755):788-791.
 25. Cai DH X, Wu X, Han J: **Non-negative Matrix Factorization on Manifold.** *IEEE* 2008, 63-72.
 26. Stefano Monti PT, Mill Mesirov, Todd Golub: **Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data.** *Machine Learning* 2003, 52:91-118.
 27. Brunet JP, Tamayo P, Golub TR, Mesirov JP: **Metagenes and molecular pattern discovery using matrix factorization.** *Proc Natl Acad Sci USA* 2004, **101**(12):4164-4169.

doi:10.1186/1471-2164-16-S7-S7

Cite this article as: Zhong *et al*: Network-based stratification analysis of 13 major cancer types using mutations in panels of cancer genes. *BMC Genomics* 2015 **16**(Suppl 7):S7.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

