Germline genomic landscapes of breast cancer patients significantly predict clinical outcomes

Jean-Sébastien Milanese[1,2], Chabane Tibiche[1], Naif Zaman[1], Jinfeng Zou[1], Pengyong Han[3], Zhigang Meng[3,5], Andre Nantel[1], Arnaud Droit[2], and Edwin Wang[1,3,4,6*]

1. National Research Council Canada, 6100 Royalmount Ave, Montreal, Canada, H4P 2R2
2. Genomics Center, Centre Hospitalier Universitaire de Québec - Université Laval Research Center, Quebec, Canada, G1V 4G2
3. Department of Biochemistry & Molecular Biology, Medical Genetics, and Oncology, University of Calgary, 3330 Hospital Dr. NW, Calgary, Canada, T2N 4N1
4. Alberta Children's Hospital Research Institute and Arnie Charbonneau Cancer Research Institute, University of Calgary, 3330 Hospital Dr. NW, Calgary, Canada, T2N 4N1
5. Chinese Academy of Agricultural Science, No. 12 Zhongguangcun South St., Haidian District, Beijing, 100086, China
6. Department of Medicine, McGill University, 3605 Mountain St, Montreal, Canada H3G 2M1

*Corresponding to: E.W. (edwin.wang@ucalgary.ca)
Department of Biochemistry & Molecular Biology, Medical Genetics, and Oncology, University of Calgary, 3330 Hospital Dr. NW, Calgary, Canada, T2N 4N1
Phone: 403-220-6609
Fax: 403-9891-9879

Abstract

Germline genetic variants such as BRCA1/2 could play an important role in tumorigenesis and even clinical outcomes of cancer patients. However, only a small fraction (i.e., 10%) of cancer patients whose outcomes have been shown to be associated with germline mutations of several genes (e.g., BRCA1/2, APC, TP53, PTEN and so on). We asked if genes' mutilations in germline genomes are associated with the clinical outcome of the majority cancer patient population. To answer this question, we applied our recently developed algorithm which enables to construct predictive models using genome sequencing data to ER+ breast cancer patients (n=755). We showed that functional gene mutations in germline genomes significantly distinguished recurred and non-recurred ER+ breast cancer patients in two independent cohorts (n=200 and 295, $P=1.0x10^{-7}$). These results suggest that pre-existing genetic variants in germlines could determine breast tumorigenesis and tumor evolution. Further, germline genomic information could be used for developing non-invasive tests for patients' outcomes and drug response in breast cancer, and even other cancer types and complex diseases.

Introduction

Cancer is a process of asexual evolution driven by genomic alterations. A single normal cell randomly acquires a series of mutations that allows it to proliferate and to be transformed into a cancer founding clone thus initiating tumor progression and recurrence. In general, cancer recurrence and metastasis are the result of the interactions of multiple mutated genes. We hypothesized that mutagenic processes are essentially blind or non-purposeful, however, to drive cancer metastasis, new mutations will be selected if they could integrate into the pre-existing genomic landscape (i.e., germline mutations or germline genetic variants) to trigger or activate a cancer metastasis network. This means that pre-existing germline genetic variants could provide a profound constraint on the evolution of founding clones and subclones, and therefore, have a contingent effect on tumor evolution and patient outcomes. Family history remains one of the major risk factors that contribute to cancer and recent studies have identified several genes whose germline mutations are associated with cancer. For example, patients suffering from Li-Fraumeni syndrome have an almost 100% chance of developing a wide range of malignancies before the age of 70. Most of them carry a missing or damaged *p53* gene, a tumor suppressor whose activity is impaired in almost 50% of all cancers. Other cancer-predisposition genes include *BRCA1* and *BRCA2*[1,2], which are associated with breast and ovarian cancer, *PTEN*, whose mutation results in Cowden syndrome, *APC*, which is linked to familial adenomatous polyposis and the Retinoblastoma gene *RB1*. Two distinct types of multiple endocrine neoplasias are associated with the *RET* and *MEN1* genes while *VHL* alterations result in kidney and other types of cancer. Finally, Lynch syndrome, a form of colorectal cancer, is linked to *MSH2*, *MLH1*, *MSH6*, *PMS2*, and *EPCAM*. Genetic tests based on these highly-penetrant gene mutations have shown their usefulness, but they can explain only a small fraction (5-10%) of the inherited cancers. Most neoplasms arise and are modulated by the interactions of multiple gene networks and there is a great diversity of genetic alterations even within tumors of the same subtypes.

Thus far, it has been unknown that to what extent the germline genomes affect tumorigenesis, evolution and even patients' clinical outcome. We have previously shown that founding clone mutations are able to predict tumor recurrence[3]. Here, we reasoned that the collective impact of germline genetic variants/mutations in cancer patients might largely determine tumorigenesis, evolution and even patients' clinical outcomes. As tumor heterogeneity becomes a more prominent concept, it is our belief that germline genetic mutations act in combination with the tumor somatic mutations to modulate tumorigenesis and metastasis. Each patient's germline genetic mutations' combination predisposes specific biological/signaling pathways (even phenotypes) that would lead to diverse clinical outcomes of cancer patients. Therefore, the germline genomic landscape of individuals could be used as a predictive tool in order to inform clinicians as to when and how the disease might progress. Thus, germline genetic mutations could offer a new non-invasive genetic testing approach because they can be determined using a blood or saliva sample. The increasing availability of genome sequencing data is creating opportunities to develop predictive models that can interpret and translate these complex genomic alterations into clinical outcomes.

To our knowledge, so far predictions using germline mutations have not been demonstrated. In this study, we showed that collective germline genetic mutations of breast cancer patients enable to predict tumor recurrence by applying a recently developed method, eTumorMetastasis[3], to 755 breast cancer patients, respectively. These results highlight the important role of germline genetic mutations in impacting on tumor evolution and

recurrences.

Results

To exam if germline genetic mutations are able to predict tumor recurrence, we used the whole-exome sequencing data of the healthy tissues from 755 ER+ breast patients by applying a recently developed method, eTumorMetastasis[3]. Breast cancer has three subtypes, ER+ subtype represents ~70% of the breast cancer patients, thus, in this study, we used ER+ subtype patients for the breast cancer. To apply the eTumorMetastasis method, for each patient, we annotated the functional mutations from the germline whole-exome sequencing data (see Methods), and the used the functionally mutated genes in the eTumorMetastasis method. A complete flowchart of eTumorMetastasis is available in the Supplementary Materials.

We used 200 ER+ breast cancer samples to identify gene signatures (i.e., because eTumorMetastasis identifies network-based gene signatures, Network Operational Signatures (NOG), we called gene signatures as NOG gene signatures here). By applying the eTumorMetastasis to the germline genomes of the 200 ER+ breast cancer patients, we identified 18 NOG gene signatures for ER+ breast cancer (see Supplementary Methods). Each of the NOG gene signature repents a cancer hallmark such as apoptosis, cell proliferation, metastasis, and so on (see Supplementary Methods). We previously showed that multiple gene signatures representing distinct cancer hallmarks could be identified from one dataset[4], furthermore, ensemble-based prediction by constructing gene signatures representing distinct cancer hallmarks enabled to significantly improve prediction performance[5]. Thus, we used the 18 NOG gene signatures to construct NOG_CSS set (i.e., NOG-based Combinatory Signature Set) using the method we developed[5].

Based on the germline genomes, the NOG_CSS sets of breast cancer significantly distinguished recurred and non-recurred tumors in the two validation ER+ breast cancer sets: 200 (Nature-Set, $P=3.4x10^{-2}$) and 295 (TCGA-CPTAC independent set, $P=1.0x10^{-7}$). The prediction results have been summarized in Table 1 and Fig 2. These results suggested that germline genetic mutations are significantly correlated with tumor metastasis and supported our hypothesis that the original germline genomic landscape of a cancer patient has a significantly impact on clinical outcome. Practically, genome sequencing a patient's blood or saliva sample could provide a convenient, timely and noninvasive way to predict patient outcome in a clinical setting.

To compare the prediction performance of the NOG_CSS sets with that of clinical factors, we conducted relapse-free survival analysis of clinical factors using the Cox proportional hazards regression model (Supplementary Table 7). The best p-value (i.e., $P=1.0x10^{-2}$ scale rank) using covariate models in Supplementary Table 7 is far less than the CSS sets-derived whole-exome sequencing data ($P=1.0x10^{-7}$ scale rank).

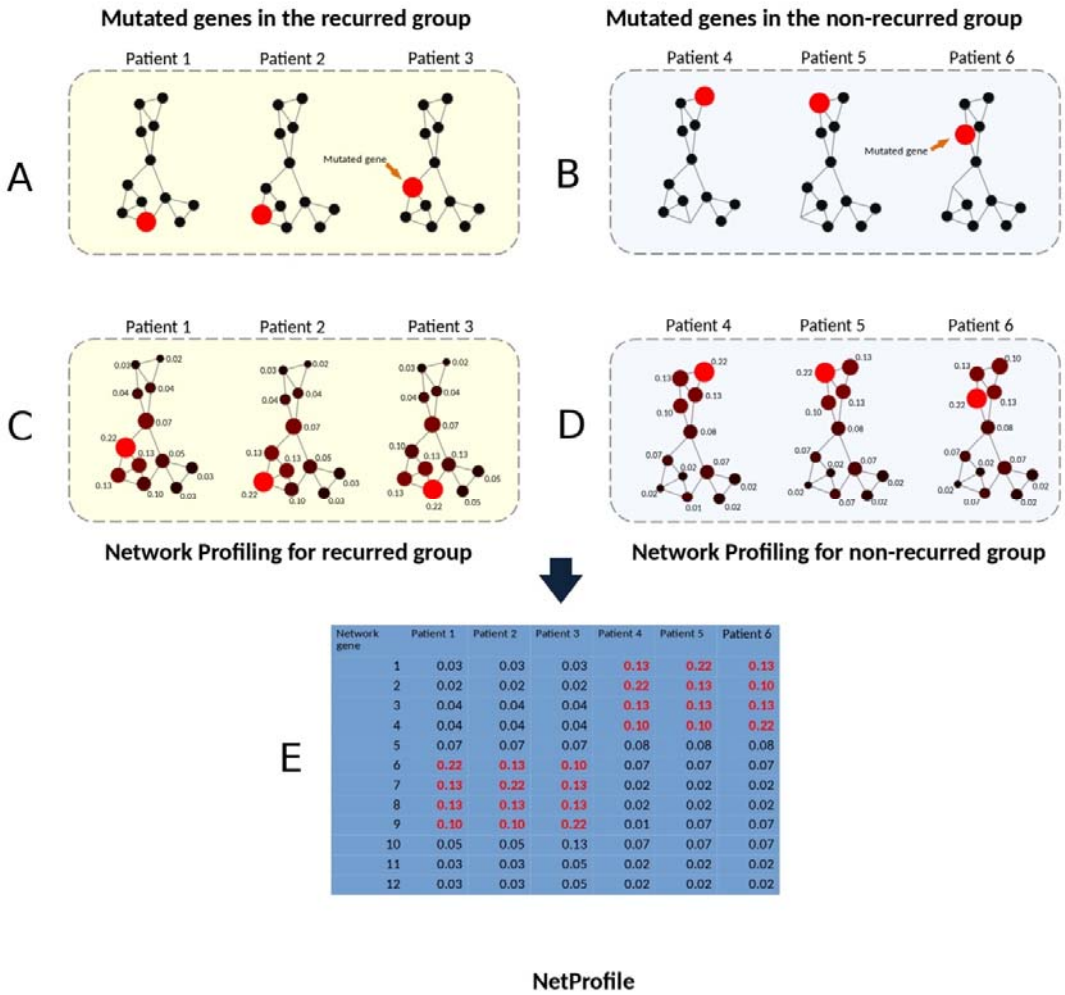
To understand why germline genomic landscapes of cancer patients are predictive for tumor recurrence, we ran a pathway enrichment analysis using PANTHER[6,7] on the NOG gene signatures derived from the germline genomes of the cancer patients. Interestingly, the top 4 scored pathways were: integrin signaling pathway, inflammation mediated by chemokine and cytokine signaling pathway, CCKR signaling map and Gonadotropin-releasing hormone receptor pathway (Supplementary Table 8), suggesting that these pathways are significantly differentiated between the recurred and non-recurred breast cancer patients' germline

genomes. Receptors like integrins have been known to play a crucial role in cell adhesion and metastasis formation. Integrins mediate a variety of stimulus from the extracellular matrix and therefore, from the microenvironment, eventually leading to multiple cellular response[8,9]. While tumor inflammation is a cancer hallmark, having chemokines/cytokines pathway enriched in breast cancer correlates with the metastatic impact played by the interactions between cancer cells and their microenvironment. Many transcriptional factors (e.g., NF-Kb, TGF-b) can stimulate inflammation and thus directly affect tumor progression[10-12]. The role of the CCKR signaling pathway in cancer is also well established. CCKs interacts with many genes (e.g., Akt, EGFR, MAPK, Bcl) for a variety of cellular response (apoptosis, proliferation, migration). Finally, the gonadotropin-releasing hormone (GnRH) receptor has also been reported to promote cell proliferation and metastasis formation (breast and ovarian cancer)[13,14]. GnRH receptor inhibits growth factors receptors (GPER) which in turn, reduces cell proliferation and cell invasion. These results suggest that the germlines of recurred patients have pre-existing biological programs which are promoting inflammation, cell adhesion and metastasis in either cancer cells or microenvironments.

Table 1 Prediction accuracy and recall rate for validation sets for breast cancer using the NOG_CSS sets derived from germline mutations

Dataset	Number of samples	Low-Risk		High-risk	
		Accuracy (%)*	Recall (%)†	Accuracy (%)**	Recall (%)††
TCGA-Nature	200	91.4	53.3	12.7	40.0
TCGA-CPTAC	295	97.4	71.3	35.4	67.7

Notes:
*Percentage of non-recurred (i.e., non-metastatic) samples in the predicted low-risk group.
†Percentage of the predicted low-risk samples from the non-recurred group.
**Percentage of recurred (i.e., metastatic) samples in the predicted high-risk group.
††Percentage of the predicted high-risk samples from the recurred group.



Network gene	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6
1	0.03	0.03	0.03	0.13	0.22	0.13
2	0.02	0.02	0.02	0.22	0.13	0.10
3	0.04	0.04	0.04	0.13	0.13	0.13
4	0.04	0.04	0.04	0.10	0.10	0.22
5	0.07	0.07	0.07	0.08	0.08	0.08
6	0.22	0.13	0.10	0.07	0.07	0.07
7	0.13	0.22	0.13	0.02	0.02	0.02
8	0.13	0.13	0.13	0.02	0.02	0.02
9	0.10	0.10	0.22	0.01	0.07	0.07
10	0.05	0.05	0.13	0.07	0.07	0.07
11	0.03	0.03	0.05	0.02	0.02	0.02
12	0.03	0.03	0.05	0.02	0.02	0.02

Figure 1 Network propagation and netProfiles for recurred and non-recurred samples. Functionally mutated genes in three recurred samples (**A**) are different but for a network cluster in the human signaling network, and so are they in three non-recurred samples (**B**). For each recurred sample, by conducting network propagation based on its mutated genes, a network cluster (nodes in red and garnet, **C**) which is similar to the cluster in A is emerged. A similar pattern (**D**) is overserved for the non-recurred samples. The network clusters (**C,** they are similar between the recurred samples) and **(D,** they are similar between the non-recurred samples) make it possible to classify recurred and non-recurred samples, respectively. In the network, nodes and lines represent genes and gene relations, respectively. Numeral numbers of each network node represent heating scores. Red nodes represent mutated genes while garnet nodes represent gaining 'energy' and node sizes represent 'heating scores' values - the amount of the gained energy. A combination of those heating scores for each gene and each patient is represented as a netProfile (**E**).

Fig 2 Kaplan–Meier curves of the risk groups for breast cancer patients with 10-year disease-free survival predicted by the NOG_CSS sets. NOG_CSS sets derived from germline mutations in (**A**) the training set, (**B**) the validation set, TCGA-Nature and (**C**) the validation set, TCGA-CPTAC. Kaplan–Meier curves have been not made for samples which do not have follow-up time. Blue and red curves represent low- and high-risk groups, respectively. P-values were obtained from the χ^2-test.

Fig 2A

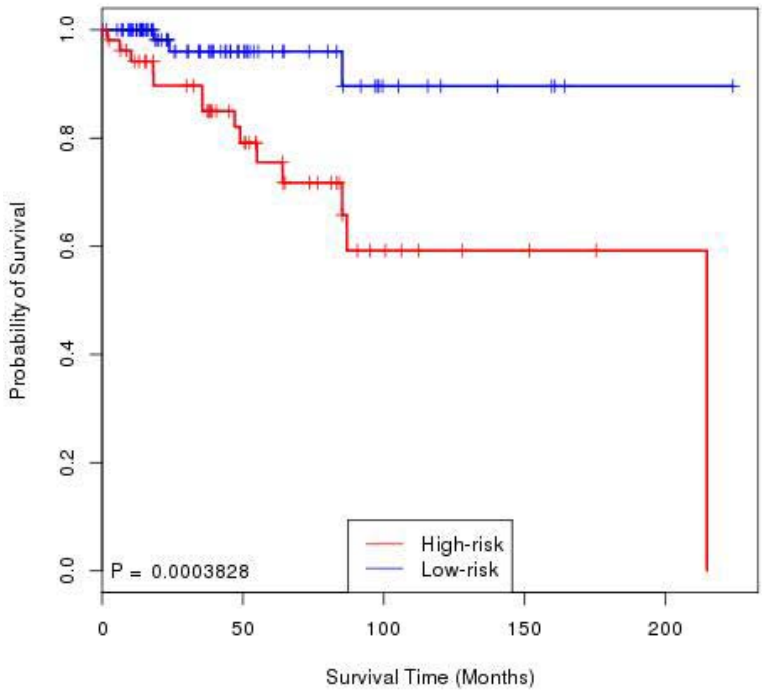


Fig 2B

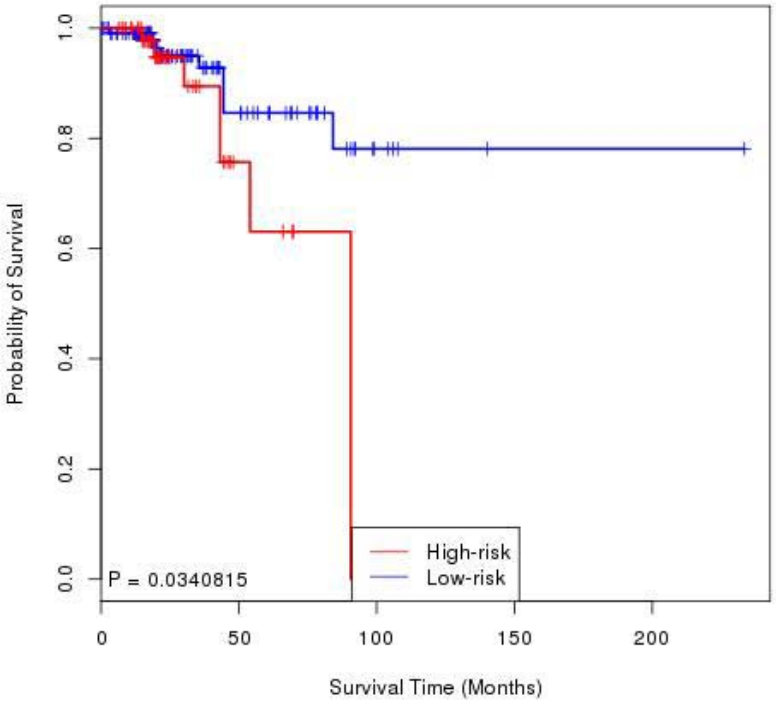
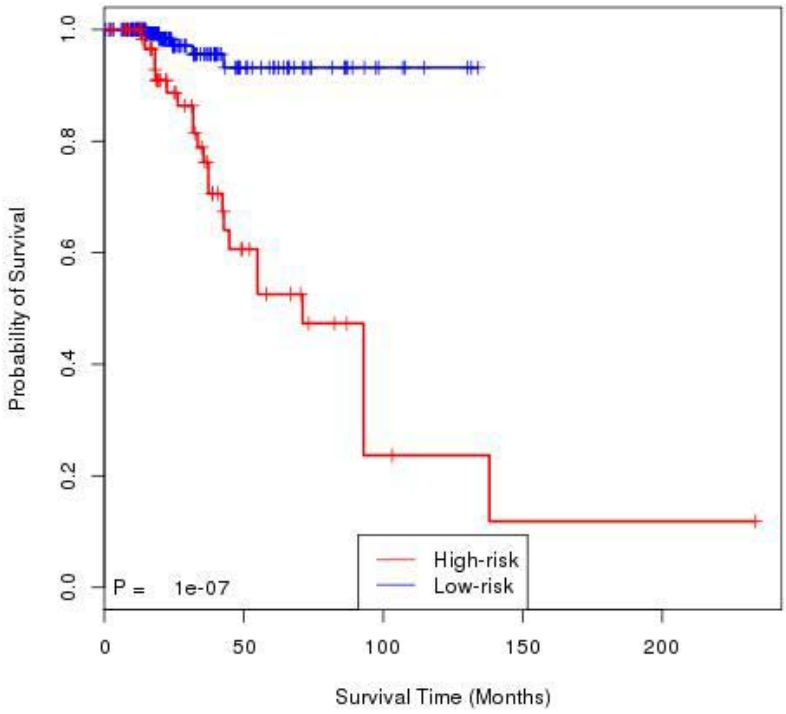

Probability of Survival
1.0
0.8
0.6
0.4
0.2
0.0
P = 0.0340815
High-risk
Low-risk
0
50
100
150
200
Survival Time (Months)

Fig 2C



Discussion

This is the first study that shows that the germline genomes of cancer patients predict clinical outcomes. These results suggest that the germline mutations influence tumor evolution and recurrence. We hypothesized that the tumor microenvironment, encoded by patients' genetic makeup, could influence tumor recurrence. The enrichment analysis of the NOG_gene signatures derived from cancer patients' germlines suggested that the germline genomes of the recurred patients have activated signaling pathways which are associated with inflammation and cell adhesion. These pathways in tumor microenvironment will promote tumor recurrence.

Thus far, our understanding of the fundamental biology of the tumor-host interaction has been limited. However, several reports provide evidence that germline mutations of cancer patients modulate metastasis through the tumor microenvironment. For example, it is known that germline-encoded receptor variants could trigger innate immune response in cancer patients[15]. The presence or absence of the T cell-inflamed tumor microenvironment phenotype in individual patients is also the result from interactions between somatic differences at the level of the tumor cells and germline polymorphisms at the level of the host. Patients with T cell-inflamed tumors often have a good prognosis[16]. Mutations in the genome of cancer patients could mediate the impact of the immune system on the tumor microenvironment. For example, lung cancer patients with a germline mutation in the Nrf2 gene have a good prognosis due to the fact that Nrf2 exerts anti-metastatic activity by regulating the inflammatory status and redox balance of the hematopoietic and immune

systems of cancer patients[17]. In prostate cancer, patients with a germline variant of the ASPN D locus are highly associated with poorer oncologic outcomes[18], probably through its interactions with TGF-b. These studies highlight the impacts of a cancer patients' genetic makeup on metastasis and more correlations will certainly come to light once future studies take gene sets in consideration instead of individual genes. Taken together, the germline genomic landscape of a cancer patient might provide not only a substantial constraint on tumor recurrence for tumor cells, but also constrains for the tumor microenvironment, which ultimately has a profound impact on patient outcomes. These insights provide rationales for further studies on the impact of germline genomic landscapes on clinical outcomes of carcinogenesis. This study further suggests that the risk of metastasis is predictable based on the germline genomic landscapes of cancer patients. These results prompted us to propose that the future study of tumor metastasis may shift to the early evolution of tumors and germline genomics. Traditionally, germline mutations have been largely ignored in cancer genomics studies; for example, the TCGA consortium has often focused on somatic cancer driving mutations, and germline mutations have been filtered out before formal analysis of tumor genome sequencing data.

In this work, the demonstration that germline exome sequencing data can predict a cancer patients' outcomes suggests that blood or saliva samples could provide a noninvasive method to determine cancer prognosis and to guide clinicians in making treatment decisions, such as the initiation of more invasive tests. In addition, germline-based tests could overcome the potential bias derived from intratumour heterogeneity. Recent genome sequencing studies of lung, renal, brain, prostate and ovarian tumor samples have indicated that intratumour heterogeneity is common in many tumors[19-24]. Furthermore, in some tumor samples, cancer subclones are organized regionally and locally so that different regions of a tumor contain distinct subclonal cells[24]. Thus, samples of different regions of a tumor may contain different genetic/epigenetic information (i.e., profiles of mutations, gene expression, copy numbers and methylation). Therefore, sampling of tumor tissues could introduce biased genetic/epigenetic information to the traditional molecular or genetic tests thereby generating incorrect testing results that misguide clinical decisions. The complexity of intratumour heterogeneity leads to huge clinical challenges in cancer diagnosis, prognosis and treatment[22]. Genome-wide germline genetic variants can be easily identified by genome/whole-exome sequencing of liquid biopsies such as blood or saliva samples. Prognostic prediction using a patient's germline genomic landscape opens up the possibility of assessing cancer patients' risk of recurrence in a non-invasive manner, which allows for the forecasting of cancer recurrences in a quick, convenient and minimally invasive manner. We showed that sequencing of patient germline exomes might provide an efficient, non-painful and convenient way for predicting tumor recurrence.

Methods

To exam if the germline genetic mutations predict tumor recurrence, we obtained whole-exome sequencing data of the germlines for breast cancer from TCGA: ER+ breast cancer, a training set of 200 samples and 2 validation sets of 200 and 295 samples (TCGA-Nature and TCGA-CPTAC, respectively). The demographic table of this dataset can be found in the supplemental (see Supplementary Table 1). Raw sequence reads from healthy samples of the ER+ breast cancer patients were processed using the methods described previously[3]. Variant calling using normal and tumor samples was then performed by Varscan2[25] (see Supplementary Materials, Supplementary Figure 1A).

To define germline mutations, we used scatter plots to visualize variant allele frequencies (VAFs) between the tumor and healthy samples. An example of allele frequencies in one sample is shown in Supplementary Figure 2. As expected, all VAFs exhibit a large cluster of sequence variants with the VAF of 45-65%. These represent heterozygous mutations whose frequency variations are affected by technical noise, sample quality and tumor heterogeneity. At the top-right corner are homozygous germline mutations with allele frequencies above 90%. What is immediately apparent is the paucity of tumor-specific homozygous and heterozygous somatic mutations that were initially absent from the healthy samples. In fact, the great majority of tumor-acquired mutations are homozygous variants that were initially present as heterozygous germline mutations. Finally, we observed a cluster of mutations with 45-65% VAFs in the healthy samples of cancer patients but a reduced frequency in the paired tumor samples. A detailed examination of these variants has shown that most lie in multigene families or segments that have undergone copy-number variations or loss of heterozygosity. This information provided the basis of choosing VAF cutoffs for germline genetic variants. Only germline functional mutations were retained for downstream analysis (Supplementary Figure 1).

To identify NOG_gene signatures using the functional mutations of the cancer patients' germline genomes, we followed the eTumorMetastasis[3] method. We used breast cancer as an example here. Briefly, we constructed a breast-specific metastasis network based on the methods described previously[26]. For each patient, we used its germline functional mutations as seeds on the breast-specific metastasis network to perform network propagation and then identify NOG_gene signatures (Fig 1)(Supplementary Figure 1 and 3).

References

1 Maistro, S. *et al*. Germline mutations in *BRCA1* and *BRCA2* in epithelial ovarian cancer patients in Brazil. *BMC Cancer* **3**, 16(1), doi: 10.1186/s12885-016-2966-x (2016).

2 Chan, SH. *et al*. Germline Mutations in Cancer Predisposition Genes are Frequent in Sporadic Sarcomas. *Sci Rep*. **7**(1), 10660, doi: 10.1038/s41598-017-10333-x (2017).

3 Milanese, J.S. *et al*. eTumorMetastasis, a network-based algorithm predicts clinical outcomes using whole-exome sequencing data of cancer patients. *bioRxiv* **268680**; doi: https://doi.org/10.1101/268680 (2018).

4 Li, J. et al. Identification of high-quality cancer prognostic markers and metastasis network modules. *Nat Commun* **1**, 34, doi:10.1038/ncomms1033 (2010).

5 Gao, S. *et al.* Identification and Construction of Combinatory Cancer Hallmark-Based Gene Signature Sets to Predict Recurrence and Chemotherapy Benefit in Stage II Colorectal Cancer. *JAMA Oncol* **2**, 37-45, doi:10.1001/jamaoncol.2015.3413 (2016).

6 Mi, H. et al. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res* **41**, 377-86, doi: 10.1093/nar/gks1118 (2013).

7 Thomas, D. *et al*. PANTHER: A Library of Protein Families and Subfamilies Indexed by Function. *Genome Res* **13**, 2129-2141, doi: 10.1101/gr.772403 (2003).

8 Ganguly, KK. *et al.* Integrins and metastasis. *Cell Adh Migr* **7(3)**, 251-61, doi: 10.4161/cam.23840 (2013).

9 Seguin, L. *et al*. Integrins and cancer: regulators of cancer stemness, metastasis, and drug resistance. *Trends Cell Biol* **25(4)**, 234–240, doi: 10.1016/j.tcb.2014.12.006 (2015).

10 Shrihari, TG. Dual role of inflammatory mediators in cancer. *Ecancermedicalscience* **11**, 721, doi: 10.3332/ecancer.2017.721 (2017).

11 Glauben, L. *et al*. Chronic Inflammation and Cytokines in the Tumor Microenvironment. *Journal of Immunology Research*, doi:10.1155/2014/149185 (2014).

12 Ben-Baruch, A. Host microenvironment in breast cancer development: Inflammatory cells, cytokines and chemokines in breast cancer progression: reciprocal tumor–microenvironment interactions. *Breast Cancer Res*. **5(1)**, 31-36, doi: 10.1186/bcr554 (2003).

13 Gründker, C. *et al*. Role of gonadotropin-releasing hormone (GnRH) in ovarian cancer. *Reprod Biol Endocrino*l **1**, 65. doi: 10.1186/1477-7827-1-65 (2003).

14 Gründker, C. *et al*. The Role of Gonadotropin-Releasing Hormone in Cancer Cell Proliferation and Metastasis. *Front Endocrinol* **8**, 187, doi: 10.3389/fendo.2017.00187 (2017).

15 Marcus, A. *et al.* Recognition of Tumors by the Innate Immune System and Natural Killer Cells. *Adv Immunol* **122**, 91-128, doi:10.1016/B978-0-12-800267-4.00003-1 (2014).

16 Gentles, A. J. *et al.* The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat Med* **21**, 938-945, doi:10.1038/nm.3909 (2015).

17 Satoh, H. *et al.* Nrf2-deficiency creates a responsive microenvironment for metastasis to the lung. *Carcinogenesis* **31**, 1833-1843, doi:10.1093/carcin/bgq105 (2010).

18 Hurley, P. J. *et al.* Germline Variants in Asporin Vary by Race, Modulate the Tumor Microenvironment, and Are Differentially Associated with Metastatic Prostate Cancer. *Clin Cancer Res* **22**, 448-458, doi:10.1158/1078-0432.CCR-15-0256 (2016).

19 Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* **366**, 883-892, doi:10.1056/NEJMoa1113205 (2012).
20 Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science (New York, N.Y.)* **344**, 1396-1401, doi:10.1126/science.1254257 (2014).
21 Gerlinger, M. *et al.* Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat Genet* **46**, 225-233, doi:10.1038/ng.2891 (2014).
22 Bedard, P. L., Hansen, A. R., Ratain, M. J. & Siu, L. L. Tumour heterogeneity in the clinic. *Nature* **501**, 355-364, doi:10.1038/nature12627 (2013).
23 de Bruin, E. C. *et al.* Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science (New York, N.Y.)* **346**, 251-256, doi:10.1126/science.1253462 (2014).
24 McPherson, A. *et al.* Divergent modes of clonal spread and intraperitoneal mixing in high-grade serous ovarian cancer. *Nat Genet* **48**, 758-767, doi:10.1038/ng.3573 (2016).
25 Koboldt, D. C. *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* **22**, 568-576, doi:10.1101/gr.129684.111 (2012).
26 Li, L. *et al.* The human phosphotyrosine signaling network: evolution and hotspots of hijacking in cancer. *Genome Res* **22**, 1222-1230, doi:10.1101/gr.128819.111 (2012).