CellPress

Integrating germline and somatic data towards a personalized cancer medicine

Miguel Angel Pujana

Breast Cancer and Systems Biology Unit, Translational Research Laboratory, Catalan Institute of Oncology (ICO), Bellvitge Institute for Biomedical Research (IDIBELL), L'Hospitalet del Llobregat, Barcelona 08908, Catalonia, Spain

In recent years there have been major advances in the knowledge of the genetic alterations that drive cancer susceptibility and progression. However, the molecular links between these alteration levels (i.e., germline for susceptibility and somatic for progression) remain mostly unknown. Here, the potential of integrating germline and somatic data for a comprehensive and personalized cancer medicine is discussed.

Genetic predisposition to cancer

The observation that the relatives of cancer patients are at higher risk of suffering the same disease than unrelated individuals from the general population was among the earliest evidence of genetic predisposition to cancer. In this context, a germline genetic mutation in a given gene may confer high or moderate risk (typically over 50% and 20–50% lifetime risks, respectively) for a particular cancer type; hereafter, 'germline mutation' will refer to disease causing genetic changes, whereas 'variation' will refer to changes or differences that *a priori* are not the precise cause of the disease. However, because of their nature, these mutations are relatively infrequent in the general population and make a limited contribution to the overall disease burden.

Following on from familial studies, the 'common disease–common variant' hypothesis led to the design of genome-wide association studies (GWAS) to identify common genetic variation associated with relatively low risk of cancer. Completion of these GWAS has substantially increased our knowledge of the genetic basis of cancer susceptibility in the general population, identifying hundreds of variants that typically increase lifetime risk by 5–20% [1]. Although some of these variants may in fact be surrogates of high/moderate-penetrance germline mutations, recent meta-analyses have suggested the involvement of thousands of such variants with very low effects [2]. This genetic information, combined with data on family history of the disease and lifestyle factors, may be used to stratify the population in terms of risk of suffering a given cancer type. However, could the acquired knowledge bring fundamental insights with broader applications than risk stratification? Because over 70 low-penetrance breast cancer susceptibility loci have

been identified to date, subsequent discussion is mainly focused on breast cancer.

Somatic alterations and cancer subtypes

Tumors arising in carriers of high-penetrance germline mutations may present specific molecular and histopathological features; for example, carriers of mutations in *breast cancer early onset 1* (*BRCA1*) commonly develop tumors with 'triple-negative' receptor status and a 'basal-like' phenotype, which are generally of poorer prognosis and have no standardized targeted therapies. Thus, for these tumor types, particularly in high-risk settings, preventive clinical interventions may be warranted. Notably, similar molecular and histopathological connections are emerging for low-penetrance germline variants. For example, specific genetic associations have been identified for triple-negative breast cancer, whereas most of the initial GWAS results corresponded to cases with estrogen receptor α (ERα)-positive tumors [3].

Intriguingly, there is population-based evidence for the heritability of prognosis and, in mice, tumor progression differs according to germline background [4]. In addition, germline genetic variants in mice and humans have been associated with metastasis susceptibility [5]. Concerning low-penetrance loci derived from GWAS, there are indications of concurrent associations with risk and prognosis or survival [6]. However, the functional relationship(s) between low-penetrance germline mutations and somatic drivers, and with clinical variables such as cancer progression and patient survival, remain mostly undetermined. Critically, most GWAS were designed to study cancer risk and there are not enough data and/or data of sufficient quality and completeness to systematically evaluate associations with somatic/clinical variables.

Integrative studies and molecular dependences

The major molecular cancer subtypes (at the somatic level) were initially identified using gene expression microarrays. Subsequent technological advances, such as next-generation sequencing and pathway signaling profiling, have facilitated the integration of huge volumes of data for genetic, genomic, and molecular alterations (including genetic/genomic somatic mutations), leading to the definition of additional or more precise clinically relevant subtypes [7,8]. Integrative studies have therefore considerably expanded our knowledge of cancer as a complex and heterogeneous disease.

Typically, integrative studies are designed to assess correlations between alterations that potentially drive

CrossMark

carcinogenesis. These correlations are principally positive (i.e., two alterations co-occurring) or negative (i.e., a given alteration impeding the acquisition of another specific alteration, and vice versa); for example, in breast tumors, epidermal growth factor receptor (EGFR) and human epidermal growth factor receptor 2 (HER2) tend to be simultaneously overexpressed and activated, whereas somatic mutations in *GATA-binding factor 3* (*GATA3*) and *forkhead box protein M1* (*FOXM1*) tend to be mutually exclusive [7,8]. Thus, the identification of such correlations is not only a potential source of information about the mechanisms of cancer development and/or progression but also highlights molecular dependences that could open the way to precise therapies. However, cancer evolution in a given case may be highly complex (involving genetically distinct clones), therefore the above correlations do not generally consider the time variable. Critically, we do not know how the initial perturbation(s) caused by germline mutations, particularly low-penetrance mutations, determine subsequent cancer evolution, which would encompass the correlations and molecular dependences described above.

Evidence for germline–somatic continuums

Pioneer studies in colorectal cancer depicted a continuum, or stepwise process, of genetic and molecular alterations [9]. However, the initial perturbation in this model is typically caused by a high-penetrance germline mutation, and other features of this setting differ from the principles that may apply to population-based analyses. In the context of GWAS and cancer in the general population, the following observations would support the existence of germline–somatic continuums: (i) several of the candidate genes identified in these studies are in fact well-known oncogenes or tumor suppressors at the somatic level, such as *myelocytomatosis viral proto-oncogene* (*MYC*) and *cyclin-dependent kinase inhibitor 2A* (*CDKN2A*), respectively [10,11]; (ii) this observation can be extended to signaling pathways and biological processes that are pivotal in cancer development and/or progression, such as the rat sarcoma/mitogen-activated protein kinase (RAS/MAPK)

pathway [12]; (iii) some loci have been associated with risk of different cancer types, and the candidate genes encode for proteins involved in fundamental processes altered in cancer cells, such as the *telomerase reverse transcriptase* (*TERT*) and telomere length deregulation [13]; (iv) low-penetrance germline mutations may differentially modify the risk associated with high-penetrance mutations, as observed in carriers of *BRCA1* or *BRCA2* mutations [14]; (v) in some instances, low-penetrance germline variants have been differentially associated with cancer subtypes defined according to known tumor markers, such as breast cancer associations by ERα tumor status [3]; and (vi) studies in mice and humans indicate that germline variation may influence metastasis susceptibility and patient survival [4–6]. Parallel evidence comes from the analysis of targeted alterations in complex molecular networks. For example, there is pathway convergence by perturbations caused by tumor virus proteins and germline (gene candidates from GWAS) and somatic cancer drivers [15]. Therefore, common germline genetic variation at canonical oncogenes and tumor suppressors may influence cancer risk, and these variants appear to stratify cases according to tumor features and systems-level properties (i.e., features of networks of gene and/or protein interactions). Because somatic alterations may be concurrent or mutually exclusive, and frequently define clinically relevant cancer subtypes, it may be reasonable to assume that the initial germline perturbation (mediated by a combination of low-penetrance germline mutations) in a given individual cooperates with a limited number of potential somatic alterations (Figure 1). However, how can we test this hypothesis and what are its implications for personalized medicine?

The potential of identifying germline–somatic continuums

The above hypothesis can only be assessed in the context of GWAS consortia and/or large-scale tumor studies: each continuum will represent a probabilistic trend, thus statistical power and comprehensive annotation of a large
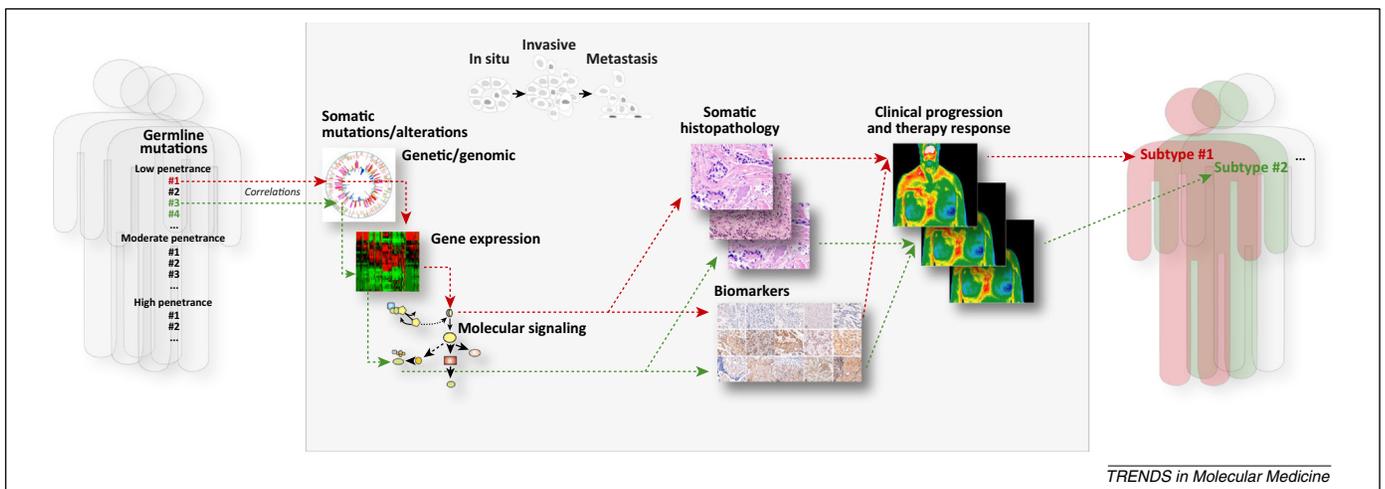

Figure 1. Illustration of the overall concept and the potential implications for personalized cancer medicine. Given an individual's germline genetic background, which includes a combination of low-penetrance germline mutations (and, less commonly, moderate- and/or high-penetrance germline mutations), a continuum (or stepwise process from an *in situ* lesion, if epithelial, to metastasis) of somatic mutations and histopathological and clinical features could be predicted to a certain extent. The continuums will help to more precisely define cancer subtypes and highlight potentially subtype-specific therapeutic approaches. The existence of these continuums would also have implications for population-level cancer screening, prevention, and treatment.

collection of cases is critical. This point links to relevant considerations regarding participatory medicine, but they fall outside the main scope of this manuscript. Thus, having envisioned the identification of germline–somatic cancer continuums, how could we benefit from this knowledge? A primary benefit would be the fundamental understanding of how each cancer case develops and progresses (Figure 1). This knowledge would further facilitate a systems-level representation of cancer, comprehensively integrating different biological levels in space and time. Next, building on this key development, two major interconnected benefits can be foreseen for the application of personalized medicine. First, information on the continuums could be harnessed for risk estimation in predictive and preventive medicine, shedding light on the probability of developing a specific cancer subtype with defined somatic alterations. Thus, the continuums would essentially help to further define cancer subtypes, which could then identify key molecular dependences. These specific dependences could be considered in the design of novel targeted therapies to prevent or impair cancer development and/or progression. Therefore, once the germline genetic variants that influence cancer development and/or progression are identified, they may be used (conditional on risk of developing a specific cancer subtype) to further guide screening, prevention, and treatment. Collectively, integrating data at the germline and somatic levels, and linking them to histopathological and clinical variables (Figure 1), potentially lays the foundations for comprehensive, personalized cancer medicine for individuals in the general population.

Concluding remarks

Considerable advances have been made in our knowledge of germline and somatic alterations in cancer. However, the germline and somatic alterations appear to be largely unconnected in a given case, particularly when considering cancer in the general population. Current evidence suggests that such connections, or germline–somatic continuums, exist and that their identification will have a fundamental impact on the development and application of personalized cancer medicine.

References

1 Varghese, J.S. and Easton, D.F. (2010) Genome-wide association studies in common cancers – what have we learnt? *Curr. Opin. Genet. Dev.* 20, 201–209
2 Michailidou, K. *et al.* (2013) Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat. Genet.* 45, 353–361
3 Garcia-Closas, M. *et al.* (2013) Genome-wide association studies identify four ER negative-specific breast cancer risk loci. *Nat. Genet.* 45, 392–398
4 Hartman, M. *et al.* (2007) Is breast cancer prognosis inherited? *Breast Cancer Res.* 9, R39
5 Hsieh, S.M. *et al.* (2009) Distinct inherited metastasis susceptibility exists for different breast cancer subtypes: a prognosis study. *Breast Cancer Res.* 11, R75
6 Azzato, E.M. *et al.* (2010) Association between a germline *OCA2* polymorphism at chromosome 15q13.1 and estrogen receptor-negative breast cancer survival. *J. Natl. Cancer Inst.* 102, 650–662
7 Cancer Genome Atlas Network (2012) Comprehensive molecular portraits of human breast tumours. *Nature* 490, 61–70
8 Curtis, C. *et al.* (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 486, 346–352
9 Vogelstein, B. *et al.* (1988) Genetic alterations during colorectal-tumor development. *N. Engl. J. Med.* 319, 525–532
10 Turnbull, C. *et al.* (2010) Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat. Genet.* 42, 504–507
11 Pomerantz, M.M. *et al.* (2009) The 8q24 cancer risk variant rs6983267 shows long-range interaction with *MYC* in colorectal cancer. *Nat. Genet.* 41, 882–884
12 Menashe, I. *et al.* (2010) Pathway analysis of breast cancer genome-wide association study highlights three pathways and one canonical signaling cascade. *Cancer Res.* 70, 4453–4459
13 Rafnar, T. *et al.* (2009) Sequence variants at the *TERT-CLPTM1L* locus associate with many cancer types. *Nat. Genet.* 41, 221–227
14 Barnes, D.R. and Antoniou, A.C. (2012) Unravelling modifiers of breast and ovarian cancer risk for *BRCA1* and *BRCA2* mutation carriers: update on genetic modifiers. *J. Intern. Med.* 271, 331–343
15 Rozenblatt-Rosen, O. *et al.* (2012) Interpreting cancer genomes using systematic host network perturbations by tumour virus proteins. *Nature* 487, 491–495