Cancer Letters 340 (2013) 227–233

Contents lists available at ScienceDirect

Cancer Letters

journal homepage: www.elsevier.com/locate/canlet

Mini-review

Genome-wide sequencing to identify the cause of hereditary cancer syndromes: With examples from familial pancreatic cancer

Nicholas J. Roberts [a], Alison P. Klein [b,c,d,*]

[a] Ludwig Center for Cancer Genetics and Therapeutics, The Johns Hopkins Kimmel Cancer Center, Baltimore, MD 21231, USA
[b] Department of Oncology, The Johns Hopkins School of Medicine, Baltimore, MD 21231, USA
[c] Department of Pathology, The Johns Hopkins School of Medicine, Baltimore, MD 21231, USA
[d] Department of Epidemiology, Bloomberg School of Public Health, The Johns Hopkins University, Baltimore, MD 21231, USA

ARTICLE INFO

Keywords:
Hereditary cancer
Genome sequencing
Cancer predisposition genes

ABSTRACT

Advances in our understanding of the human genome and next-generation technologies have facilitated the use of genome-wide sequencing to decipher the genetic basis of Mendelian disease and hereditary cancer syndromes. However, the application of genome-wide sequencing in hereditary cancer syndromes has had mixed success, in part, due to complex nature of the underlying genetic architecture. In this review we discuss the use of genome-wide sequencing in both Mendelian diseases and hereditary cancer syndromes, highlighting the potential and challenges of this approach using familial pancreatic cancer as an example.

1. Introduction

The publication of the draft human reference genome and advent of high-throughput sequencing technology have ushered in the genomics era. These new sequencing technologies have driven down the cost of genome sequencing to within a hair's breadth of the $1000 dollar genome. Over the last decade, per genome cost of sequencing has dropped precipitously, from approximately $70 million dollars in 2002, to nearly $7000 dollars in 2012 [1]. Similarly, the time to generate genomic sequence has also decreased, with over 10^{10} bases per instrument per day with the latest next generation sequencing technology. These advances have resulted in the application of genome-wide sequencing to elucidate the underlying genetic basis of hereditary disease. Numerous examples of high-throughput, genome-wide approaches for disease gene identification are available for Mendelian diseases and these provide a starting framework for the analysis of complex diseases. Such gene discovery approaches share common elements, such as case selection and sample acquisition, genome-wide sequencing, data analysis to identify candidate genes with subsequent functional or population-based validation (Fig. 1). However, the application genome-wide sequencing in the background of complex disease faces unique challenges, in particular the identification of candidate genes using either filter-based or statistical based tests in the face of underlying genetic heterogeneity. In this review, we: (1) discuss the current use of genome-wide sequencing to identify the genetic basis of Mendelian diseases and hereditary cancer syndromes, (2) outline current knowledge about the genetic basis of familial pancreatic cancer, and (3) highlight the challenges of using genome-wide sequencing to identify pancreatic cancer susceptibility genes. We focus on familial pancreatic cancer because while genomic sequencing approaches have successfully identified familial pancreatic cancer genes, the challenges of using these approaches, specifically underlying genetic heterogeneity and its impact in study size requirements and candidate variant selection are faced by researchers using these approaches to study other cancer syndromes or other complex diseases.

2. Genome-wide sequencing and hereditary disease

The first use of genome-wide sequencing to identify the genetic cause of a hereditary disease used Sanger sequencing to analyze the coding genes in the germline and tumor of a patient with familial pancreatic cancer [2]. Jones and colleagues employed a filter-based approach using the hypothesis that any susceptibility gene will be inactivating and heterozygous in the germline of the patient and contain a second mutation in the gene of the tumor, that is, it obeys the classical two-hit model of a tumor suppressor gene. Using this filter-based approach, 15,461 germline variants not seen in the human reference genome were narrowed to 3 in genes *SERPINB12*, *RAGE* and *PALB2*. *SERPINB12* and *RAGE* were eliminated from further analysis as nonsense variants were common in healthy individuals. This left *PALB2* as the putative susceptibility gene, highlighting the power of genome-wide sequencing to iden-

* Corresponding author at: Department of Oncology, The Johns Hopkins School of Medicine, Baltimore, MD 21231, USA.
E-mail address: aklein1@jhmi.edu (A.P. Klein).

0304-3835/$ - see front matter © 2012 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.canlet.2012.11.008

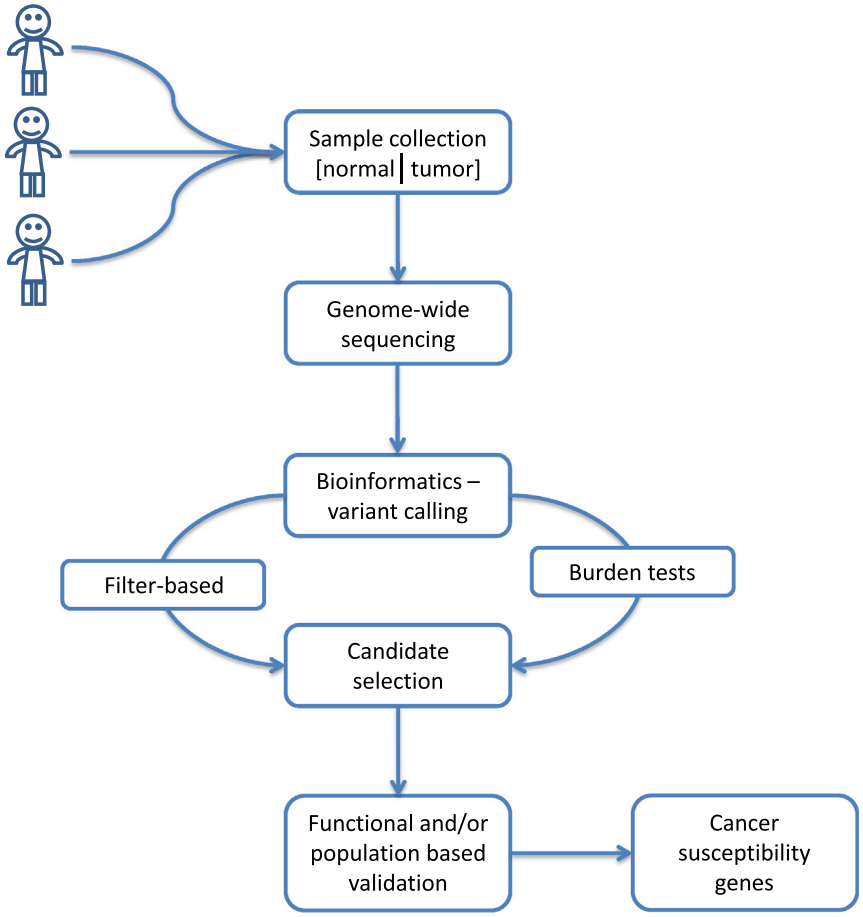

Fig. 1. Schematic representation of cancer susceptibility gene discovery with genome-wide sequencing. Discovery relies on careful case identification and collection of matched tumor and normal samples. Genome-wide sequencing is conducted on a cohort of samples and bioinformatic analyses performed to identify individual variation. Candidate selection carried out with either filterbased approaches or burden tests before functional or population-based validation. Validated cancer susceptibility genes may be used for screening, risk assessment, prognostics and therapeutic targets.

tify the genetic cause of a hereditary disease and the need of a good detective when approaching such a study.

Shortly after the discovery of *PALB2* as a pancreatic cancer susceptibility gene, genome-wide sequencing was successfully employed to identify the genetic basis of a Mendelian disorder. Ng and colleagues used next-generation technology to sequence the exome of 4 unrelated individuals with Freeman-Sheldon Syndrome, an autosomal dominant disorder of congenital arthrogryposis [3]. This study provided a powerful demonstration of the ability of next-generation technology to identify the genetic cause of a Mendelian disorder. Central to this proof-of-principle study were computational methods to find the single causative variant in the 20,000 rare variants per exome, the proverbial needle in the haystack. To do this, Ng and colleagues used a filter-based approach, eliminating common variants using control samples and the database of single nucleotide polymorphisms (dbSNP), and searching for genes where each affected individual harbored a rare variant. Their efforts identified one gene, MYH3, previously identified as the cause of Freeman-Sheldon Syndrome [4]. Subsequently, numerous Mendelian disorders where traditional linkage studies were unsuitable have been subjected to genome-wide sequencing approaches and the genetic basis of the phenotype defined [5–8]. Such studies predominantly use whole-exome sequencing which limits search to exons and small segments of adjacent sequence, that is, coding and splice site variants. Fewer studies use whole-genome sequencing due to added cost and computational difficulties associated with a 20-fold increase in data per individual. However, studies utilizing whole-genome sequencing are not limited to coding variants and as such can find associations between copy number variation or non-coding variation and disease. A study by Jaeger and colleagues [9] used whole-genome sequencing in individuals with hereditary mixed polyposis syndrome (HMPS) to identify a duplication, 5′ to *GREM1*, that results in ectopic expression of *GREM1* in colonic epithelium. As the authors note, further evidence for the causal nature of *GREM1* duplications in HMPS patients is given by juvenile polyposis syndrome (JPS). The majority of JPS patients harbor germline mutations in *SMAD4* or *BMPR1A* [10], members of the bone morphogenetic protein (BMP) signaling pathway [11]. Interestingly, *GREM1* is thought to negatively regulate BMP pathway signaling, providing a contemporaneous link between HMPS and JPS.

Hot on the heels of these successes were further large scale genome-wide sequencing studies to identify novel susceptibility genes in hereditary cancer syndromes. In 2012, Roberts and colleagues [12] reported on the whole-genome and whole-exome sequencing of 38 familial pancreatic cancer cases. Here they describe the aggregation of deleterious *ATM* variants in the germline of 6 pancreatic cancer cases from 2 different kindreds. Similar to studies of Mendelian disorders described previously, Roberts and colleagues employ a filter-based approach using assumptions about the genetic architecture of susceptibility genes in familial pancreatic cancer to narrow down candidate genes for further study. In this study, germline variants present in the Human Gene Mutation Database (HGMD) [13] and associated with disease in people, were