Clinical relevance of cancer genome sequencing

Chee Seng Ku, David N Cooper, Dimitrios H Roukos

Chee Seng Ku, Department of Medical Epidemiology and Bio-statistics, Karolinska Institute, SE-17177 Stockholm, Sweden
David N Cooper, Institute of Medical Genetics, School of Medicine, Cardiff University, Cardiff CF14 4YU, United Kingdom
Dimitrios H Roukos, Department of Surgery, Ioannina University School of Medicine, 45110 Ioannina, Greece
Dimitrios H Roukos, Centre for Biosystems and Synthetic Genomic Network Medicine, Centre for BioSystems and Genomic Network Medicine, Ioannina University, 45110 Ioannina, Greece

Author contributions: All authors contributed to the writing of this manuscript, and all of them have read and approved the final version to be published.

Correspondence to: Dimitrios H Roukos, Associate Professor, Department of Surgery, Ioannina University School of Medicine, 45110 Ioannina, Greece. droukos@cc.uoi.gr

Telephone: +30-26510-07423 Fax: +30-26510-07094

Received: June 7, 2012 Revised: September 21, 2012
Accepted: September 29, 2012
Published online: April 7, 2013

INTRODUCTION

Sporadic cancers are complex diseases that are caused by the accumulation of somatic mutations that are acquired by the genomes of the cells of the tissue in which the cancer originated[1]. The importance of identifying the “driver” (causal) somatic mutations amongst the much more numerous “passenger” mutations has long been recognized. However, previous cancer genome sequencing studies have also been constrained by technological limitations. Although several early attempts were made to sequence the coding regions of the majority of the consensus coding sequence and/or RefSeq genes in several cancers (e.g., breast and colorectal), these studies were conducted in “brute-force” mode employing traditional low-throughput polymerase chain reaction-Sanger sequencing methods[3-5]. The advent of next-generation sequencing (NGS) technologies has revolutionized the sequencing of cancer genomes, the first example of which employed whole-genome sequencing (WGS) to characterize an acute myeloid leukemia (AML) genome thereby identifying numerous tumor-specific mutations[4]. This study clearly demonstrated the technical feasibility of applying NGS to interrogate the genome-wide somatic mutational spectra of entire cancer genomes in tandem with paired constitutional DNA samples.

In parallel, the development of a variety of exome enrichment methods to selectively capture the entire collection of exons in the human genome has made whole-exome sequencing (WES) technically feasible[6]. Coupling this development to the high-throughput NGS techniques has allowed the exome to be sequenced very...
rapidly and in unprecedented detail. In comparison to WGS, WES is more affordable for larger sample sizes and is analytically less challenging since only 1%-2% of the entire genome needs to be sequenced\(^\text{[6-8]}\). As a result, a larger number of cancer DNA samples have been sequenced by WES than WGS in attempts to identify recurrent mutations (i.e., identical mutations that recur in different samples) and highly mutated genes (genes harboring multiple mutations in a significant proportion of the cancer samples)\(^\text{[6-12]}\). The other reason for the more widespread adoption of WES has been that the mutations identified within protein coding regions are inherently easier to interpret than those in the non-coding regions, which still remain largely uncharacterized in functional terms. In addition to the advances being made in characterizing the somatic mutational landscape in various cancers, the applications of cancer genome sequencing in a clinical setting have also become increasingly numerous.

**SOMATIC MUTATIONS IN SPORADIC CANCERS**

WGS and WES have been commonly applied to study the patterns of somatic mutation in a range of different cancers\(^\text{[13-16]}\). Collectively, these endeavors have generated new insights into the mutational landscape of various cancers, and have resulted in the identification of a large number of recurring mutations as well as many highly mutated genes. For example, in the context of gastrointestinal cancers, WES of 15 gastric adenocarcinomas and their matched normal DNAs succeeded in identifying several frequently mutated genes such as TP53, PIK3CA and ARID1A\(^\text{[18]}\). In addition, it was found that cell adhesion was the most enriched biological pathway among the frequently mutated genes. More importantly, mutations in three chromatin remodeling genes (ARID1A, MLL and MLL) were detected in almost half of the gastric cancers examined\(^\text{[18]}\). In fact, an earlier study which performed WES in 22 gastric cancer samples also identified frequent inactivating mutations in ARID1A, which encodes a member of the switch/sucrose non-fermenting chromatin remodeling family. Further, the mutational spectrum for ARID1A was found to differ between molecular subtypes of gastric cancer\(^\text{[7]}\). In similar vein, mutations in multiple chromatin regulator genes such as ARID1A, ARID1B, ARID2, MLL and MLL3 were also found in about half of hepatocellular carcinomas through WGS\(^\text{[16]}\). The consistent finding of mutations in chromatin remodeling genes in different cancers, which also included renal carcinoma and glioblastoma multiforme, further highlights a close inter-relationship (or possibly a “synergy” interaction effect) between somatic mutations and aberrant epigenetic regulation in the pathogenesis of cancers\(^\text{[8,17-19]}\).

In addition to individual studies, technological advances have made possible large-scale international projects such as the International Cancer Genome Consor-
analytical tool for the entire genome. More importantly, the molecular confirmatory diagnosis carries important clinical implications for the treatment and management of the patient\cite{23}. Similarly, WGS was also employed to resolve the genetic basis of a suspected cancer susceptibility syndrome based upon the early onset of several primary tumors\cite{24}. Further, therapeutic prediction has also benefited from NGS as a powerful discovery tool. For example, a recent study employed a targeted NGS approach to sequence 138 cancer genes in melanomas derived from a patient (before and after relapse) and succeeded in identifying the underlying genetic mutation in the \textit{MEK1} gene responsible for acquired resistance to PLX4032 ( vemurafenib) after an initial dramatic response, revealing a novel mechanism of acquired drug resistance\cite{25}.

The potential applications of cancer genome sequencing in the clinical arena are promising, but what are the challenges associated with their adoption? As WGS and WES are high-throughput methods which generate huge amounts of data, our ability to perform both the analysis and the interpretation of the data in a clinically relevant time-frame is critical. This challenge is being addressed in the context of a “comprehensive genomic approach” where WGS, WES and transcriptome sequencing are applied to cancer samples to evaluate their clinical utility and feasibility (in terms of technical, time and cost)\cite{26}. In particular, the time required, from biopsy sampling and wet-lab experiments to computational analysis and initial results, was streamlined to just 24 d with the cost of all the sequencing and analysis estimated to be only USD5400. An obvious advantage of this “integrative genomic approach” is that the findings can be cross-validated more efficiently. For example, both WGS and WES detected an amplification event on chromosome 13q spanning the \textit{CDK8} gene in a metastatic colorectal carcinoma; the over-expression of \textit{CDK8} was confirmed by transcriptome sequencing. Although this “comprehensive genomic approach” was shown to be both time- and cost-effective, the handling and interpretation of the huge amount of genomic data remains a key issue. To address this challenge, it was proposed that a multi-disciplinary “sequencing tumor board” (which included professionals from multiple disciplines such as clinicians, geneticists, pathologists, biologists, bioinformatic specialists and bioethicists) should take responsibility for the clinical interpretation of the sequencing data obtained from each patient\cite{26}.

**FAMILIAL CANCER SYNDROMES**

In addition to the investigation of somatic mutations in sporadic cancers, cancer genome sequencing has also made significant advances in relation to the study of the germline mutations underlying familial cancer syndromes. The early successes in the identification of causal mutations and genes for familial cancer syndromes (e.g., \textit{RB1} and \textit{APC}) were achieved by painstaking family linkage analysis and positional cloning. However, the genetic causes of many familial cancer syndromes have remained elusive. For example, \textit{CDH1} was the first and only causal gene identified for hereditary diffuse gastric cancer through linkage analysis\cite{27}, but germline mutations in this gene account for only a proportion of hereditary diffuse gastric cancer cases. Thus germline mutations in \textit{CDH1} were found in 30%-40% of clinically defined families with hereditary diffuse gastric cancer from different ethnic backgrounds\cite{28,29}. This suggests that an as-yet-to-be identified gene(s) is likely to be responsible for the remaining cases unexplained by \textit{CDH1}. On the other hand, whereas most Lynch syndrome cases can be accounted for by mutations in DNA mismatch repair genes, the genetic basis of familial colorectal cancer type X still remains elusive\cite{30,31}. Similarly, the genetic causes of other familial cancer syndromes, such as familial pancreatic cancer, still remain largely unknown\cite{32,33}. In a fashion similar to that noted with the identification of somatic mutations, cancer genome sequencing provides new opportunities to identify germline mutations for familial cancer syndromes. This is well exemplified by the case of hereditary pheochromocytoma, a rare neural crest cell tumor; by harnessing the latest technological advances, germline mutations in \textit{MAX} were identified in three unrelated individuals with hereditary pheochromocytoma by WES\cite{34}. The segregation of two \textit{MAX} gene variants with hereditary pheochromocytoma was observed in families from whom DNA from affected relatives was available. Further, additional data to support the causative role of the \textit{MAX} variants came from their absence (or non-detection) in more than 750 population-matched control chromosomes. Additional screening for \textit{MAX} mutations in 59 cases lacking germline mutations in known genes for hereditary pheochromocytoma then identified two additional truncating mutations and three missense variants in the gene\cite{34}. Following this discovery, a recent study found that germline mutations in \textit{MAX} are responsible for 1.12% pheochromocytomas or paragangliomas (both are genetically heterogeneous neural crest-derived neoplasms) by sequencing \textit{MAX} in 1694 patients\cite{35}.

In addition to its role in research discovery, cancer genome sequencing has also been used as a diagnostic tool to detect known germline mutations for familial cancer syndromes. Indeed, by leveraging technological advances in genomic sequence enrichment methods and NGS technologies, studies have developed NGS-based diagnostic tests for breast and ovarian cancers and Lynch syndrome. For example, Walsh et al\cite{36} designed custom oligonucleotides in solution to capture 21 genes responsible for hereditary breast and ovarian cancers, and the enriched genomic DNA was then subjected to sequencing using an NGS platform. This NGS-based test was evaluated in 20 women diagnosed with breast or ovarian cancer and with a known mutation in one of the genes responsible for inherited predisposition to these cancers. The results were very promising in that all the known
point mutations and small indel mutations (ranging from 1 bp to 19 bp), as well as large genomic duplications and deletions (ranging from 160 bp to 101,013 bp), were detected in all the samples. The potential to detect different mutations of various sizes has further demonstrated the technical advantages of NGS-based tests over conventional PCR-Sanger sequencing methods. For example, two different tests were offered separately to detect point mutations and large deletions/amplifications for genetic testing of BRCA1/2 genes, respectively. Similarly, attempts have also been made to incorporate custom genomic enrichment and NGS methods into the genetic diagnostic testing of Lynch syndrome by capturing every exon in a panel of 22 genes (most of which are known to be associated with hereditary colorectal cancer) followed by NGS.

Technological advances have facilitated the accessibility of cancer genome sequencing in the clinical arena. In addition to the custom sequence enrichment methods (i.e., either based on polymerase chain reaction amplification such as Fluidigm and RainDance technologies, or based on target-probe hybridization such as the Agilent and Nimblegen technologies) that allow one to selectively capture the genomic regions of interest, the arrival of several bench-top NGS instruments has not only made the sequencing of a panel of genes highly feasible technically but also cost-effective. This is an important step towards the development and adoption of NGS-based diagnostic tests in the clinic. The bench-top NGS instruments (Roche 454 Genome Sequencer Junior, Ion Torrent Personal Genome Machine Sequencer and Illumina MiSeq) have a much lower throughput (ranging from > 10 Mb to > 1 Gb per run) than the conventional high-throughput NGS machines. The bench-top NGS instruments are therefore more suitable in terms of their throughput for sequencing panels of genes (as discussed earlier for the panels of genes for breast/ovarian cancers and Lynch syndrome) than WES or WGS. Further, sample indexing (or barcoding) is also available for the bench-top NGS instruments which can further optimize sample throughput and cost-effectiveness by multiplexing up to several tens of samples for sequencing. However, the level of multiplexing is dependent on the sizes of the regions to be sequenced and the throughputs of the instruments being used. Although it remains to be demonstrated in the context of cancer, WES has been widely assessed and shown as a promising diagnostic tool for various Mendelian disorders. In addition to diagnosis, WGS has also been applied to optimize patient treatment regimens, although not in the context of cancer. In the context of inherited disease, WGS has been applied to sequence a fraternal twin pair diagnosed with dopa (3,4-dihydroxyphenylalanine)-responsive dystonia (OMIM 128230); germline compound heterozygous mutations were identified in the SPR gene encoding sepiapterin reductase. As a result, supplementation of L-dopa therapy with 5-hydroxytryptophan has led to clinical improvements in both twins.

PERSPECTIVES AND CONCLUSIONS

NGS technologies have already made a major contribution to characterizing somatic mutations in cancer genomes. This endeavor will be further accelerated by international initiatives such as the International Cancer Genome Consortium. Although the number of studies is currently still very limited, NGS should also be applied to identify germline mutations in those familial cancer syndromes whose genetic causes have not yet been fully characterized. On the other hand, the successful demonstration of the applications of NGS/WGS/WES in a clinical setting such as cancer diagnostics are likely to be just the first examples of how the new technologies will prove their worth; the numbers are expected to increase in the coming year.

So far, the applications of NGS in a clinical setting have been very promising. However, challenges ranging from technical, analytical and interpretational, to the need for a considerable number of well-trained professionals from a range of disciplines in these genomic technologies must be further addressed before the adoption of NGS-based tests in the clinic. The technical challenges include, for example, incomplete capture of the exons in WES and uneven sequencing across the genome which might result in poor sequence coverage in some of the regions and affect both the sensitivity and specificity of variant detection. Having specialists trained in genomic technologies is critical to (1) obtaining fully informed consent from patients in relation to the genomic tests; (2) ensuring accurate and reliable interpretation of the data for clinical decision-making; and (3) counselling the patients on the basis of the results obtained. It is also evident from the Global Cancer Genomics Consortium (GCGC) that the translation of emerging cancer genomics knowledge into clinical applications can only be achieved through the integration of multidisciplinary expertise. The GCGC is an international collaborative platform that brings cancer biologists and cutting edge high-throughput genomics expertise together with medical oncologists and surgical oncologists to address the most important translational questions that are central to cancer research, diagnosis and treatment.

As to test affordability, although the total cost of sequencing for several genomie experiments was only a few thousands of USD per patient, it should be appreciated that this is unlikely to be the final chargeable cost to the patients. The cost of sequencing is currently plummeting and will become ever cheaper in the future with new developments. However, it should be appreciated that hidden costs are likely to be incurred for data storage, interpretation of results and subsequent clinical consultation.

Further, handling of the complex ethical issues such as revealing findings that might be considered incidental to the initial testing (WGS and WES) procedure must also be given serious consideration. Determining what to disclose and what not to disclose to the patients is...
likely to be quite challenging, e.g., those results which are
deemed clinically important i.e., those which could have a
direct impact on the patient’s care or management, but
which are irrelevant to the initial purpose of the diagnostic
test, (i.e., incidental findings). Have the patients the
righ to be informed about those results which are/might be
clinically important but not actionable e.g., mutations
that are considered likely to predispose to certain inher-
ited diseases, although preventive treatments are not yet
available? Adequate consultation must also be given to
the reporting of results that are of unknown clinical im-
portance. This raises concerns as to whether periodic re-
analysis of the WGS/WES data might be needed, which
in turn would lead to some practicality issues potentially
incurring additional costs. Finally, any results from WGS-
and WES-based tests that would affect clinical decision-
making must be properly validated or the tests must be
performed in a heavily regulated clinical setting accord-
ing to the College of American Pathologists/Clinical
Laboratory Improvement Amendments.

It is widely anticipated that cancer genome sequenc-
ing or the NGS-based tests will gradually become more
accessible in clinical practice once the associated chal-
enges and ethical issues have been adequately addressed.
Irrespective of the challenges that still remain to be over-
come, the application of NGS in the clinic appears inevi-
table.

REFERENCES

1 Stratton MR, Campbell PJ, Futreal PA. The cancer genome.
Nature 2009; 458: 719-724 [PMID: 19360079 DOI: 10.1038/nature07943]

2 Sjöblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber
consensus coding sequences of human breast and colorectal cancers.
Science 2006; 314: 268-274 [PMID: 16959974 DOI: 10.1126/
science.1133427]

haults P, Markowitz SD, Parmigiani G, Kinzler KW, Velculescu VE. The
genome landscapes of human breast and colorectal cancers.
Science 2007; 318: 1108-1113 [PMID: 17932254 DOI: 10.1126/science.1145720]

4 Ley TJ, Mardis ER, Ding L, Fulton B, McLellan MD, Chen
K, Dooling D, Dunford-Shore BH, McGrath S, Hickenbo-
DNA sequencing of a cytogenetically normal acute myeloid
DOI: 10.1038/nature07485]

5 Mertes F, Elsharawy A, Sauer S, van Helvoort JM, van
der Zaag PG, Franke A, Nilsson M, Lehrach H, Brookes AJ.
Targeted enrichment of genomic DNA regions for next-gen-
eration sequencing. Brief Funct Genomics 2011; 10: 374-386
[PMID: 22121152 DOI: 10.1093/bfgp/elr033]

6 Yan XJ, Xu J, Gu ZH, Pan CM, Lu G, Shen Y, Shi JY, Zhu
YM, Tang L, Zhang XW, Liang WX, Mi JQ, Song HD, Li QK, Chen Z, Chen SJ. Exome sequencing identifies somatic mu-
tations of DNA methyltransferase gene DNMT3A in acute
21398634 DOI: 10.1038/ng.788]

7 Wang K, Kan J, Yuan ST, Shi ST, Chu KM, Law S, Chan TL,
Kan Z, Chan AS, Tsui WY, Lee SP, Ho SL, Chan AK, Cheng
GH, Roberts PC, Rejto PA, Gibson NW, Pocolyko DJ, Mao M,
Xu J, Leung SY. Exome sequencing identifies frequent muta-
tion of ARID1A in molecular subtypes of gastric cancer. Nat
Genet 2011; 43: 1219-1223 [PMID: 22037554 DOI: 10.1038/
ng.9682]

8 Varela I, Tarpey P, Raine K, Huang D, Ong CK, Stephens
Mancera PA, Mustonen V, Fischer A, Adams DJ, Rust A, Chan-on W, Subimerb C, Dykema K, Furge K, Campbell
PJ, Teh BT, Stratton MR, Futreal PA. Exome sequencing
identifies frequent mutation of the SWI/SNF complex gene
[PMID: 21248752 DOI: 10.1038/nature09639]

9 Wei X, Walia V, Lin JC, Teer JK, Prickett TD, Gartner J, Davis
S, Stemke-Hale K, Davies MA, Gershewndal JE, Robinson
W, Robinson S, Rosenberg SA, Samuel Y. Exome sequenc-
ing identifies GRN2A as frequently mutated in melanoma.
Nat Genet 2011; 43: 442-446 [PMID: 21492947 DOI: 10.1038/
ng.9180]

10 Barbieri CE, Baca SC, Lawrence MS, Demichelis F, Blattner
M, Theurillat JP, White TA, Stojanov P, Van Allen E, Stran-
sky N, Nickerson E, Chae SS, Boysen G, Auclair D, Onofrio
RC, Park K, Kitabayashi N, MacDonald TY, Sheikh K, Vu-
ong T, Guiducci C, Cibulskis K, Sivachenko A, Carter SL,
Saksena G, Voet D, Hussain WM, Ramos AH, Winckler W,
Redman MC, Ardile K, Tewari AK, Mosquera JM, Rupp N,
Wild PJ, Moch H, Morrissey C, Nelson FS, Kantoff PW, Ga-
briel SB, Golub TR, Meyerson M, Lander ES, Getz G, Robin
MA, Garraway LA. Exome sequencing identifies recurrent
SPOP, FOXA1 and MED12 mutations in prostate cancer.
Nat Genet 2012; 44: 685-689 [PMID: 2261019 DOI: 10.1038/
ng.2279]

11 Nikolaeff SI, Rimoldi D, Iseli C, Valsesia A, Robyr D, Geh-
rig C, Harshman K, Guipponi M, Buchak O, Zoete V, Mich-
kelin O, Muehlethaler K, Speiser D, Beckmann JS, Xenarios
I, Halazonetis TD, Jongeneel CV, Stevenson BJ, Antonarakis
SE. Exome sequencing identifies recurrent somatic MAP2K1
and MAP2K2 mutations in melanoma. Nat Genet 2012; 44:
133-139 [PMID: 22197931 DOI: 10.1038/ng.1026]

12 Stark MS, Woods SL, Gartside MC, Bonazzi VF, Dutton-Re-
gerest K, Aoude LG, Chow D, Sereduk C, Niemi NM, Tang
N, Ellis JJ, Reid J, Zismann V, Tyagi S, Muzny D, Newsham I,
Kan Z, Chan AS, Tsui WY, Lee SP, Ho SL, Chan AK, Cheng
GH, Roberts PC, Rejto PA, Gibson NW, Pocolyko DJ, Mao M,
Xu J, Leung SY. Exome sequencing identifies recurrent
SPOP, FOXA1 and MED12 mutations in prostate cancer.
Nat Genet 2012; 44: 685-689 [PMID: 2261019 DOI: 10.1038/
ng.2279]

13 Meyerson M, Gabriel S, Getz G. Advances in understanding
cancer genomes through second-generation sequencing. Nat
Rev Cancer 2010; 11: 685-696 [PMID: 20847746 DOI: 10.1038/
nr2841]
April 7, 2013 | Volume 19 | Issue 13

Ku CS et al. Cancer genome sequencing
Ku CS et al. Cancer genome sequencing


P- Reviewers Lindblom A, De Petro G
S- Editor Gou SX L- Editor A E- Editor Xiong L