REVIEW ARTICLE

An emerging place for lung cancer genomics in 2013

Marissa G. Daniels[1], Rayleen V. Bowman[1], Ian A. Yang[1], Ramaswamy Govindan[2], Kwun M. Fong[1]

[1]The University of Queensland and the Prince Charles Hospital Thoracic Research Centre, the Prince Charles Hospital, Chermside 4032, Australia; [2]Division of Oncology, Washington University School of Medicine, St. Louis, MO 63021, USA

ABSTRACT

Lung cancer is a disease with a dismal prognosis and is the biggest cause of cancer deaths in many countries. Nonetheless, rapid technological developments in genome science promise more effective prevention and treatment strategies. Since the Human Genome Project, scientific advances have revolutionized the diagnosis and treatment of human cancers, including thoracic cancers. The latest, massively parallel, next generation sequencing (NGS) technologies offer much greater sequencing capacity than traditional, capillary-based Sanger sequencing. These modern but costly technologies have been applied to whole genome-, and whole exome sequencing (WGS and WES) for the discovery of mutations and polymorphisms, transcriptome sequencing for quantification of gene expression, small ribonucleic acid (RNA) sequencing for microRNA profiling, large scale analysis of deoxyribonucleic acid (DNA) methylation and chromatin immunoprecipitation mapping of DNA-protein interaction.

With the rise of personalized cancer care, based on the premise of precision medicine, sequencing technologies are constantly changing. To date, the genomic landscape of lung cancer has been captured in several WGS projects. Such work has not only contributed to our understanding of cancer biology, but has also provided impetus for technical advances that may improve our ability to accurately capture the cancer genome. Issues such as short read lengths contribute to sequenced libraries that contain challenging gaps in the aligned genome. Emerging platforms promise longer reads as well as the ability to capture a range of epigenomic signals. In addition, ongoing optimization of bioinformatics strategies for data analysis and interpretation are critical, especially for the differentiation between driver and passenger mutations.

Moreover, broader deployment of these and future generations of platforms, coupled with an increasing bioinformatics workforce with access to highly sophisticated technologies, could see many of these discoveries translated to the clinic at a rapid pace. We look forward to these advances making a difference for the many patients we treat in the Asia-Pacific region and around the world.

KEY WORDS

High-throughput nucleotide sequencing; genomics; lung neoplasms; non-small cell lung carcinoma (NSCLC); small cell lung carcinoma (SCLC)

J Thorac Dis 2013;5(S5):S491-S497. doi: 10.3978/j.issn.2072-1439.2013.10.06

Introduction

Lung cancer is the most common and deadly (1) human cancer, called "a global scourge" (2) with a dismal prognosis. The relative 5-year survival for patients with this disease is 14% (3), and has remained largely unchanged for years.

Only recently had research interest focused on lung cancer, as more effective therapies have become available. Indeed, some consider lung cancer to be a "poster boy" (4) for personalized medicine. Much of this hope and hype comes from the potential for advances in translational genomics to improve understanding and management of this cancer.

Deoxyribonucleic acid (DNA) sequencing technologies were introduced in the mid 1970s (5,6). When the first human, naturally occurring tumorigenic somatic mutation was discovered in 1982 (7), it became clear that sequencing the cancer genome was a necessary next step (8). Completion of the first human reference genome (9,10) then stimulated technological advances which enabled the first human cancer genome to be sequenced only four years later (11).

Sanger sequencing was first described in 1975 (5) as a "rapid method for determining sequences in DNA", by providing accuracy and sequence contiguity that remains unmatched(12).

Corresponding to: Marissa G. Daniels. The University of Queensland and The Prince Charles Hospital Thoracic Research Centre, The Prince Charles Hospital, Chermside 4032, Australia. Email: m.daniels@uq.edu.au.

Submitted Oct 10, 2013. Accepted for publication Oct 12, 2013.
Available at www.jthoracdis.com

ISSN: 2072-1439
© Pioneer Bioscience Publishing Company. All rights reserved.

Nowadays, the rapidity and scalability of next generations sequencing (NGS) technologies are quickly improving our ability to explore the cancer genome and exponentially advancing our understanding of lung cancer pathogenesis, its diagnosis and treatment (13-16).

Next generation technologies for whole genome sequencing (WGS)

Past techniques for exploration of the cancer genome, such as capillary electrophoresis based "Sanger" sequencing, array-based genome-wide analysis of amplifications and deletions (17), gene expression arrays (18) and retrovirus mediated expression screening techniques, have successfully identified genomic drivers of cancer (19-21), and novel therapeutic targets (20,22). Commercially available since 2004, next generation sequencing (NGS) platforms now provide high throughput, massively parallel techniques able to generate sequence data orders of magnitude more quickly, and at lower cost, than traditional techniques (13,23).

Comprehensive genome sequencing techniques are now used to study a range of genetic diseases (15). While many NGS platforms require significant investments in infrastructure, they share methodological similarities. Massively parallel sequencing begins with the generation of a DNA library onto which platform-specific "adaptors" are bound. This library is then fixed to a solid surface and each fragment is amplified so that the sequencing reaction produces a detectable signal. The library is sequenced in a series of automated, repetitive steps (14).

There are a large number of potential applications for NGS technologies (24), such as WGS and whole exome sequencing (WES) to find novel mutations (25); paired-end and mate-pair sequencing to identify structural variations (26); targeted resequencing for mutation discovery and validation (27); transcriptome sequencing for quantification of gene expression and discovery of transcribed mutations (28); small RNA-sequencing for microRNA profiling (29); large scale analysis of DNA methylation (30) and chromatin immunoprecipitation for genomic mapping of DNA-protein interactions (31). It is likely that there will be refinements in future that are not even envisaged in these early days of genomic research.

NGS WGS of lung cancer

The advances of the past decade of genomics research demonstrate the power of massive genomic surveys (15). The passion and dedication with which large scale, comprehensive studies of the human genome have been undertaken is second only to the fervor and fecundity driving advances in requisite research infrastructure, such as experimental technology, bioinformatics and ethics (32). The spectrum of genomic changes seen in lung cancer is beyond the scope of this brief overview, which will outline the methods and studies used so far in the application of NGS technologies to the study of this cancer.

Although early tools for genome-wide exploration provided putative therapeutic targets, the majority of tumours still lacked an identified molecular driver. The subsequent introduction of NGS technologies has revolutionized our understanding of cancer biology, the processes of carcinogenesis as well as its molecular drivers, and the range of techniques with which to explore the genome. The application of these platforms to the optimization of lung cancer outcomes promises the potential to transform lung cancer care.

Lung cancer—early steps towards personalized medicine

Before the development of massively parallel sequencing, much of our understanding of the molecular pathology of lung cancer was based on techniques such as mismatch repair detection (33), sequencing of candidate genes (34), single nucleotide polymorphism (SNP) arrays (21) and gene expression analysis (35). High throughput sequencing technologies now enable comprehensive examination of the lung cancer exome (36) and genome (37,38).

An important and key global, collaborative approach using these technologies is led by The International Cancer Genome Consortium (ICGC) and The Cancer Genome Atlas (TCGA) (39), which bring the promise of a truly personalized approach to cancer care, including the two major subtypes of lung cancer, adenocarcinoma (AC) and squamous cell carcinoma (SCC), in the first instance.

WGS of lung cancer

The first applications of massively parallel sequencing to the study of lung cancer were understandably carefully designed and focused studies (Table 1). In 2008, Campbell and colleagues published the genomes of 2 lung cancer cell lines, 1 derived from a neuroendocrine tumour and another from a small cell lung cancer (SCLC) (26). The second sequence of a SCLC cell line was published in 2010 (40). The first application of WGS to DNA extracted from lung tumour paired with normal lung was the same year by Lee *et al.* (38) Then, a second paired non-small cell lung cancer (NSCLC)/normal lung sequence was published in 2011, from a never-smoking patient with AC (37).

These initial WGS studies were performed using a range of sequencing technologies (Table 2), including combinatorial probe anchor ligation (cPAL) as well as the platforms of Illumina and Life Technologies.

Illumina's massively parallel sequencing technology is based on sequencing by synthesis (47). After library preparation, sequencing begins with the addition of a mixture of four nucleotides, each labeled with a reversible terminator and base-specific fluorescent

Table 1. Next generation sequencing studies of lung cancer to date.

First author	Year	Project design	Platform
Campbell (26)	2008	2 lung cancer cell lines WGS	Illumina Genome Analyzer
Pleasance (40)	2010	1 SCLC cell line WGS	SOLiD
Lee (38)	2010	1 NSCLC/normal lung pair WGS	Combinatorial probe anchor ligation
Ju (37)	2011	1 NSCLC/normal lung pair WGS	Illumina HiSeq and Genome Analyzer IIx
Imielinski (41)	2012	183 ACs and matched normal tissue; 159 WES, 23 WES + WGS, 1 WGS	Illumina HiSeq
The Caner Genome Atlas Research Network (42)	2012	Integrated analysis of 178 SCCs and matched germline; 19 WGS, 178 WES	Illumina HiSeq
Govindan (43)	2012	17 WGS; 16 adenocarcinoma and 1 large cell carcinoma	Illumina Genome Analyzer II

WGS, whole genome sequencing; WES, whole exome sequencing; CNV, copy number variation; SNP, single nucleotide polymorphism; AC, adenocarcinoma; SCC, squamous cell carcinoma; SCLC, small cell lung cancer; cPAL, combinatorial probe anchor ligation.

Table 2. Commercial next generation sequencing platforms for human whole genome sequencing.

	Read length/read time	Data processing	Advantages	Disadvantages
Illumina: optical detection of fluorescently labeled nucleotides				
Illumina: HiSeq 2500/2000 (44)[2]	2× 150 bp read length; 11 days	600 Gb; 6 billion reads[1]	Widely used; target difficult genomic regions[3]	Limited scalability; DNA polymerase errors, slow
Illumina: Genome Analyzer IIx (45)[2]	2× 150 bp	85-95 Gb; 640 million reads[1]	Widely used; target difficult genomic regions[3]	Lower data processing capacity than HiSeq
Life Technologies: optical detection of fluorescently labeled nucleotides using sequencing by oligonucleotide ligation and detection				
Life Technologies: SOLiD 5,500xl (46)	Mate paired: 2× 60 bp; Paired-end: 75 bp and 35 bp; fragment: 75 bp; 7 days	10-15 Gb/day;	2-base encoding reduces errors	Short read lengths

[1], paired end reads; [2], with TruSeq v3 kit; [3], TruSeq DNA PCR-Free Kit; bp, base pairs; Mb, megabase; Gb, gigabase; PCR, polymerase chain reaction.

tag, and DNA polymerase. Fragments are extended one nucleotide at a time, and after nucleotide incorporation, laser excitation and image acquisition allow identification of the newly incorporated nucleotide. The fluorescent tag and terminator are then removed, and the next cycle of sequencing proceeds (47). Illumina's range of sequencing platforms available for WGS include the HiSeq and Genome Analyzer IIx. The HiSeq offers both high output and rapid run modes with paired end reads up to 150 bp generating up to 600 Gb data in 11 days. By contrast, the Genome Analyzer IIx can generate up to 95 Gb data by sequencing paired end reads of 150 bp in 14 days from as little as 50 ng DNA (45).

This platform is vulnerable to imperfections in DNA polymerase activity which can result in nucleotide incorporation errors (13), increasing the need for bioinformatic interpretation of sequence results (48). Ongoing developments promise to improve coverage of difficult genomic regions such those rich in repeats, or in guanine (G) and cytosine (C). The TruSeq DNA PCR-Free Kit removes the need for PCR amplification during library preparation, reducing the risk of DNA polymerase-related nucleotide incorporation errors and mis-representation of C-G rich regions (49). In addition, Illumina acquired from Moleculo in 2012 an innovative technology capable of reconstructing short read data into long reads (50).

By contrast, Sequencing by Oligonucleotide Ligation and Detection (SOLiD; Life Technologies) is driven by DNA ligase, not DNA polymerase (51). The DNA library is amplified onto paramagnetic beads immobilized on a solid substrate, a universal primer is hybridized and sequencing commences. Nucleotide octamers, fluorescently tagged at a specific base, are incorporated. After ligation, image acquisition occurs in 4 channels to document the nucleotide present at the identified base. The octamer is then cleaved between bases 5 and 6, and the tag is removed in preparation for the next round of sequencing. Consecutive cycles sequence every 5th base (i.e., 5, 10, 15, 20 etc.); when complete, the DNA is denatured and the process is repeated to sequence a different set of nucleotide positions (i.e., 4, 9, 14, 19 etc.) (13, 24). This system offers the potential for 2-base encoding. This is achieved by incorporating fluorescent tags for adjacent bases in the nucleotide octamer thus sequencing 2 adjacent bases. Each base can then be interrogated twice, allowing detection of

sequencing errors (24).

cPAL is a unique sequencing strategy that is commercially available as a complete sequencing solution. Rolling circle amplification is used to generate nanoballs of genomic DNA that are adsorbed to a solid substrate in a nanoarray. Anchor molecules bind to nucleotide adaptor sequences introduced during library preparation. A fluorescently labeled probe hybridizes to the template DNA and is ligated to the anchor. Sequence is then read by detecting the fluorescence generated at each ligation with great accuracy (52).

In a 2012 landmark paper, Imielinski *et al.*, reported the results of WGS of 36 pulmonary ACs/normal lung pairs coupled with WES of an additional 92 pairs (41). WGS and WES were performed with mean coverage of 69x and 91x, respectively. A number of different techniques were used to identify potential driver mutations. It was found that a combination of 4 bioinformatics strategies optimally identified their ultimate 25 significantly mutated genes. Structural variation was also explored in the 24 matched pairs for which WGS data was available. This was the first publication of WGS data for multiple lung tumours, and postulated a number of novel molecular drivers.

It was closely followed by the TCGA's publication of the first WGS analysis of SCC (42). This study investigated 178 WES and 19 WGS performed on paired SCC/germline DNA. Whole transcriptome profiling was also performed using integrated RNA-sequencing and microarray data. Cases were distributed among gene expression subtype signatures (53) that classified tumours according to functional themes described in terms of gene overexpression relative to the other subtypes.

WGS has also been used, in association with whole transcriptome sequencing, to study the genomic differences between NSCLC in smokers and never smokers (43). Govindan *et al.,* performed paired end WGS sequencing on 17 tumour/normal lung pairs, including 16 ACs and a single large cell carcinoma, with mean haploid coverage of 30x.

These studies suggest that application of NGS technologies to the study of lung cancer genomes will assist us to unlock the mysteries of this disease, and lead to improvements in outcomes for patients. In addition to the identification of a number of new putative therapeutic targets, WGS technology is fuelling advances in metabolomics, epigenomics and transcriptomics. On the other hand, in addition to the high cost currently, a number of technical challenges are yet to be overcome.

Challenges facing lung cancer genomics

NGS technologies offer significant advantages over traditional sequencing techniques for the field of WGS. However, a number of the current platforms' limitations have been identified and are driving technological advances.

Traditional Sanger sequencing is a widely available, but meticulous and time-consuming process, and generally delivers reads of around 1,000 bp with raw accuracy of 99.999% (24). Newer NGS techniques generate large numbers of relatively short reads using techniques which often require amplification by PCR; these techniques are particularly vulnerable to systematic error when applied to comprehensive genomic analysis (12). Long sequences of repeated bases, degraded or damaged DNA and C-G rich regions are particularly problematic (12). Indeed, each sequencing platform has a unique profile of strengths and weaknesses (54).

Short sequence read lengths are characteristic of the early NGS technologies. Unfortunately, these can be troublesome to assemble in a complete genome, and can create a biased aligned read that is insensitive to repeat content and hinders systematic exploration of the genetic basis of disease. Even the "complete reference" is reported to contain up to 350 gaps (12). Regions of the genome rich in repeats, such as those in proximity to centromeres and telomeres, are particularly challenging to map with small reads lengths. Sanger sequencing remains the method of choice for characterising regions where NGS is suboptimal (55).

Genome regions that are rich in C and G can be prone to erroneous replication by NGS technologies that are based on PCR amplification because DNA polymerase ineffectively amplifies these regions. Consequently, these can result in errors in the DNA template delivered to the sequencer, and cause systematic error in the generated reads. Assembly programs are essential for sequence alignment and mapping, and share the potential of sequencing platforms to confound results (56).

Without a doubt, these challenges are driving researchers to achieve exponential advances in experimental and computational technology. A variety of sample preparation technologies are now available to help eliminate the need for template amplification and generate reads of sufficient length to bridge repeat regions. Another technological issue facing WGS bioinformaticians is the intricate complexity of the mutational profile of lung cancer. The computational bottleneck has progressed from generating and aligning sequencing to bioinformatic analytics including discrimination between driver and passenger mutations or background 'noise'.

The extremely large datasets generated by WGS projects have meant that many investigators around the world are working on strategies to optimally store, analyze and handle the "big data". A tiered approach to mutation classification appears of be quite a useful way to classify the large number of mutations found in these genomic interrogations (57). Tier 1 mutations include changes in the coding regions of annotated exons, consensus splice-site regions and RNA genes (including miRNA). Tier 2 features changes in conserved regions of the genome or those with regulatory potential. Tier 3 is characterized by mutations in nonrepetitive part of the genome that are not included in Tier 2. The remainder of the genome is allocated to Tier 4. Common

strategies for the differentiation of driver from passenger mutations vary, but often include the selection of biologically relevant, recurrent mutations. However, the definition of "recurrent" remains inconsistent (41,58).

While these approaches assist in the interpretation of mutation significance, they do not differentiate mutations from background genome variation. The high mutation rate observed in lung cancers disqualifies the bioinformatic assumption of a uniform background mutation rate. However, computational approaches adopted by TCGA in their WGS studies were able to map the structure of variation in the background mutation rate and account for this during data analysis (42).

Hanahan and Weinberg (59,60) proposed a framework of "cancer hallmarks" which is increasingly extended and validated by genomic discoveries. Systematically optimising methods for genome sequencing, alignment, mapping and variant calling will allow us to fully realise benefits of these advances for the study of lung cancer biology.

Future directions

Continuing advances in experimental and bioinformatics technologies are increasing our armamentarium for genomic exploration. Today's platforms, which are essentially limited to replication of DNA's nucleotide sequence, may be replaced by systems able to capture DNA with epigenetic modifications (12). Better analytical tools are needed to facilitate sensitive detection of driver mutations and comprehensive catalogues of genomic data will facilitate dissemination of this knowledge to the scientific, medical and lay communities (61). Other technological advances, such as single cell sequencing, are likely to help us understand the complexity and heterogeneity of the tumour genome (62).

Furthermore, a comprehensive understanding of the intricate relationships among cellular pathways and mutations therein, along with the processes underlying mutagenesis, will guide the development of molecularly targeted therapy. As genome sequencing becomes widely available and interpretable with high accuracy at reduced cost, this information will be increasingly useful for clinicians, and more importantly, their patients.

Acknowledgments

We thank the patients and staff of The Prince Charles Hospital for their involvement and contribution to the TPCH Lung Research Program.

Funding: This work was supported by National Health and Medical Research Council (NHMRC) Project Grants (KF); NHMRC Practitioner Fellowship (KF); NHMRC Career Development Fellowship (IY); NHMRC PhD Scholarship (MD); Cancer Council Queensland (CCQ) Senior Research Fellowship (KF); CCQ PhD Scholarship (MD); CCQ project grants; Health and Medical Research (HMR) project grants; The Prince Charles Hospital Foundation.

Disclosure: The authors declare no conflict of interest.

References

1. Ferlay J, Shin HR, Bray F, et al. GLOBOCAN 2008 v2.0, Cancer Incidence and Mortality Worldwide Lyon, France: International Agency for Research on Cancer, 2010. Available online: http://globocan.iarc.fr
2. Lung cancer: a global scourge. Lancet 2013;382:659.
3. Cancer survival and prevalence in Australia: Cancers diagnosed from 1982-2004. Canberra: Australian Institute of Health and Welfare; Cancer Australia; The Australasian Association of Cancer Registries, July 2008. Report No.: Contract No.: Sept 12.
4. Parikh P, Puri T. Personalized medicine: Lung Cancer leads the way. Indian J Cancer 2013;50:77-9.
5. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol 1975;94:441-8.
6. Maxam AM, Gilbert W. A new method for sequencing DNA. Proc Natl Acad Sci U S A 1977;74:560-4.
7. Reddy EP, Reynolds RK, Santos E, et al. A point mutation is responsible for the acquisition of transforming properties by the T24 human bladder carcinoma oncogene. Nature 1982;300:149-52.
8. Dulbecco R. A turning point in cancer research: sequencing the human genome. Science 1986;231:1055-6.
9. The International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature 2004;431:931-45.
10. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. Nature 2001;409:860-921.
11. Ley TJ, Mardis ER, Ding L, et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature 2008;456:66-72.
12. Marx V. Next-generation sequencing: The genome jigsaw. Nature 2013;501:263-8.
13. Ross JS, Cronin M. Whole cancer genome sequencing by next-generation methods. Am J Clin Pathol 2011;136:527-39.
14. Mardis ER. A decade's perspective on DNA sequencing technology. Nature 2011;470:198-203.
15. Lander ES. Initial impact of the sequencing of the human genome. Nature 2011;470:187-97.
16. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. Nat Rev Genet 2010;11:685-96.
17. Beroukhim R, Mermel CH, Porter D, et al. The landscape of somatic copy-number alteration across human cancers. Nature 2010;463:899-905.
18. Tomlins SA, Rhodes DR, Perner S, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science 2005;310:644-8.
19. Soda M, Choi YL, Enomoto M, et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature 2007;448:561-6.

20. Paez JG, Jänne PA, Lee JC, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science 2004;304:1497-500.
21. Weir BA, Woo MS, Getz G, et al. Characterizing the cancer genome in lung adenocarcinoma. Nature 2007;450:893-8.
22. Kwak EL, Bang YJ, Camidge DR, et al. Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. N Engl J Med 2010;363:1693-703.
23. Daniels M, Goh F, Wright CM, et al. Whole genome sequencing for lung cancer. J Thorac Dis 2012;4:155-63.
24. Shendure J, Ji H. Next-generation DNA sequencing. Nat Biotechnol 2008;26:1135-45.
25. Wheeler DA, Srinivasan M, Egholm M, et al. The complete genome of an individual by massively parallel DNA sequencing. Nature 2008;452:872-6.
26. Campbell PJ, Stephens PJ, Pleasance ED, et al. Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. Nat Genet 2008;40:722-9.
27. Hodges E, Xuan Z, Balija V, et al. Genome-wide in situ exon capture for selective resequencing. Nat Genet 2007;39:1522-7.
28. Sugarbaker DJ, Richards WG, Gordon GJ, et al. Transcriptome sequencing of malignant pleural mesothelioma tumors. Proc Natl Acad Sci U S A 2008;105:3521-6.
29. Morin RD, O'Connor MD, Griffith M, et al. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res 2008;18:610-21.
30. Ordway JM, Budiman MA, Korshunova Y, et al. Identification of novel high-frequency DNA methylation changes in breast cancer. PLoS One 2007;2:e1314.
31. Robertson G, Hirst M, Bainbridge M, et al. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat Methods 2007;4:651-7.
32. Yuille M. Infrastructure vital to genome success. Nature 2011;471:166.
33. Kan Z, Jaiswal BS, Stinson J, et al. Diverse somatic mutation patterns and pathway alterations in human cancers. Nature 2010;466:869-73.
34. Ding L, Getz G, Wheeler DA, et al. Somatic mutations affect key pathways in lung adenocarcinoma. Nature 2008;455:1069-75.
35. Tanaka H, Yanagisawa K, Shinjo K, et al. Lineage-specific dependency of lung adenocarcinomas on the lung development regulator TTF-1. Cancer Res 2007;67:6007-11.
36. Liu P, Morrison C, Wang L, et al. Identification of somatic mutations in non-small cell lung carcinomas using whole-exome sequencing. Carcinogenesis 2012;33:1270-6.
37. Ju YS, Lee WC, Shin JY, et al. A transforming KIF5B and RET gene fusion in lung adenocarcinoma revealed from whole-genome and transcriptome sequencing. Genome Res 2012;22:436-45.
38. Lee W, Jiang Z, Liu J, et al. The mutation spectrum revealed by paired genome sequences from a lung cancer patient. Nature 2010;465:473-7.
39. Wu K, Huang RS, House L, et al. Next-generation sequencing for lung cancer. Future Oncol 2013;9:1323-36.
40. Pleasance ED, Stephens PJ, O'Meara S, et al. A small-cell lung cancer genome with complex signatures of tobacco exposure. Nature 2010;463:184-90.
41. Imielinski M, Berger AH, Hammerman PS, et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. Cell 2012;150:1107-20.
42. Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. Nature 2012;489:519-25.
43. Govindan R, Ding L, Griffith M, et al. Genomic landscape of non-small cell lung cancer in smokers and never-smokers. Cell 2012;150:1121-34.
44. Systems/Sequencing Systems/A sequencer for every need. Every Budget. Every lab. : Illumina; 2013 [cited 2013 Sept 14]. Available online: http://www.illumina.com/systems/sequencing.ilmn
45. Systems/Genome Analyzer IIx/Specifications: Illumina Inc.; 2013 [cited 2013 Sept 14]. Available online: http://www.illumina.com/systems/genome_analyzer_iix/performance_specifications.ilmn
46. 5500 Series Genetic Analysis Systems. In: Technologies L, eds. Carlsbad CA: Life Technologies, 2011.
47. Sequencing Video: The Genome Analyzer: Illumina, 2009.
48. Quail MA, Kozarewa I, Smith F, et al. A large genome center's improvements to the Illumina sequencing system. Nat Methods 2008;5:1005-10.
49. Applications/Sequencing/DNA Sequencing/Whole-Genome Sequencing: Illumina Inc.; 2013 [cited 2013 Sept 14]. Available online: http://www.illumina.com/applications/sequencing/dna_sequencing/whole_genome_sequencing.ilmn
50. Technology/Moleculo Technology: Illumina Inc.; 2013 [cited 2013 Sept 14]. Available online: http://www.illumina.com/technology/moleculo-technology.ilmn
51. Dressman D, Yan H, Traverso G, et al. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. Proc Natl Acad Sci U S A 2003;100:8817-22.
52. Drmanac R, Sparks AB, Callow MJ, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science 2010;327:78-81.
53. Wilkerson MD, Yin X, Hoadley KA, et al. Lung squamous cell carcinoma mRNA expression subtypes are reproducible, clinically important, and correspond to normal cell types. Clin Cancer Res 2010;16:4864-75.
54. Quail MA, Smith M, Coupland P, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics 2012;13:341.
55. Kirby A, Gnirke A, Jaffe DB, et al. Mutations causing medullary cystic kidney disease type 1 lie in a large VNTR in MUC1 missed by massively parallel sequencing. Nat Genet 2013;45:299-303.
56. Kim D, Kim WY, Lee SY, et al. Revising a personal genome by comparing and combining data from two different sequencing platforms. PLoS One 2013;8:e60585.
57. Mardis ER, Ding L, Dooling DJ, et al. Recurring mutations found by sequencing an acute myeloid leukemia genome. N Engl J Med 2009;361:1058-66.
58. Vignot S, Frampton GM, Soria JC, et al. Next-generation sequencing reveals high concordance of recurrent somatic alterations between primary tumor and metastases from patients with non-small-cell lung cancer. J Clin Oncol 2013;31:2167-72.

59. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell 2000;100:57-70.
60. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144:646-74.
61. Green ED, Guyer MS, National Human Genome Research Institute. Charting a course for genomic medicine from base pairs to bedside. Nature 2011;470:204-13.
62. Leary RJ, Sausen M, Diaz LA Jr, et al. Cancer detection using whole-genome sequencing of cell free DNA. Oncotarget 2013;4:1119-20.

Cite this article as: Daniels MG, Bowman RV, Yang IA, Govindan R, Fong KM. An emerging place for lung cancer genomics in 2013. J Thorac Dis 2013;5(S5):S491-S497. doi: 10.3978/j.issn.2072-1439.2013.10.06