Genomics of lung cancer

Saber Hosseinabadi, Ali

Chapter 1

General introduction

Lung cancer

Lung cancer is the main cause of mortality in cancer patients with worldwide incidence of approximately 1.4 million per year[1]. In the Netherlands, 11,669 new cases were diagnosed and 10,544 deaths were registered in 2011 (www.ikc.nl). The low 5-year survival rate (16%) is mainly due to late diagnosis with metastasized disease at time of diagnosis and aggressive behavior of the tumor[2]. Lung cancer is classified into non-small cell lung cancer (NSCLC), which accounts for approximately 85% of all cases, and small cell lung cancer (SCLC)[3]. There are three main NSCLC subtypes, i.e. adenocarcinoma (AC), squamous cell carcinoma (SQCC) and large cell carcinoma (Figure 1)[4]. Computed tomography (CT) and positron emission tomography (PET) are used to detect the location of the primary tumor and presence of metastases. This is followed by histological examination of a biopsy of the tumor taken by bronchoscopy, endoscopic ultrasound guided biopsies or transthoraxic lung biopsies.

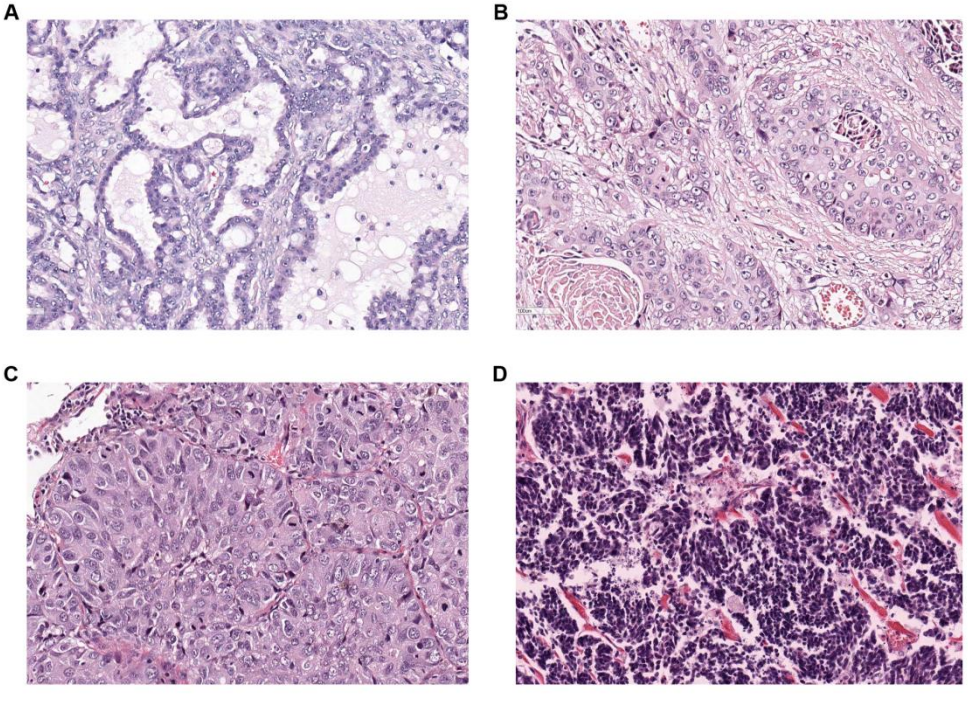

Figure 1: Different subtypes of lung cancer. Formalin fixed paraffin embedded (FFPE) tissues stained by standard hematoxylin and eosin protocol. Magnification: 20x. A) Adeno B) Squamous cell C) Large cell D) Small cell carcinoma.

The staging system of lung cancer is based on the size and location of the primary tumor in the thorax, the number and location of tumor involved lymph nodes and presence of distant metastases (TNM staging system)[5]. In this 7th edition of the lung cancer staging system NSCLC is classified as local, locally advanced or advanced disease. Stage of the disease, performance status, presence of metastasis, histological subtype and presence of activating mutations are the key elements used to select the most optimal treatment strategy for each patient[6]. Surgery is the primary treatment for patients with stage I/II NSCLC[7]. (Neo)adjuvant chemotherapy increases the survival of resectable NSCLC patients[8]. Stage III NSCLC is usually treated with chemoradiotherapy and stage IV NSCLC with systemic treatments such as chemotherapy, targeted therapy or immunotherapy.

There are several risk factors that contribute to the development of lung cancer including cigarette smoking[9], chronic obstructive pulmonary disease (COPD)[10], asbestos exposure[11], air pollution such as fine particulate matter with an aerodynamic diameter of 20.5-10 (PM_{10}), and genetic risk factors[9]. Smoking is the main risk factor[12-13]. A smoking history of more than 40 years increased the relative risks (RR) of SCLC and SQCC with 77 and 22, respectively[14]. However, only 10% of the smokers will develop lung cancer[12]. For AC, besides smoking, air pollution also has a significant contribution. According to a large European study, the lung cancer risk (including AC) was associated with air pollution (Hazard ratio (HR) 1.22 [95% CI 1.03-1.45] per 10 $\mu g/m^3$ increase in PM_{10})[13]. Despite the low HR, the total air pollution impact may be high since everybody is exposed.

Hereditary factors also seem to affect susceptibility of smoking individuals to develop lung cancer. Several single nucleotide polymorphisms (SNPs) such as rs1051730 (*CHRNA3*), rs8034191 (*AGPHD1*) and rs1048943 (*CYP1A1*) on chromosome 15q have been reported to be associated with the risk of lung cancer development[15-16]. There are also some reports showing germline mutations in *EGFR* such as T790M, V843I and P848L in lung cancer patients[17-18]. In addition, a germline *EGFR* T790M mutation has been detected in two family members of a lung AC patient, one of which was diagnosed with lung AC as well[19]. Another case report identified a germline V843I mutation in family members with a history of lung cancer[20].

Gene mutations in lung cancer

Lung cancer is a result of multiple genetic alterations that accumulate during life[21]. In 1982, a human gene with transforming activity was identified in a lung carcinoma cell line. This gene is homologous to the Kirsten Rat Sarcoma virus[22] and was referred to as *KRAS*. *KRAS* is involved in regulating cell proliferation[21] and mutations in *KRAS*, mostly found in codons 12, 13, and 61, result in abnormal activation of the protein[23]. *KRAS* is mutated more frequently in AC of smoking patients (5-40%) than in the other subtypes of lung cancer[21, 23]. The *TP53* gene, first described in 1979[24], has been implicated in DNA repair, cell cycle arrest and apoptosis[25-26]. It is an important tumor suppressor gene that is mutated in approximately 45% of NSCLC[26], most frequent in SQCC[27]. Epidermal growth factor receptor (*EGFR*) shows a high mutation frequency in lung cancer, especially in non-smoking Asiatic females with AC[28]. It is mutated in 5-40% of NSCLC patients depending on the ethnicity[2]. The EGFR protein is involved in cell proliferation, differentiation and survival[29] and activating mutations in the *EGFR* tyrosine kinase domain, i.e. exons 19-21 of the *EGFR* gene[30], triggers a cascade of signaling pathways leading to uncontrolled cell proliferation, inhibition of apoptosis, and metastasis[31].

Next Generation Sequencing

During the last three decades, automated Sanger sequencing has been extensively used both in characterization of the human genome and identification of genetic aberrations in human cancer samples. Despite improvements, technical limitations (cost and time) of Sanger sequencing urged development of improved technologies[32]. Automated Sanger sequencing is called “first generation sequencing” and newer methods like massively parallel sequencing are referred to as “next generation sequencing” (NGS)[33]. NGS is much more cost effective and allows generation of large amount of sequencing data within a shorter period of time. These developments led to increased output of sequencing studies[34]. Analysis of every single nucleotide of the human genome is referred to as whole genome sequencing (WGS), while determining the sequence of all exons, i.e. approximately 1% of the genome, is referred to as whole exome sequencing (WES)[34-35]. WGS is more costly than WES with the same coverage[34], but gives a complete overview of all variations of the entire genome. For example, chromosomal rearrangements and gene fusions can be detected using paired-end WGS[36], but not by WES. On the other hand, the huge amount of data

from WGS and its interpretation is a great challenge both in clinical and research settings[37].

In recent years, massively parallel sequencing has been applied also in RNA sequencing (RNA-seq) studies allowing researchers to investigate the entire transcriptome at high resolution[38-39]. Besides quantifying gene expression, RNA-seq also allows detection of splice variants, identification of transcript start and end sites, novel transcripts and discovery of fusion transcripts[38, 40]. Moreover, using new advances in RNA-seq, novel small and long non-coding RNAs can be identified as well[41].

Tumor evolution and heterogeneity

Cancer is a complex disease that is initiated by acquiring and accumulation of mutations in a single cell through time resulting in a cancerous cell population[42]. It is a genetically heterogeneous disease and several subpopulations coexist in a single tumor[43]. In 1976, cancer has been defined as an evolutionary process which is driven by somatic mutations and selection pressures that help outgrowth of some clones over others[44]. This definition led to two different concepts of evolution, i.e. linear versus branched. In the first model, cancer cells contain all driving mutations that accumulate during evolution, while in the branched model each tumor cell can acquire different mutations and multiple subclones can grow out within a tumor[43]. A branched evolution can result in extensive intra-tumor heterogeneity and this can affect clinical outcome of targeted therapy[45-46]. Intra-tumor heterogeneity has been shown in different cancer types including NSCLC[47-48], clear cell renal carcinoma[49] and pancreatic cancer[50]. Genomic analysis using multiple samples of a patient and analysis at single cell level can improve our understanding about this complex disease and help to further optimize treatment strategies.

Targeted therapy in lung cancer

Utilizing WGS and/or WES provided detailed information about genomic aberrations in lung cancer and this approach has led to identification of potential new targets for therapy[36, 51]. These NGS developments were introduced at the same time frame as the development of novel chemical compounds that can target proteins derived from oncogenic driver mutations. New technologies enhanced generation of structurally adapted compounds to optimally inhibit specific target kinase receptors[52-53]. Discovery of *EGFR* mutations and their

predictive value on tumor response to targeted TKI treatment in lung cancer patients is a revolution in so called “personalized therapy”, i.e. specific drugs targeting these drivers of lung cancer (targeted therapy)[54]. Despite prolonged survival of lung cancer patients with these targeted drugs, tumor cells develop resistance. This indicates the need for novel targeted drugs to treat patients with resistant tumors[55]. This resistance might occur due to intra-tumor heterogeneity and emergence of resistant minor subpopulations via selective pressure applied by treatment[46].

Several promising phase I and II studies are currently being executed that will improve treatment results of lung cancer patients. One can envisage that more driver mutations will be detected in tumor cells and that new or combinations of drugs will enhance treatment outcome in the future[56-58]. Of course, toxicity and induced resistance may still limit their efficiency. Currently there is a growing number of targeted treatments available for NSCLC patients in clinical setting such as gefitinib, erlotinib and afatinib for activating *EGFR* mutations, rociletinib and AZD9291 for resistant *EGFR* T790M mutations, crizotinib, ceritinib, brigatinib and alectinib for EML4-ALK fusion proteins and combination of dabrafenib and trametinib for *BRAF* mutations.

Scope of the study

The aim of this thesis is to investigate genetic changes in lung cancer and to link these changes to tumor evolution and resistance to targeted treatment. In **chapter 2**, we present an overview of known and novel mutated genes in NSCLC, and provide an overview of clinically available targeted therapies including known resistance mechanisms. In **chapter 3**, we examined the association between presence and type of *EGFR* and *KRAS* mutations in NSCLC patients with and without COPD. In **chapter 4**, we applied a combination of NGS techniques such as WES and single cell-WGS on normal, tumor and multiple metastases samples from five lung cancer patients to explore clonal evolution of tumor cells and tumor heterogeneity in different subtypes of metastatic lung cancer. In **chapter 5A**, we performed a pilot RNA-sequencing study to optimize the approach to distinguish between true fusion genes and false positives. In **chapter 5B**, we performed RNA-seq on frozen resistant tumor samples of three patients with a known *EML4-ALK* translocation, who progressed on crizotinib. The goal was to identify new fusion transcripts that may play a role in ALK-TKI resistance. At the end, we summarize the main findings and discuss future perspectives (**chapter 6**).

References

1. 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;22:436-445.
2. Nana-Sinkam SP, Powell CA. Molecular Biology of Lung Cancer: Diagnosis and Management of Lung Cancer, 3rd ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest* 2013;143:e30S-39S.
3. Francis H, Solomon B. The current status of targeted therapy for non-small cell lung cancer. *Intern Med J* 2010;40:611-618.
4. Travis WD. Pathology of lung cancer. *Clin Chest Med* 2011;32:669-692.
5. Goldstraw P. Staging Manual in Thoracic Oncology. In: Cancer IAftSoL, ed.Denver, CO: Editorial Rx Press; 2009.
6. Socinski MA, Evans T, Gettinger S, et al. Treatment of stage IV non-small cell lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest* 2013;143:e341S-368S.
7. van der Drift MA, Karim-Kos HE, Siesling S, et al. Progress in standard of care therapy and modest survival benefits in the treatment of non-small cell lung cancer patients in the Netherlands in the last 20 years. *J Thorac Oncol* 2012;7:291-298.
8. Lim E, Harris G, Patel A, et al. Preoperative versus postoperative chemotherapy in patients with resectable non-small cell lung cancer: systematic review and indirect comparison meta-analysis of randomized trials. *J Thorac Oncol* 2009;4:1380-1388.
9. Molina JR, Yang P, Cassivi SD, et al. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. *Mayo Clin Proc* 2008;83:584-594.
10. Collins LG, Haines C, Perkel R, et al. Lung cancer: diagnosis and management. *Am Fam Physician* 2007;75:56-63.
11. O'Reilly KM, McLaughlin AM, Beckett WS, et al. Asbestos-related lung disease. *Am Fam Physician* 2007;75:683-688.
12. Chin LJ, Ratner E, Leng S, et al. A SNP in a let-7 microRNA complementary site in the KRAS 3' untranslated region increases non-small cell lung cancer risk. *Cancer Res* 2008;68:8535-8540.
13. Raaschou-Nielsen O, Andersen ZJ, Beelen R, et al. Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE). *Lancet Oncol* 2013;14:813-822.
14. Kenfield SA, Wei EK, Stampfer MJ, et al. Comparison of aspects of smoking among the four histological types of lung cancer. *Tob Control* 2008;17:198-204.
15. Amos CI, Wu X, Broderick P, et al. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat Genet* 2008;40:616-622.
16. Larsen JE, Colosimo ML, Yang IA, et al. Risk of non-small cell lung cancer and the cytochrome P4501A1 Ile462Val polymorphism. *Cancer Causes Control* 2005;16:579-585.
17. Oxnard GR, Miller VA, Robson ME, et al. Screening for germline EGFR T790M mutations through lung cancer genotyping. *J Thorac Oncol* 2012;7:1049-1052.
18. Prim N, Legrain M, Guerin E, et al. Germ-line exon 21 EGFR mutations, V843I and P848L, in nonsmall cell lung cancer patients. *Eur Respir Rev* 2014;23:390-392.
19. Yu HA, Arcila ME, Harlan Fleischut M, et al. Germline EGFR T790M mutation found in multiple members of a familial cohort. *J Thorac Oncol* 2014;9:554-558.
20. Ohtsuka K, Ohnishi H, Kurai D, et al. Familial lung adenocarcinoma caused by the EGFR V843I germ-line mutation. *J Clin Oncol* 2011;29:e191-192.
21. Okudela K, Woo T, Kitamura H. KRAS gene mutations in lung cancer: particulars established and issues unresolved. *Pathol Int* 2010;60:651-660.
22. Der CJ, Krontiris TG, Cooper GM. Transforming genes of human bladder and lung carcinoma cell lines are homologous to the ras genes of Harvey and Kirsten sarcoma viruses. *Proc Natl Acad Sci U S A* 1982;79:3637-3640.

23. Riely GJ, Marks J, Pao W. KRAS mutations in non-small cell lung cancer. *Proc Am Thorac Soc* 2009;6:201-205.
24. Linzer DI, Levine AJ. Characterization of a 54K dalton cellular SV40 tumor antigen present in SV40-transformed cells and uninfected embryonal carcinoma cells. *Cell* 1979;17:43-52.
25. Liloglou T, Ross H, Prime W, et al. p53 gene aberrations in non-small-cell lung carcinomas from a smoking population. *Br J Cancer* 1997;75:1119-1124.
26. Mogi A, Kuwano H. TP53 mutations in nonsmall cell lung cancer. *J Biomed Biotechnol* 2011;2011:583929.
27. Subramanian J, Govindan R. Molecular genetics of lung cancer in people who have never smoked. *Lancet Oncol* 2008;9:676-682.
28. Mitsudomi T, Kosaka T, Yatabe Y. Biological and clinical implications of EGFR mutations in lung cancer. *Int J Clin Oncol* 2006;11:190-198.
29. Herbst RS. Review of epidermal growth factor receptor biology. *Int J Radiat Oncol Biol Phys* 2004;59:21-26.
30. Ma ES, Ng WK, Wong CL. EGFR gene mutation study in cytology specimens. *Acta Cytol* 2012;56:661-668.
31. Piperdi B, Perez-Soler R. Role of erlotinib in the treatment of non-small cell lung cancer: clinical outcomes in wild-type epidermal growth factor receptor patients. *Drugs* 2012;72 Suppl 1:11-19.
32. Metzker ML. Emerging technologies in DNA sequencing. *Genome Res* 2005;15:1767-1776.
33. Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet* 2010;11:31-46.
34. Bick D, Dimmock D. Whole exome and whole genome sequencing. *Curr Opin Pediatr* 2011;23:594-600.
35. Wiesner T, Obenauf AC, Murali R, et al. Germline mutations in BAP1 predispose to melanocytic tumors. *Nat Genet* 2011;43:1018-1021.
36. Imielinski M, Berger AH, Hammerman PS, et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* 2012;150:1107-1120.
37. Cordero P, Ashley EA. Whole-genome sequencing in personalized therapeutics. *Clin Pharmacol Ther* 2012;91:1001-1009.
38. Nagalakshmi U, Waern K, Snyder M. RNA-Seq: a method for comprehensive transcriptome analysis. *Curr Protoc Mol Biol* 2010;Chapter 4:Unit 4 11 11-13.
39. Ozsolak F, Milos PM. RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet* 2011;12:87-98.
40. Tsuchihara K, Suzuki Y, Wakaguri H, et al. Massive transcriptional start site analysis of human genes in hypoxia cells. *Nucleic Acids Res* 2009;37:2249-2263.
41. Ghildiyal M, Zamore PD. Small silencing RNAs: an expanding universe. *Nat Rev Genet* 2009;10:94-108.
42. Yates LR, Campbell PJ. Evolution of the cancer genome. *Nat Rev Genet* 2012;13:795-806.
43. de Bruin EC, Taylor TB, Swanton C. Intra-tumor heterogeneity: lessons from microbial evolution and clinical implications. *Genome Med* 2013;5:101.
44. Nowell PC. The clonal evolution of tumor cell populations. *Science* 1976;194:23-28.
45. Swanton C. Cancer evolution: the final frontier of precision medicine? *Ann Oncol* 2014;25:549-551.
46. Bedard PL, Hansen AR, Ratain MJ, et al. Tumour heterogeneity in the clinic. *Nature* 2013;501:355-364.
47. de Bruin EC, McGranahan N, Mitter R, et al. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science* 2014;346:251-256.
48. Zhang J, Fujimoto J, Wedge DC, et al. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science* 2014;346:256-259.
49. Gerlinger M, Rowan AJ, Horswell S, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* 2012;366:883-892.

50. Yachida S, Jones S, Bozic I, et al. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* 2010;467:1114-1117.
51. Liu P, Morrison C, Wang L, et al. Identification of somatic mutations in non-small cell lung carcinomas using whole-exome sequencing. *Carcinogenesis* 2012;33:1270-1276.
52. Abidin AZ, Garassino MC, Califano R, et al. Targeted therapies in small cell lung cancer: a review. *Ther Adv Med Oncol* 2010;2:25-37.
53. Dempke WC, Suto T, Reck M. Targeted therapies for non-small cell lung cancer. *Lung Cancer* 2010;67:257-274.
54. Janku F, Garrido-Laguna I, Petruzelka LB, et al. Novel therapeutic targets in non-small cell lung cancer. *J Thorac Oncol* 2011;6:1601-1612.
55. Shaw AT, Engelman JA. Ceritinib in ALK-rearranged non-small-cell lung cancer. *N Engl J Med* 2014;370:2537-2539.
56. Savas P, Hughes B, Solomon B. Targeted therapy in lung cancer: IPASS and beyond, keeping abreast of the explosion of targeted therapies for lung cancer. *J Thorac Dis* 2013;5:S579-S592.
57. Shaw AT, Ou SH, Bang YJ, et al. Crizotinib in ROS1-rearranged non-small-cell lung cancer. *N Engl J Med* 2014;371:1963-1971.
58. Planchard D, Mazieres J, Riely GJ, et al. Interim results of phase II study BRF113928 of dabrafenib in BRAF V600E mutation-positive non-small cell lung cancer (NSCLC) patients. *J Clin Oncol* 2013;31:abstr 8009.