

Targeted Hot Spot Sequencing of Uzbek Lung Cancer Patients

Z Mirakbarova¹ and Sh Turdikulova²

¹Institute of Biophysics and Biochemistry at the National University of Uzbekistan named after Mirzo Ulugbek.

²Centre for Advanced Technologies under the Ministry of Innovative Development of the Republic of Uzbekistan.

*Corresponding author E-mail: zebyniso@gmail.com

<https://dx.doi.org/10.13005/bpj/2138>

(Received: 03 December 2020; accepted: 04 March 2021)

Carcinoma of lung is a widespread multifactorial disease with a poor prognosis. Uzbek lung cancer patients screening for prevalent mutations, associated with cancer development by targeted next generation sequencing was the main objective of current research. Preparation of sequencing libraries was conducted using the Illumina Cancer Hotspot Panel v2 covering mutational hot spot regions of 50 cancer genes, from DNA extracted from formalin-fixed paraffin-embedded tumor samples of 10 patients. Sequencing was performed with the MiSeq Next Generation Sequencing machine (NGS). Hot spot mutations data analysis revealed clinically benign mutations in all studied samples, with 70% having 3 or more mutations stored in the COSMIC database. According to ClinVar FGFR3 gene was associated with Crouzon syndrome with acanthosis nigricans, PDGFRA, EGFR and TP53 gene mutations were correlated with Idiopathic hypereosinophilic syndrome, Lung cancer and Glioma susceptibility respectively. KIT and KDR gene mutations were, respectively, associated with Gastrointestinal stromal tumor and Hemangioma, capillary infantile. Synonymous variants in RET and HRAS genes, were associated with congenital central hypoventilation and epidermal nevus syndrome respectively. EGFR gene mutations occurred exclusively in patients with adenocarcinoma G2, whereas MET and BRAF mutations were identified in bronchioalveolar cancer samples. In conclusion, particular polymorphisms and mutations affect treatment response and the toxicity level among patients with lung cancer, undergoing chemotherapy. Variety of SNPs were reported to have impact on disease outcomes of those patients, therefore they should be validated in an independent population, prior to being adopted for pre-personalised therapy screening.

Keywords: Uzbek Lung cancer, SNPs, Therapy screening.

Majority of cancer deaths worldwide both in men and women occur as an outcome of lung carcinoma¹. Non-small cell lung cancer (NSCLC) represents the from 80 to 85% of diagnosed lung cancer cases^{2,3}. Despite having the same aetiology, types of lung cancer can vary on molecular level: epigenetic changes, protein expression, mutations

etc, which are reported as diagnostic markers and can influence the therapy outcome. Although platinum based doublet therapy has been used as the standard treatment for late stages for many years, accumulating evidence suggests that lung non-small cell lung cancer patients with activated EGFR, KRAS, BRAF and some other genes

represent positive outcome, while treated with targeted therapy methods^{4,5}. Thus, it is important to identify the mutation status of NSCLC patients prior to treatment tactic selection. Advanced stages of lung cancer are currently unpreventable. Clinical assays define using molecular subtype, clinical and histological features of cancer as a basis of therapy methods¹. Large-scale investigations are proving lung cancer heterogeneity and its development into progressive disease^{2,3} and, paired with the dramatical advancement of therapy methods, illustrate the need for more complex diagnostics in cancer management⁴.

Diagnostics based on next generation sequencing (NGS) are creating an opportunity for researchers and clinicians to identify specific modification occurring in tumor cells. Basically, identifying pathologic transformation on gene and protein level enables clinicians to diagnose oncologic diseases on early stages as well as predict therapy responses more exactly. At the meantime current methods comprise of realtime polymerase chain reaction (PCR) and Sanger sequencing, where each selected region of studied gene in a given sample is studied on separate basis. Sanger sequencing method has yet several challenges to overcome while detecting somatic mutations, associated with cancer development. These mutations are detected at low prevalence, due to specific localisations features of the tumor or the variations in mixture rates of normal and tumor tissues in studied samples, therefore, require reliable DNA library preparation and highly sensitive methods for successful detection. To accomplish this, highly sensitive next-generation sequencing, with its remarkable throughput, is dramatically entering the clinical testing area. However to date, limited use of this method in clinical diagnostics is observed⁵.

Primary lung, colon, pancreatic and other cancers were shown to be conveniently detected by the AmpliSeq Cancer Hotspot Panel⁷,

MATERIALS AND METHODS

Subjects

The subjects were recruited from people, undergoing treatment in the Department of Thoracic surgery National Cancer Center of Uzbekistan. 20 people were asked to participate

and 12 accepted. Subjects with the history of small cell carcinoma and tuberculosis were excluded from the study. Finally, the baseline data of 10 Uzbek adults (7 males and 3 females), 38-66 years of age were analyzed in this study. Written informed consent in a form approved by the Ethics Committee under the Ministry of Health of the Republic of Uzbekistan was obtained from all subjects for participation in the current study. Genotyping and data analyses were performed according to lab protocols approved by the same Ethics Committee. Patients with histologically confirmed lung adenocarcinomas and squamous cell carcinoma of different stages were eligible for enrolment in his study. All patients provided clinic-pathological as well as demographic data, including age at diagnosis, gender, nationality, profession, workplace, place of birth, concomitant diseases history, tumor stage, prior therapy, family history. Formalin fixed paraffin embedded tumor tissue specimens, previously studied for adequacy via histopathological assessment were included in this study.

DNA extraction

Tumor cells from archive FFPE tissue samples were initially micro dissected manually. DNA from 10 FFPE tissue slides was extracted according to ReliaPrep™ FFPE gDNA Miniprep System Promega protocol at the laboratory of Biotechnology Centre for advanced Technologies under the Ministry of Innovative development of the Republic of Uzbekistan. Assessment of DNA quality and quantity was conducted using a Shimadzu BioSpec-nano Micro-volume UV-Vis Spectrophotometer (Japan) and Qubit dsDNA HS Assay Kit and a Qubit 2.0 fluorometer. All DNA samples were aliquoted and stored at +20 °C until analysis.

Library Preparation and Quality Control

Illumina AmpliSeq Cancer Hotspot Panel v2, containing probes to generate 207 amplicons from 50 genes, associated with cancer, to identify around 2800 mutations in hotspot regions was used to prepare sequencing amplicon libraries. The amplicons of target regions were generated using 207 oligonucleotide pairs, during thermocycling in GeneAmp® PCR-amplifier Verity AB. Further target amplicons were partially digested to form and phosphorylate sticky ends to facilitate bar-coded adapter ligation. The amplicons were

flanked with index sequences, during multiplex PCR to create unique index combinations for dual index sequencing. After library clean-up, quality of library was assessed on a Qubit dsDNA HS Assay Kit and a Qubit 2.0 fluorometer. Afterwards, libraries were reamplified to ensure required quantity for Illumina sequencing with further two round clean-up. In order to provide optimum cluster density on the flow cell the library was checked on Qubit 2.0 fluorometer, normalized, diluted to 2.8ng/μl, before pooling equal volumes for final sequencing library generation, after which libraries were quantified and diluted to starting concentration. MiSeq instrument was used for pooled library sequencing applying a 2 150 paired-end sequencing design

Data analysis

Raw data was processed in FASTQ Generation Version: 1.0.0 within Base Space to generate fastqfile. The IGV16 (Broad Institute, Cambridge, Massachusetts) was applied to align paired-end fastq format raw reads to the hg19 reference genome and produce VCF files. Illumina Variant Studio version 1.0 (Illumina) vs Annovar were used for variants annotation.

Variant effect predictor (VEP) on ENSEMBLE and CRAVAT version 4.3 (<http://hg19.cravat.us/CRAVAT/>) was implemented for identification of variants from aligned reads.

Variants representing a global minor allele frequency more than 0.1% were excluded from further studies, being considered as common SNPs as well as non-coding variants.

RESULTS

Molecular-genetic analysis was finalised, and reports formed in 30 days, after receiving agreement consent from patients to be enrolled in this study. Variant Effect Predictor was used to perform annotation, by classifying variants into 8 various classes (“Frame Shift Insertion”, “Missense Mutation”, “Frame Shift Deletion”, “In Frame Insertion”, “In Frame Deletion”, “Nonsense Mutation” and “Splice Site”). Variants having allele frequency less than 0.05% and SIFT score <0.05 (Deleterious) were retained. All patients, included in the study had mutations in genes, responsible explicitly or implicitly for carcinogenesis and drug sensitivity and involved in Hot Spot Cancer v2 panel. Overall mutations in 18 genes, 11 of which were reported as oncogenes and 7 as tumor suppressor genes were identified. All the mutations identified except 5 appeared to be initiating amino acid change in coded proteins.

Hot-spot mutations were identified in below mentioned genes during data analysis: ALK, APC,

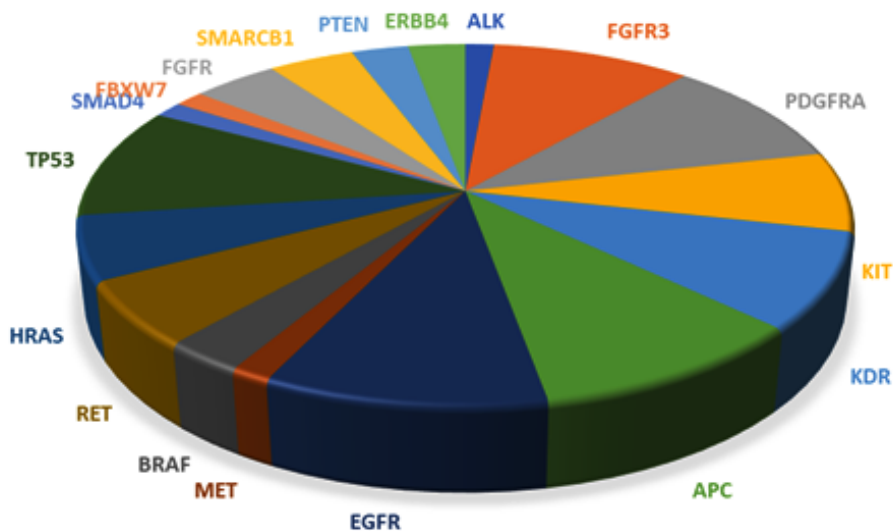


Fig. 1. Illustration of genes with hot spot mutations in Uzbek NSCLC patients

Table 1. Single-Nucleotide Polymorphisms inAmpliSeq Cancer Hotspot Panel with MAF greater than or equal to 0.05%

Gene	rs number	cDNA change	Aminoacid change	RefNumber	MAF %*
APC	rs41115	c.4479G>A	T1493T	NM_000038	0.66
EGFR	rs1050171	c.2361G>A	Q787Q	NM_005228	0.43
FGFR3	rs7688609	c.1953A>G	T653T	NM_000142	0.95
HRAS	rs12628	c.81T>C	H27H	NM_005343	0.29
KDR	rs1870377	c.1416A>T	Q472H	NM_002253	0.21
KIT	rs3822214	c.1621A>C	M541L	NM_000222	0.06
PDGFRA	rs1873778	c.1701A>C	P567P	NM_006206	0.95
RET	rs1800861	c.2307T>G	L769L	NM_020975	0.71
TP53	rs1042522	c.215C>G	P72R	NM_000546	0.54

*Minor allele frequency (MAF) according to the 1000 Genomes database

BRAF, ERBB4, EGFR, FGFR1, FBXW7, HRAS, FGFR3, KIT, KDR, MET, PDGFRA, PTEN, RET, SMAD4, SMARCB1, TP53, illustrated in COSMIC database.

Identified SNPs population minor allele frequency range was from 0,06% to 0,95% according to the 1000 Genomes database.

All participants underwent adjuvant chemotherapy as standard treatment method prior to analysis. Mutations in PDGFRA, FGFR3, APC, EGFR and TP53 genes being the most frequently were revealed in 70% of studied samples. According to ClinVar database rs7688609 in FGFR3 gene was associated with Crouzon syndrome with acanthosis nigricans, rs1873778 in PDGFRA gene, rs1050171 in EGFR gene and rs1042522 in TP53 gene were associated with Idiopathic hypereosinophilic syndrome, Lung cancer and Glioma susceptibility respectively. The second and third most common mutations with 50% and 60% were reported to be missense variants rs3822214 and rs1870377 in KIT and KDR genes respectively, associated with Gastrointestinal stromal tumor and Hemangioma, capillary infantile. Synonymous variants in rs1800861 in RET and rs12628 in HRAS genes, giving association with Congenital central hypoventilation and Epidermal nevus syndrome respectively came fourth with the index of 40%. Worth noting that each participant of this study had at least one variant, associated with lung cancer development, in addition, involved in drug metabolism.

DISCUSSION

Lung cancer is multifactorial disease, occurring in majority cases because of genetic translocations in organism. Each type of lung cancer has specific treatment methods, based on histological subtype, stage, aetiology. Over the past decade, targeted therapy of lung cancer has become one of the standard methods of treatment for non-small cell lung cancer. Mutational spectra identified during the research was found to be potential targets in the implementation of targeted therapy. Validation of identified EGFR mutations was conducted by classic mutant enriched PCR method, showing 96% concordance, proving trustfulness of Hot Spot Cancer Sequencing results. Study of c.2361A>G mutation in patients with colorectal, breast and lung cancer revealed better clinical outcome after anti-EGFR treatment³. The frequency of missense mutation in EGFR gene c.2361A>G, leading glycine to histidine change in amino acid strain of coded protein, was rather high than reported in other studies, researching European (4,9-15%)^{2,6,7,12,15} and Asian lung cancer patients (3,8-49,1%), however no correlation was identified between the mutation status and patients' characteristics of age, gender, smoking history^{16, 17}. Although, several studies were conducted showing the role of exon 20 EGFR mutation in chemotherapy response, it wasn't included in PharmGKB database up to date. CC genotype in rs1042522 of TP53 gene was reported as one of the

risk factors, leading to lung cancer development in Bangladeshi population⁴. According to PharmGKB, rs1042522 is associated with decreased response to capecitabine and paclitaxel, in patients with stomach neoplasms, while patients with ovarian neoplasms, carrying this mutation, suffered from more severe form of Neutropenia while being treated with cyclophosphamide and cisplatin. According to 1000Genomes, studies conducted in aggregated populations showed the frequency of TP53 mutations from 30 to 68%⁹. In our study the frequency of TP53 mutation was higher than reported elsewhere, although further studies are yet to be conducted. Regarding the study of rs1873778 in PDGFRA gene and rs7688609 in FGFR3, which was present in 70% of studied samples, association with cervical adenosquamous carcinoma and colorectal cancer to add withafatinib resistance in non-small cell lung cancer patients was reported¹. KIT gene mutation rs3822214 identified in studied samples was shown to have clinical significance in partial albinism, gastrointestinal stromal tumor, hereditary cancer-predisposing syndrome, mastocytosis, chronic myelogenous leukemia¹⁴. According to results of multivariate analysis, conducted in 163 patients with various types of cancer, major homozygous rs1870377 genotype of KDR had negative effects on both TTF and OS, when compared to the effects of the heterozygous or minor homozygous genotypes, while being treated with sunitinib as an complementary therapy¹⁰, in addition kinase insert domain receptor polymorphisms are associated with the increased protein concentration in serum plasma and correlate with an increased risk of stomach, lung and breast neoplasms progression^{8,11,13}.

ACKNOWLEDGEMENTS

We thank Professor M. Tillyashakhov Director of the Republic Specialized Scientific Practical Medical Centre of Oncology and Radiology and his vice-director Professor A.A. Yusupbekov and the staff of the departments of Thoracic surgery and Pathomorphology for clinical data and sample collection support and D. Dalimov the head of the laboratory of Biotechnology Centre for Advanced Technologies under the Ministry of Innovative Development of the Republic of Uzbekistan and the staff of the laboratory for practical support in

organizing the initial sample preparation works. Authors are grateful to the Ministry of Innovative Development of the Republic of Uzbekistan for providing funding opportunity to conduct this research.

Availability of data and materials

The datasets during and/or analysed during the current study are available from the corresponding author on reasonable request.

Consent for publication

The manuscript doesn't include any individual person's data.

Ethics approval and consent to participate

Formalin fixed paraffin embedded specimens and clinical information's were collected after informed consent from patients.

Conflict of interest

Authors of the manuscript Mirakbarova Z and Turdikulova Sh don't have a potential conflict of interest to declare. The funders of current research had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript and have no potential conflict of interest to declare.

Funding source

This research was supported by Scientific Research Projects PZ-2014-0915210258 and MU-PZ-20171025473, funded by the Ministry of Innovative Development of the Republic of Uzbekistan. The funders had no role in study design, data collection.

Author's contributions

ZM: Conception and design of the work. Acquisition and interpretation of data. Drafting manuscript, FFPE sequencing. Statistical analysis of data. ShU: Design of work, approval of the final version of the manuscript.

CONCLUSION

In conclusion, particular polymorphisms and mutations affect treatment response and the toxicity level among patients with lung cancer, undergoing chemotherapy. Variety of SNPs were reported to have impact on disease outcomes of those patients therefore, they should be validated in an independent population, prior to being adopted for pre-personalised therapy screening.

REFERENCES

1. Adhemar Longatto-Filho I, Pinheiro C, Martinho O, Marise A, Moreira R, Luiz F, Ribeiro J, Queiroz G, Schmitt F, Baltazar F, Rui M Reis Molecular Characterization of EGFR, PDGFRA and VEGFR2 in Cervical Adenosquamous Carcinoma. *BMC Cancer*; **9**:212 (2009). doi: 10.1186/1471-2407-9-212.
2. Boch C, Kollmeier J, Roth A, *et al.*, The frequency of EGFR and KRAS mutations in non-small cell lung cancer (NSCLC): routine screening data for central Europe from a cohort study, *BMJ Open* **3**: (2013). <http://dx.doi.org/10.1136/bmjopen.2013-002560>
3. Bonin, S., Donada, M., Bussolati, G., Nardon, E., Pichler, M., Chiaravalli, A. M., ... Stanta, G. The EGFR synonymous polymorphism rs1050171 predicts responsiveness to anti-EGFR therapy in metastatic colorectal cancer patients. *European Journal of Cancer*, **61**: S181–S182 (2016). doi:10.1016/s0959-8049(16)61643-x
4. Chowdhury MK, Moniruzzaman M, Emran AA, Mostafa MG, Kuddus RH, Uddin MA. TP53 Codon 72 Polymorphisms and Lung Cancer Risk in the Bangladeshi Population. *Asian Pac J Cancer Prev*; **16**:3493–98 (2015). doi: 10.7314/APJCP.2015.16.8.3493.
5. Dong L., Wang W., Li A., Kansal R., Chen Y., Chen H., Li X. Clinical next generation sequencing for precision medicine in cancer (2015).
6. Gahr S, Stoehr R, Geissinger E, *et al.*, EGFR mutational status in a large series of Caucasian European NSCLC patients: data from daily practice, *Br. J. Cancer*, **109**: 1821–1828 (2013).
7. Hogdall, P. Clementsen, *et al.*, The prevalence of EGFR mutations in non-small cell lung cancer in an unselected Caucasian population, *APMIS*, **123**: 108–115 (2015).
8. Langsenlehner, U., Hofmann, G., Renner, W., Gerger, A., Krenn-Pilko, S., Thurner, E.M., Krippel, P. and Langsenlehner, T. Association of Vascular Endothelial Growth Factor: A Gene Polymorphisms and Haplotypes with Breast Cancer Metastases. *Acta Oncologica*, **54**: 368-376 (2015).
9. Liu X, Lin X, Wang C, *et al.*, Association between smoking and p53 mutation in lung cancer: a meta-analysis, *Clin. Oncol. (R. Coll. Radiol.)* **26**: 18–24 (2014).
10. Maeng C, Yi J, Lee J, Hong J, Choi M, Jung H, *et al.* Effects of single nucleotide polymorphisms on treatment outcomes and toxicity in patients treated with sunitinib. *Anticancer research*; **33**(10):4619–26 (2013). Epub 2013/10/15. PMID:24123039.
11. N Naykoo, Dil-Afroz, R Rasool, S Shah, AAhangar, I Bhat, I Qasim, M Siddiqi, Z Shah. Single Nucleotide Polymorphisms, Haplotype Association and Tumour Expression of the Vascular Endothelial Growth Factor (VEGF) Gene with Lung Carcinoma. *Gene*, **608**: 95-102.
12. Ramlau, T. Cufer, P. Berzinec, *et al.*, Epidermal growth factor receptor mutation-positive non-small-cell lung cancer in the real-world setting in central Europe: the INSIGHT study, *J. Thorac. Oncol.* **10**: 1370–1374 (2015).
13. Sa-Nguanraksa, Kooptiwut D, Chuangsuwanich S, Pongpruttipan T, Malasit T, Charoenrat P and O, Vascular Endothelial Growth Factor Polymorphisms Affect Gene Expression and Tumor Aggressiveness in Patients with Breast Cancer. *Molecular Medicine Reports*, **9**: 1044-1048 (2014).
14. Siqueira E, de Sousa S, Franca J, *et al.* Targeted next-generation sequencing of glandular odontogenic cyst: a preliminary study. *Oral Surg Oral Med Oral Pathol Oral Radiol.*; **124**: 490-494 (2017).
15. Smits A, Kummer J, Hinrichs J, *et al.*, EGFR and KRAS mutations in lung carcinomas in the Dutch population: increased EGFR mutation frequency in malignant pleural effusion of lung adenocarcinoma, *Cell Oncol.* **35**: 189–196. [18] (2012).
16. Xu S, Duan Y, Lou L, *et al.* Exploring the impact of EGFR T790M neighboring SNPs on ARMS-based T790M mutation assay. *Oncology Letters*; **12**: 4238–44 (2016).
17. Y Wang, W Bao, H Shi, Y Zhang. Epidermal Growth Factor Receptor Exon 20 Mutation Increased in Post-Chemotherapy Patients with Non-Small Cell Lung Cancer Detected with Patients' Blood Samples. *Translational Oncology*, **6**: 4: P. 504–510 (2013).