Molecular Detection of *EGFR* Mutation in Lung Cancer Using Bioinformatics Database with PCR Technique

Ayad Mohammed Ali* and Kameran Mohammed Ali**

*Biotechnology Specialist- Kalar General Hospital- Garmian General Directorate of Health- Ministry of Health- Kalar- Sulaimani.

** MLT Dept.- Kalar Technical Institute- Sulaimani Polytechnic University- Kalar- Sulaimani.

Received 3 July 2016 ; Accepted 24 August 2016

Abstract

The aim of this study was to design a new model of polymerase chain reaction (PCR) based diagnostic method using mouse genome for one rare epidermal growth factor receptor (*EGFR*) gene mutations with diagnostic and prognostic values in lung cancer (this mutation are not covered by commercial test kits). From the systematic database search, one rare mutations identified which is within the tyrosine kinase (TK) domain of *EGFR*, with an insertion mutation in exon 20 namely A763_Y764insFQEA. For designing primers to detect the targeted regions surrounding the rare mutations, Primer3 Plus was carried out to design two sets of PCR primers for exon 20 of mouse *EGFR* gene. UCSC (University of California Santa Cruz) *in silico* PCR testing along with BLASTN search were used for primer specificity in terms of predicted target location (chromosome 11 for exon 20) and predicted amplicon sizes (276 bp for exon 20). PCR was partially optimised for the exon 20 with the presence of expected amplicon bands, along with unspecific and primer-dimer bands. The amplicon was sequenced and revealed the presence of the mutation. The mutation of interest selected was A763_Y764insFQEA (insertion of phenylalanine, glutamine, glutamic acid and alanine in between codon 763 and 764) in exon 20, the mutation was in the kinase domain of *EGFR*. The Identification and Analysis for rare

mutations via this way is sensitive, simple, accurate and inexpensive technique. It is used as genetic markers for allelic and mutational sequence variation.

Keywords: Lung cancer, EGFR, mutation, Exon 20.

الكشف الجزيئي لطفرة *EGFR* في سرطان الرئة باستخدام قاعدة بيانات معلومات حيوية مع تقنية تفاعل البلمرة المتسلسل

أياد محمد علي* و كامران محمد علي**

*مستشفى كلار العام- مديرية صحة طرميان- وزارة الصحة

**قسم التحليلات المرضية- المعهد التقني في كلار- جامعة بوليتكنيك السليمانية

الخلاصـــة

هدفت هذه الدراسة هو تصميم نموذج جديد لتقنية التشخيص المعتمد على تفاعل البلمرة المتسلسل (PCR) باستخدام جينوم الفأر لاحد الطفرات الجينية النادرة لمستقبل عامل نمو الادمة (*EGFR*) مع القيم التشخيصية والنذيرية في سرطان الرئة (هذه الطفرة لم تغطى باستخدام عدد تجارية). من البحث في نظام قاعدة البيانات، حددت واحدة من الطفرات النادرة التي تقع ضمن نطاق انزيم التيروسين كاينيز (TK) من *EGFR* كطفرة مدرجة في اكسون 20 وهي A763_Y764insFQEA.لتصميم الشعيلة للكشف عن المناطق المستهدفة المحيطة الطفرات النادرة نفذت Primer3 Plus لتصميم مجموعتين من شعيلات لاكسون 20 من جين *EGFR* للفار .استخدمت تقنية جامعة كاليفورنيا سانت كروز (UCSC) في silico PCR مع البحث BLASTN لخصوصية الشعيلات من حيث تنبأ الموقع المستهدف (الكروموسوم 11 لاكسون 20) وتوقع حجم الامبليكون بـ (276 لاكسون 20) و حسنت جزئيا تقنية PCR للاكسون 20 مع وجود حزم امبليكون متوقعة، جنبا إلى جنب مع الخطوط غير محددة. حيث الامبليكون تسلسل وكشف وجود الطفرة. وكانت الطفرة المهم اختيار ها كانت A763_Y764insFQEA المدرج من الفذيل الأنين، الجلوتامين، حمض الجلوتاميك وألالنين بين كودون 763 و 764 في اكسون 20، كانت طفرة في المجال كاينيز من *EGFR*. ان تحديد وتحليل الطفرات النادرة باستخدام هذه التقنية كانت حساسة وبسيطة ودقيقة و غير مكلفة، وهو يستخدم كدلائل وراثية لمختلف التسلسلات الأليلة والطفرية.

الكلمات المفتاحية: سرطان الكبد، طفرة ***EGFR***، اكسون 20

Introduction

Lung cancer, a leading cancer in terms of number of patients and deaths recorded globally, is driven by multiple, complex and heterogeneous genetic abnormalities (1). Late detection is the attribution of Poor survival rates (10% for 5 years), often when the tumour has metastasised to other parts of the body during which surgical intervention is ruled out, and resistance to drugs (2). Generally, a large number of oncogenes such as epidermal growth factor receptor tumour (*EFGR*) and Kristen rate sarcoma (*KRAS*), and suppressor genes such as breast cancer 1 (*BRCA1)* and phosphatase and tensin homolog (*PTEN)* (3). Lung cancer is also known for wide ranges of mutations, for example, tumor protein p53 *(TP53)* which encodes a transcription factor that plays a role in arresting cell cycle, initiating DNA repair and apoptosis, and is inactivated in 90% of small cell lung cancer (SCLC) and 50% of non-small cell lung cancer (NSCLC). This inactivation has a role in DNA binding domain and point mutations (4).

The most significant gene convoluted in lung cancer is *EGFR* oncogene which is mutated in around 25% of NSCLC. This gene encodes for a cellular transmembrane receptor tyrosine kinase named *EGFR*. This gene will activate after binding of ligands such as transforming growth factor-α (TGFα) and epidermal growth factor (EGF). The activation result in autophosphorylation and intracellular activation leading to the signalling, cell proliferation and differentiation (5). Based on the importance of this gene, this study was aimed to develop a model of PCR based diagnostic technique using mouse genome for one rare gene mutation for *EGFR* that have diagnostic and prognostic values in lung cancer and not covered by commercial test kits.

Materials and Methods

Over all work of the research starting from bioinformatics database to laboratory work is shown in (figure 1).

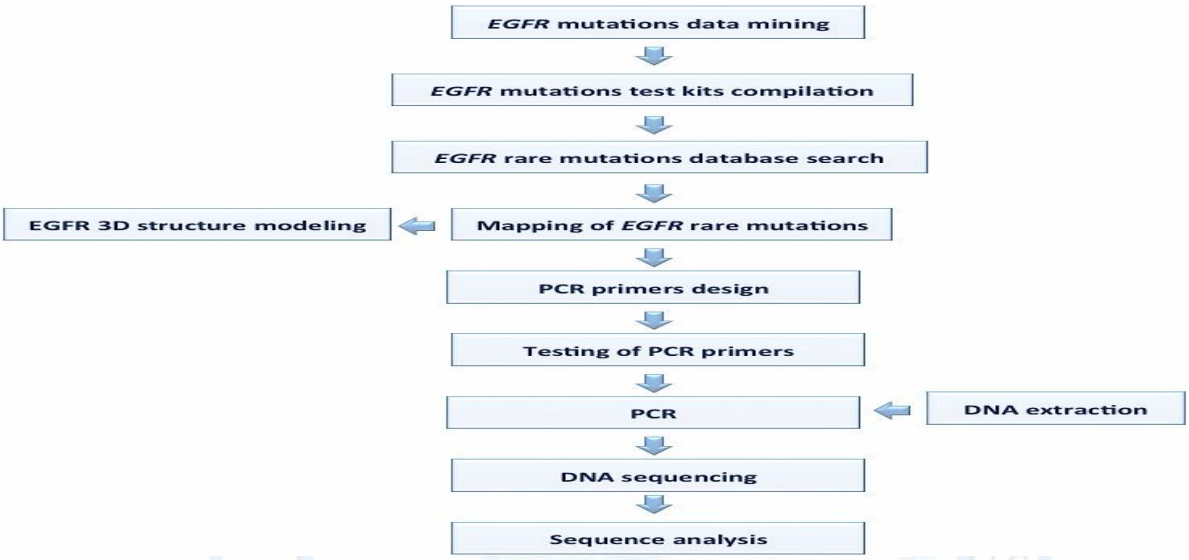


Figure1: Flowchart of steps undertaken in this study.

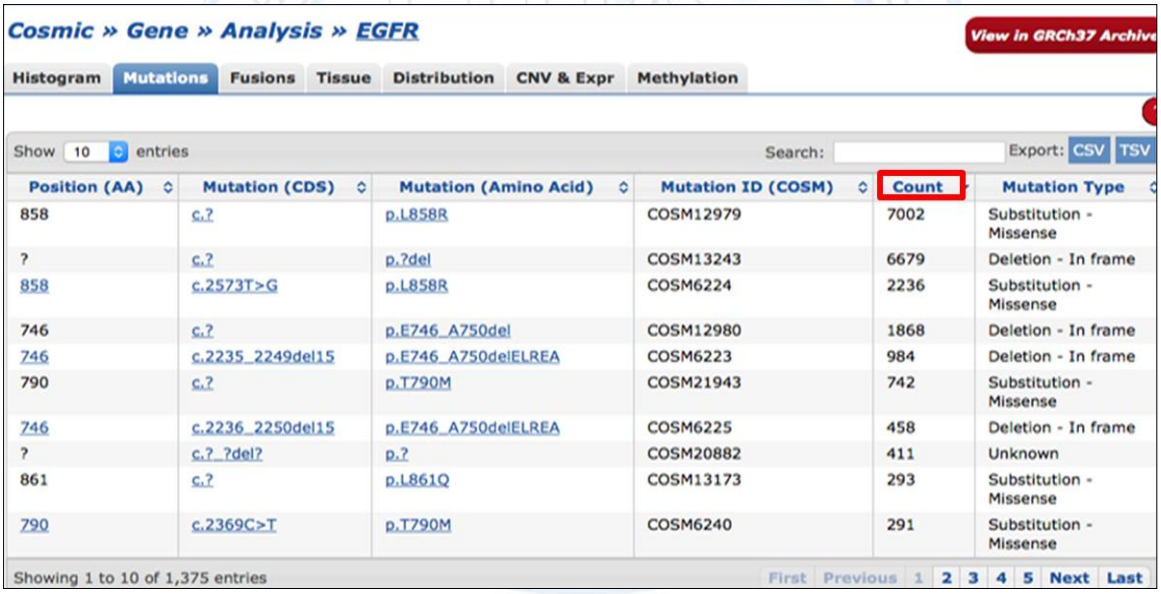
***EGFR* Mutation Data Mining:**

***EGFR* Mutations Test Kits Data collection:**

The Google search box was used for the term '*EGFR* mutation test kit. The detailed information like exon count, amino acid changes nucleotide changes and Catalogue of Somatic Mutation in Cancer (COSMIC) were saved taken from the Website links for *EGFR* mutation test kit companies were accessed. Microsoft Excel spreadsheets were used for collection of mutations test kits according to the companies and arranged by exon count.

EGFR rare mutations database search

COSMIC database (http://cancer.sanger.ac.uk/cosmic) was accessed and the term 'EGFR' was typed in the COSMIC homepage text search box. The result in the second table was used; the Histogram link was clicked for the next page results. The diagnostic and prognostic data like amino acid changes, mutation ID and nucleotide changes was taken from the mutation tab. The number of unique mutated samples was obtained from the Count box data, as shown in (Figure 2). The table was crossed by clicking the Next box below the table to view all the mutation entries. The mutation which selected for this study was cross-checked with that of *EGFR* mutation test kits in the spreadsheets.



Cosmic » Gene » Analysis » EGFR

View in GRCh37 Archive

Histogram | Mutations | Fusions | Tissue | Distribution | CNV & Expr | Methylation

Show 10 entries Search: Export: CSV TSV

Position (AA)	Mutation (CDS)	Mutation (Amino Acid)	Mutation ID (COSM)	Count	Mutation Type
858	c.?	p.L858R	COSM12979	7002	Substitution - Missense
?	c.?	p.?del	COSM13243	6679	Deletion - In frame
858	c.2573T>G	p.L858R	COSM6224	2236	Substitution - Missense
746	c.?	p.E746_A750del	COSM12980	1868	Deletion - In frame
746	c.2235_2249del15	p.E746_A750delELREA	COSM6223	984	Deletion - In frame
790	c.?	p.T790M	COSM21943	742	Substitution - Missense
746	c.2236_2250del15	p.E746_A750delELREA	COSM6225	458	Deletion - In frame
?	c.?_?del?	p.?	COSM20882	411	Unknown
861	c.?	p.L861Q	COSM13173	293	Substitution - Missense
790	c.2369C>T	p.T790M	COSM6240	291	Substitution - Missense

Showing 1 to 10 of 1,375 entries First Previous 1 2 3 4 5 Next Last

Figure 2: The Count heading of the fifth column of the table (shown in red box) was clicked twice to sort the mutations entries in descending order according to the number of unique mutated samples.

Mapping of Mutations to *EGFR* Gene Structure:

Intron and exon sequences were found from Ensemble database (http://www.ensembl.org/index.html). Human was typed in homepage and the 'egfr' was typed in the text search box. The information was obtained by selecting the drop down menu, then Transcript ID column from transcript table was selected, as shown in (Figure 3). This led to opening of a new side panel of Transcript-based displays on the left of the page. Information on amino acid sequence and cDNA sequence were obtained under the sequence menu from exon and cDNA link. These steps were repeated for mouse *EGFR* where in the first step in the Ensembl homepage, Mouse was selected in the Search drop down menu and 'egfr' was typed in the text search. The subsequent steps were followed as above and mouse *EGFR* DNA sequences and amino acid sequences of exons, introns.

National Center for Biotechnological Information (NCBI) (http://www.ncbi.nlm.nih.gov/) database was used for finding the complete *EGFR* DNA sequence. In the NCBI homepage, Gene was selected in the drop down menu, 'egfr' was typed in the text search box and search button was clicked. The first results entry corresponding to *EGFR* (ID:1956) was selected which led to a new page. Microsoft panel was used for obtaining the mutation mapping for egfr gene for both human and mouse mutations.

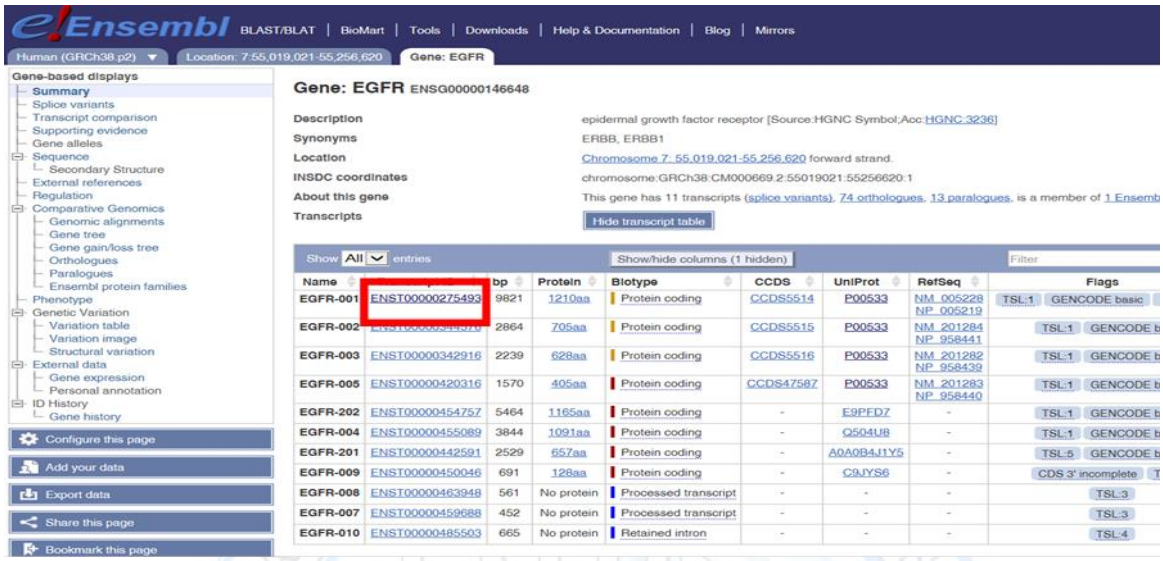


Figure 3: The first entry under the Transcript ID column (shown in red box) corresponding to transcript for *EGFR* protein with 1210 amino acids was clicked.

Designing Of PCR Primers:

PCR primers were designed using Primer3Plus software package from (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/). *EGFR* exon 20 sequence (186bp) with upstream intron sequence (71bp) and downstream intron sequence (93bp) were pasted into sequence query box of the Primer3Plus homepage. Two sets of primers were generated by using the Square brackets from the beginning and at the end of the exon sequence to mark the targeted region and general Settings tab were clicked and parameters were adjusted, then Pick Primers button was selected.

University of California Santa Cruz (UCSC) *in silico* PCR:

UCSC website (https://genome.ucsc.edu/) was used and *in Silico* PCR link on the side menu was clicked which led to a new page. The search box were used for both forward and reverse primer were obtained from Primer3 Plus for human and mouse *EGFR* exon 20 and 21 respectively

DNA extraction from blood sample:

There were 6 samples of mouse genomic DNA extracted from mouse blood, using the protocol from Center for Cancer Research (2005) for DNA extraction with some modification was undertaken (protocols prepared by Williams and Rapley, 2000) (6). 4 ml of lysis buffer was added to 2 ml of mouse (*Mus musculus)* EDTA treated whole blood and shaken gently by hand for 1 min. Then the mixture was incubated on ice for 30 minutes and centrifuged at 2400 rpm for 5 minutes. The supernatant was discarded and the pellet was treated again with 3.5 ml of lysis buffer and then centrifuged again at 2400 rpm for 5 min. Then the pellet was treated with SE buffer and centrifuged again for obtaining a pellet. Then 10 µl of proteinase K (20mg/ml) from Invitrogen and 500 µl of 10% sodium dodecyl sulphate (SDS) were added to the pellet and incubated for 18 hrs at 37°C in a water bath. After that, 1.5 ml of SE buffer and 4 ml of phenol were added and the supernatant was collected using 3000 rpm centrifuge. 10 ml of phenol/isoamyl alcohol was added then centrifuged for 10 minutes at 2500 rpm. The collected supernatant then treated with 1.5ml of isopropanol and 250 µl of 3 M sodium acetate from Fisher. Then the mixture was put at the refrigerator for 30 minutes and then the DNA precipitation (white colour) was appeared by shaking and centrifuging gently. The pellet then washed with 70% isopropanol. DNA was dissolved in 0.4 ml TE buffer. DNA concentration and wavelength at 230 nm, 260 nm and 280 nm were measured in eppendorf Biophotometer. 1% agarose gel was used for finding the purity of the extracted DNA. The destined gel was then visualised on UV transilluminator to evaluate the quality and quantity of the DNA.

PCR Analysis:

PCR amplifications were carried out according to protocols from (Kumari *et al.*, 2015) (7) using selected primers from Primer3 Plus:

Forward: 5'-AAAGGGATATGCGTGCCTCT-3'

Reverse: 5'-AGACCTCCCAACGTGCTTAC-3'

The PCR reaction for targeted sequence from exon 20 was set up in 0.2 ml eppendorf tube, 25 µl PCR mixture containing: 1 µl of 50 mM $MgCl_2$, 0.5 µl of 10 mM dNTPs, 2.0 µl of 10X reaction buffer, 0.5 µl of each of forward and reverse primers, 0.2 U of *Taq* polymerase (Invitrogen), and then the mixture was completed to 25 µl with 1.5 µl DNA template and sterile distilled water. Positive and negative controls were also prepared for the PCR set up. PCR conditions were consisted of 1 cycle of 5 min at 94°C, followed by 40 cycles of 1 minute at 94°C, 1 minute at 55°C, 1 minute extension at 72°C, and a final cycle of 10 min at 72°C. Amplified products were visualized on 1.5% agarose gel under UV light.

DNA Sequencing and Analysis:

The bands corresponding to targeted sequence for exon 20 were cut with scalpel visualised by UV light, transferred to eppendorf tube and 400 µl of TE buffer was added. The targeted amplicons were obtained from 1% agarose gel were sent to Bio-Rad Development Laboratories for DNA sequencing in Hemel Hempstead using Sanger sequencing method. Sequence data saved were analysed using BLAST software package. BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) website was accessed and BLASTN under the Basic BLAST section was clicked. This led to a new page where DNA sequence for exon 20 was pasted in Enter Query Sequence query box. BLAST button at the bottom of the page was clicked and results for BLAST search were generated. Under the Descriptions panel, a table of sequences producing significant alignment is produced and results such as Query cover and E-value were evaluated for the aligned sequences.

Results and Discussion:

***EGFR* mutations Test Kits**

Currently, lung cancer patients are tested for a total of 132 common *EGFR* mutations in hospitals worldwide (8). These mutations are tested using commercially available test kits from various companies that are approved for diagnostic purpose by relevant authorities. The most utilised technique for these mutations test kits are real-time PCR (by Qiagen, Amoy, Roche and

EntroGen), followed by PCR clamping (by Panagene and DiaCarta), pyrosequencing (by Qiagen) and NGS (by Arup Laboratories). These mutations span across exon 2, 6, 11, 13, 18, 19, 20 and 21 with 1, 3, 1, 2, 19, 60, 29 and 17 number of mutations, respectively. Over 88% of these mutations are located in exon 18 to 24 and within the TK domain of *EGFR* (between amino acid residues of 712 to 979).

Mutation of Interest Mapping on *EGFR* Exons:

The mutation of interest selected was A763_Y764insFQEA (insertion of phenylalanine, glutamine, glutamic acid and alanine in between codon 763 and 764) in exon 20, the mutation was in the kinase domain of *EGFR* (figure 4)

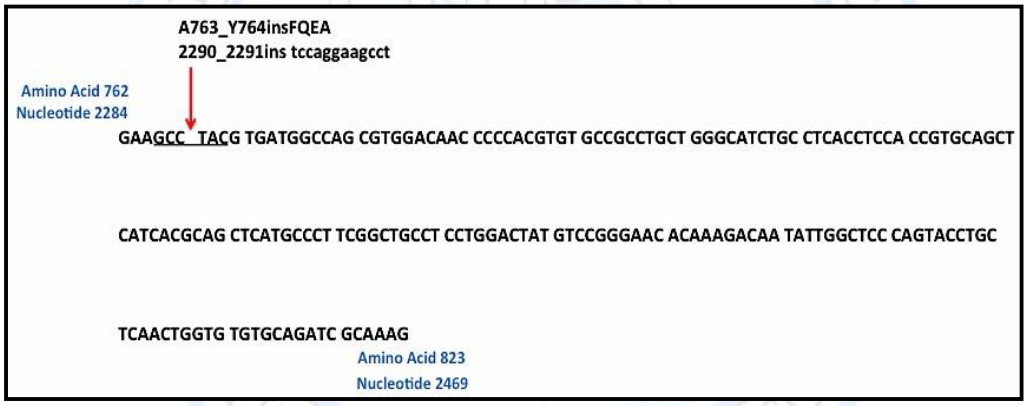


Figure 4: Human *EGFR* exon 20 nucleotide sequence (186 nucleotides from position 2284-2469) and the position of 1 mutation of interest, 2290_2291 ins tccaggaagcct (A763_Y764insFQEA).

DNA Quantity and Quality Analysis:

Six samples of genomic DNA extracted from mouse blood with concentration and absorbance ratio of 260/280 nm and 260/230 nm measured on eppendorf Biophotometer as shown in Table 3.8. DNA sample 1 had the highest concentration, 38 µg/ml. DNA sample 2 had the closest 260/280 nm absorbance ratio to 1.8, which was 1.84 followed by DNA sample 1 with 1.85 (figure 5).

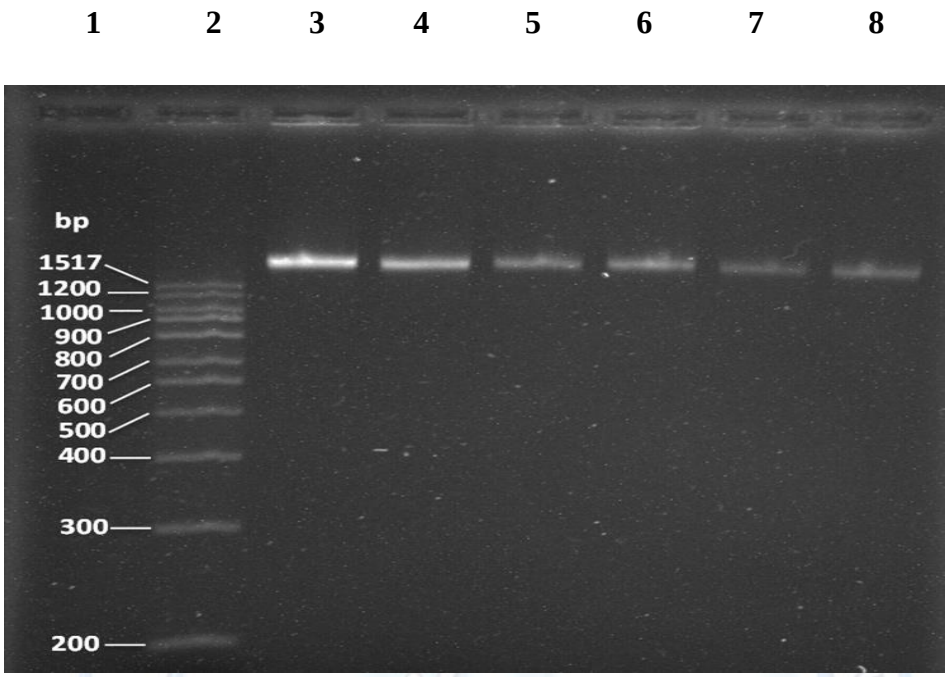


Figure 5: 1.5% agarose gel electrophoresis of DNA extracted from mouse blood. Lane 2 was loaded with 100 bp DNA ladder, lane 3, 4, 5, 6, 7 and 8 were mouse DNA with concentration of 38 µg/ml, 18 µg/ml, 33 µg/ml, 38 µg/ml, 30 µg/ml and 24 µg/ml, respectively. The migration distance of all 6 DNA samples were approximately similar, which were shorter than the first DNA ladder fragment of 1517 bp. The intensity of the DNA bands was the highest for DNA sample in lane 3 and decreased from lane 3 to lane 8.

PCR Analysis:

PCR products for *EGFR* exon 20, exon 21, β-actin positive control and negative control were run in 1.5% agarose gel electrophoresis and results are as shown in Figure 6. In lane 2 and 3, clear thick bands of 200-300 bp was observed which corresponded to the expected mouse *EGFR* exon 20 PCR product of 276 bp. In lane 4, a clear band was observed which corresponded to β-actin positive control of approximately 400 bp. No band was observed in lane 5 negative control.

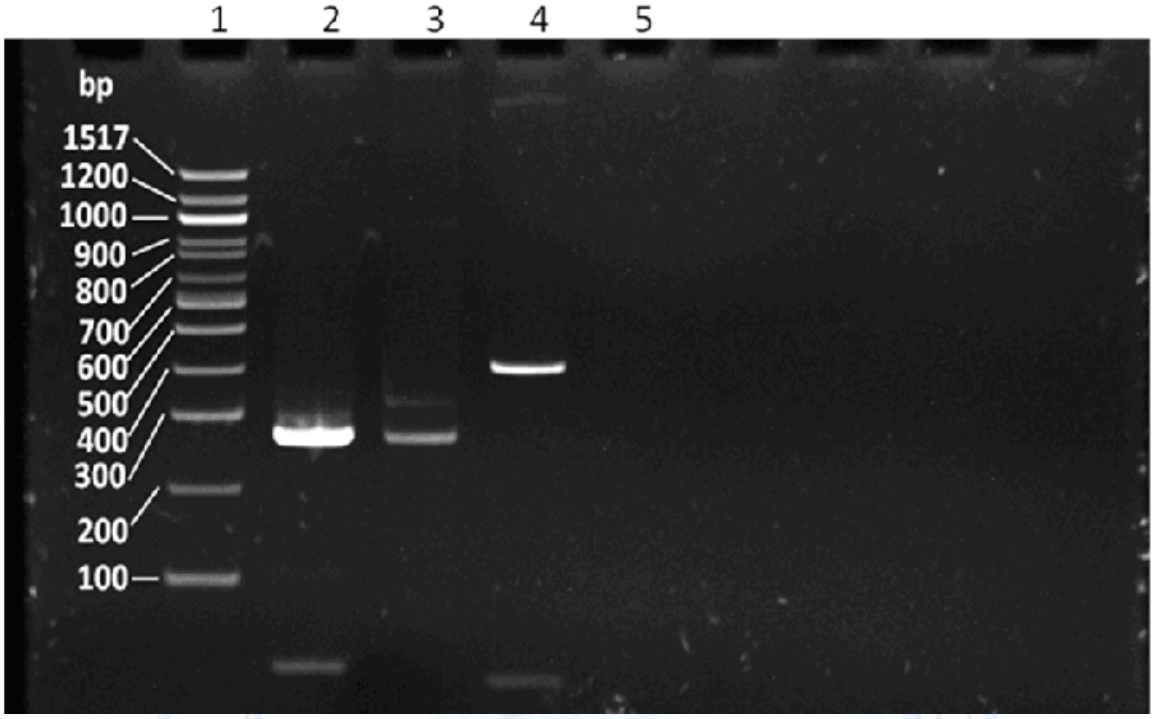


Figure 6: 1.5% Agarose gel of mouse *EGFR* targeted sequence. Lane 1 was loaded with 100 bp DNA ladder, Lane 2 and 3 with targeted sequence (clear expected band from 200-300bp), lane 4 with amplicon for β-actin positive control and lane 5 with negative control.

DNA Sequence and Analysis:

Before cutting the bands, the UV transluminator was covered with plastic wrap. The UV light made the bands more clear for cutting. Then the left side of the DNA bands was cut and putted in a 3ml of plastic TE buffer for sanger sequencing. DNA sequencing (Sanger sequencing) for amplicon of exon 20 produced a partial DNA sequence of (79 bp out of 276 bp predicted) with 16 unassigned bases within the sequence marked as 'N'. BLASTN search showed that the partial exon 20 sequence corresponds to mouse *EGFR* sequence on chromosome 11 as expected as shown in (table 1).

Table 1. Sanger sequencing results for amplicon of exon 20.The size of the sequence is 79 bp as compared to the predicted size of 267 bp. Matching bases between Sanger sequencing and predicted sequences are highlighted in red, while 'N' represents the unassigned bases.

Amplicon	Predicted		Sanger sequencing	
	DNA sequence	Size (bp)	DNA sequence	Size (bp)
Exon 20	AAAGGGATATGCGTGCCTCTCCCACCAGCCAGGAAACAATTGTGATTCATCTATTGTCCTTACCTTGTAGGAAGCCTATGTGATGGCTAGTGTGGACAACCCTCATGTATGCCGCCTCCTGGGCATCTGTCTGACCTCCACTGTCCAGCTCATTACACAGCTCATGCCCTACGGTTGCCTCCTGGACTACGTCCGAGAACACAAGGACAACATTGGCTCCCAGTACCTCCTCAACTGGTGTGTGCAGATTGCAAAGGTAAGCACGTTGGGAGGTCT	276	NNNNTGNNNNCTCCTGNNCATCTGTCTGACCTCCACTGTCCAGCTCATTACACAGCTCATNCCCTACGGTTNNTCCNNN	79

Bioinformatics resources that house *EGFR* mutations were searched systematically to identify rare *EGFR* mutations. COSMIC database provided all the *EGFR* mutations in City of Hope. Hence, COSMIC database can be considered to be a principal database with most comprehensive EGFR mutations coverage, delivering deep mutational profiles to support cancer genetics research as a whole (9). The City of Hope database is also important for cancer research due to its collection of somatic mutations in *EGFR* and the associated epidemiological data which relates to the prognostic indicators of NSCLC (10). There are a huge number of mutations covered by commercially test kits, but also there are rare mutations with diagnostic and prognostic values are not analysed. This is due to the large number of total *EGFR* mutations (over 1,300 *EGFR* mutation entries in COSMIC) and partly explains the ever-growing

importance of *EFGR* mutations in lung cancer (11). The mutation which selected in this research was A763_Y764insFQEA in exon 20. This mutation is another mutation that has shown good response to gefitinib and erlotinib (12). In this study, a key bioinformatics tool, Primer3Plus software was used in the designing of the primers for exon 20. In that it considered the GC-rich region and Tm (13). Additionally, Primer3Plus takes in to account the number of hydrogen bonds and the duplex thermodynamics based on nearest neighbour (NN) model to evaluate the binding stabilities between primer and template. Researchers have suggested *in silico* testing of PCR primer sequences using bioinformatics tools such as UCSC *in silico* PCR and BLAST to provide and confirm certainty on the primer specificity for the amplicon before laboratory-based PCR is undertaken (14, 15).The ration of A_{260}:A_{280} of above 1.8 possibly indicates contamination with RNA. Hence, to observe the DNA integrity and possible impurities with RNA, the DNA samples were run in 1% agarose gel and visualised by ethidium bromide under UV light (16). Hence, the selected DNA sample in lane 3 was selected for this study in that it possesses the highest band intensity as compared to the other samples, with no RNA smear observed as shown in Figure 5. Overall, the most efficient and cost effective way leading to generation of sequencing data for mutational detection was aimed for in this study. Taking into account the results obtained from the laboratory work, there is certainly room for improvement to fulfil the objective. In the DNA extraction section, the purity of the genomic DNA might have negatively affected the subsequent steps. The ratio A_{260}:A_{230} of 1.65 was slightly lower than the ideal ratio of 1.8, indicating the presence of impurities such as phenol and salts, which in certain cases, may inhibits PCR and Sanger sequencing efficiency. The impurities can be treated using extra 70% ethanol washing steps or by modifying the isopropanol precipitation step (as salts tends to precipitate together with DNA with the use of isopropanol) in this study by precipitating the DNA with 1/10 volume of 3M acetate and 2 to 2.5 volume of 100% ice-cold ethanol for 20 minutes at -20°C as recommended by Sambrook and Russell (2001) (17). The assessment of DNA concentration can be further improved by using fluorescent dye based method such as PicoGreen assay. While single stranded DNA (ssDNA), RNA and degraded DNA may affect the spectrophotometric reading at 260 nm and

280 nm, Picogreen assay selectively detect the double stranded DNA (dsDNA) in the presence of these nucleic acid impurities (18, 19).

Mutational detection depends on the steps of successful PCR amplification by the subsequent downstream step of DNA sequencing. Optimisation of PCR requires careful troubleshooting so that strategies can be conceived in a timely and efficient manner (20). Direct sequencing of PCR products is still a widely used method and a gold standard to detect *EGFR* mutations in spite of its low sensitivity and emergence of more sensitive methods. Many studies have highlighted its inability to detect *EGFR* mutations with higher false negative rate as compared to methods such as amplification refractory mutation system and high-resolution melting analysis (21, 22).

References

1. Sekido, Y., Fong, K. M., and Minna, J. D. (2003). Molecular Genetics of Lung Cancer. *Annual Review of Medicine*, 54(1):73-87.
2. Brennan, P., Hainaut, P., and Boffetta, P. (2011). Genetics of Lung-Cancer Susceptibility. The Lancet Oncology, 12(4): 399-408.
3. Strachan, T., and Read, A. (2011). Human Molecular Genetics. (4th ed.). New York: Garland Science.
4. Sato, M., Shames, D. S., Gazdar, A. F., and Minna, J. D. (2007). A translational view of the molecular pathogenesis of lung cancer. Journal of Thoracic Oncology, 2(4): 327-343.
5. Parker, B. C., and Zhang W. (2013). Fusion genes in solid tumors: an emerging target for cancer diagnosis and treatment. Chin J Cancer, 32(11): 594–603.
6. Williams, D. R., and Rapley, R. (2000). Agarose gel electrophoresis of nucleic acid. In R. Rapley (Ed.). The Nucleic Acid Protocols Handbook. (pp. 67-70). New Jersey: Humana Press Inc.

7. Kumari, N., Thakur, S. K., Kumar, D., & Kumari, K. (2015). Single Strand Conformation Polymorphism (SSCP) – A Review. Indian Research Journal of Genetics and Biotechnology, 7(1): 27-34.

8. Sag, S. O., Gorukmez, O., Ture, M., Gorukmez, O., Deligonul, A., Sahinturk, S., Topak, A., Gulten, T., Kurt, E., and Yakut, T. (2016). Spectrum of EGFR gene mutations and ALK rearrangements in lung cancer patients in Turkey. Springerplus, 5(482): 1-6.

9. Forbes, S. A., Beare, D., Gunasekaran, P., Leung, K., Bindal, N., Boutselakis, Ding, M., Bamford, S., Cole, C., Ward, S., Kok, C. Y., Jia, M., De, T., Teague, J. W., Stratton, M. R., McDermott, U., and Campbell, P. J. (2015). COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Research, 43(D1), D805-D811.

10. Gu, D., Scaringe, W. A., Li, K., Saldivar, J. S., Hill, K. A., Chen, Z., Gonzalez, K. D., and Sommer, S. S. (2007). Database of somatic mutations in EGFR with analyses revealing indel hotspots but no smoking-associated signature. Hum Mutat. 28(8):760-770.

11. Angulo, B., Conde, E., Suárez-Gauthier, A., Plaza, C., Martínez, R., Redondo, P., Izquierdo, E., Rubio-Viqueira, B., Paz-Ares, L., Hidalgo, M., and López-Ríos, F. (2012). A Comparison of *EGFR* Mutation Testing Methods in Lung Carcinoma: Direct Sequencing, Real-time PCR and Immunohistochemistry. PLoS One. 7(8): e43842.

12. Oxnard, G. R., Lo, P., Nishino, M., Dahlberg, S., Lindeman, N. I., Butaney, M., Jackman, D. M., Johnson, B. E., and Jänne, P. A. (2013). Natural history and molecular characteristics of lung cancers harboring EGFR exon 20 insertions. Journal of Thoracic Oncology, 8(2): 179-184.

13. Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., and Leunissen, J. A. (2007). Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Research, 35(suppl 2), W71-W74.

14. Whitehouse, D. B. (2009). Genes and Genomes. In J.M. Walker & R. Rapley (Eds.), Molecular Biology and Biotechnology (5th ed.). (pp. 112-153). Cambridge: Royal Society of Chemistry.

15. Cao, Y., Wang, L., Xu, K., Kou, C., Zhang, Y., Wei, G., He, J., Wang Y., and Zhao L. (2005). Information theory-based algorithm for *in silico* prediction of PCR products with whole genomic sequences as templates. BMC Bioinformatics. 6: 190-195.

16. Clark, W., and Christopher, K. (2000). An Introduction to DNA: Spectrophotometry, degradation, and the 'Frankengel' experiment. Tested Studies for Laboratory Reaching, 22:81-99.

17. Sambrook, J., and Russell, D. W. (2001). Molecular cloning. A Laboratory Manual. (3rd ed.). New York: Cold spring Harbor Laboratory Press,

18. Psifidi, A., Dovas, C. I., Bramis, G., Lazou, T., Russel, C. L., Arsenos, G., and Banos, G. (2015). Comparison of Eleven Methods for Genomic DNA Extraction Suitable for Large-Scale Whole-Genome Genotyping and Long-Term DNA Banking Using Blood Samples. PLoS One. 2015; 10(1): e0115960.

19. Olson, N. D., and Morrow, J. B. (2012). DNA extract characterization process for microbial detection methods development and validation. BMC Res Notes. 5: 668-681.

20. Lorenz, T. C. (2012). Polymerase chain reaction: basic protocol plus troubleshooting and optimization strategies. Journal of Visualised Experiments, 63: e3998.

21. Da Cunha Santos, G., Shepherd, F. A., and Tsao, M. S. (2011). EGFR mutations and lung cancer. Annual Review of Pathology: Mechanisms of Disease, 6: 49-69.

22. Pao, W., and Ladanyi, M. (2007). Epidermal growth factor receptor mutation testing in lung cancer: searching for the ideal method. Clinical Cancer Research, 13(17): 4954-4955.