
REZUMATUL TEZEI DE DOCTORAT

Utilizarea tehnologiilor de Secvențiere de Nouă Generație (NGS) în medicina personalizată

Student doctorand **Diana-Cecilia Pop (Bica)**

Conducător de doctorat **Prof. univ. dr. Ioana Neagoe**



UMF
UNIVERSITATEA DE
MEDICINĂ ȘI FARMACIE
IULIU HAȚIEGANU
CLUJ-NAPOCA

CUPRINS

INTRODUCERE	18
STADIUL ACTUAL AL CUNOAȘTERII	
1. Caracterizarea patologiei oncologice	23
1.1. Patologia oncologică și rata de supraviețuire la nivel mondial	23
1.2. Profilul genomic în patologia oncologică	23
1.2.1. Tipuri de mutații în patologia oncologică	24
1.2.2. ARN-urile noncodificatoare în cancer	25
1.2.3. Tipuri de ARN-uri noncodificatoare și rolul acestora în cancer	26
1.2.4. Medicina de precizie în oncologie	30
2. Cancerul pulmonar: diagnostic și management	
2.1. Statistici în cancerul pulmonar	32
2.2. Clasificarea și caracterizarea cancerului pulmonar	33
3. Tehnologiile de tip "high-throughput" și utilizarea acestora în patologia oncologică	36
3.1. Secvențierea de Nouă Generație (NGS)	36
3.2. NGS în aplicații de tip screening al cancerului	38
3.3. NGS în diagnosticul și prognosticul cancerului	39
3.4. NGS în medicina de precizie	39
3.4.1. NGS în caracterizarea profilului molecular al cancerului pulmonar	41
CONTRIBUȚIA PERSONALĂ	
1. Ipoteza de lucru/Obiective	47
2. Metodologia generală	49
3. Studiul 1. Caracterizarea profilului mutațional al liniilor celulare de plămân – tumorale și normale	
3.1. Conținutul studiului	53
3.2. Ipoteza de lucru și obiectivele	54
3.3. Materiale și metode	55
3.4. Rezultate	
3.4.1. Caracteristicile liniilor celulare analizate	56
3.4.2. Mutații identificate în experimentul de secvențiere	57

3.4.3. Concordanța dintre profilul ATCC și mutațiile identificate în experimental de secvențiere	59
3.5. Discuții	60
3.6. Concluzii	61
4. Studiul 2. Identificarea profilului mutațional al pacienților cu cancer pulmonar prin aplicarea tehnologiei NGS de secvențiere-țintită	
4.1. Conținutul studiului	63
4.2. Ipoteza de lucru și obiectivele	64
4.3. Materiale și metode	64
4.4. Rezultate	
4.4.1. Caracteristicile clinico-patologice ale pacienților din cohorta de studiu	66
4.4.2. Mutații identificate în experimentul de secvențiere	70
4.4.3. Analiza statistică	79
4.5. Discuții	81
4.6. Concluzii	84
5. Studiul 3. Validarea variațiilor rs3822214 și rs1042522 identificate și caracterizarea profilului mutațional somatic al cancerului pulmonar prin analiza seturilor de date din TCGA.	
5.1. Conținutul studiului	85
5.2. Ipoteza de lucru și obiectivele	86
5.3. Materiale și metode	86
5.4. Rezultate	
5.4.1. Caracteristicile clinico-patologice ale pacienților din cohorta de studiu	88
5.4.2. Rezultatele SNP Genotyping assay de validare a mutațiilor	90
5.4.3. Alterări genetice identificate în seturile de date de cancer pulmonar din TCGA	90
5.4.4. Analiza statistică	94
5.5. Discuții	95
5.6. Concluzii	96
6. Discuții generale	99
7. Concluzii generale	103
8. Originalitatea și contribuția inovativă a tezei de doctorat	105
BIBLIOGRAFIE	109

Cuvinte cheie: secvențiere de nouă generație (NGS), secvențiere-țintită, profil mutațional, medicină personalizată, cancer pulmonar, terapie țintită.

INTRODUCERE

Patologia oncologică reprezintă o problemă majoră la nivel mondial, tratamentele existente fiind eficiente la un număr redus de pacienți, fapt ce duce la o mortalitate ridicată în rândul pacienților cu afecțiuni oncologice. Această situație este datorată atât heterogenității în rândul patologiilor oncologice, dar mai ales mecanismelor moleculare care stau la baza dezvoltării și progresiei oncologice.

Evoluția celulelor normale din diferite localizări tisulare în celule tumorale apare descrisă în literatura de specialitate, iar caracteristicile acestor celule transformate includ instabilitatea genomică cauzată de mutații, potențial replicativ nelimitat, inducerea procesului de angiogeneză la situsul tumoral, mecanisme de eludare a supravegherii imune și a moleculelor de tip supresor de creștere, activarea mecanismelor de invazie și metastazare, dar și de un metabolism energetic alterat. În consecință, numeroase studii sunt axate pe elucidarea transformărilor moleculare care duc la validarea unui fenotip malign al acestor celule pentru a identifica potențiale ținte terapeutice în medicina personalizată, care să crească rata de răspuns la terapie la pacienții cu cancer și să reducă efectele adverse care apar în cazul chimio- și radioterapiei. Odată cu progresele tehnologice din ultimele decenii și dezvoltarea domeniului bioinformatic care permite valorificarea informației biologice generată prin tehnologiile de tip "high-throughput", apare posibilitatea diagnosticării unui pacient la nivel molecular care favorizează alegerea unui tratament țintit care să asigure un prognostic favorabil și/sau remisia completă la pacienții cu cancer. Secvențierea de Nouă Generație (NGS) reprezintă una din tehnologiile de tip "high-throughput" ce sunt folosite la momentul actual, atât în laboratoarele de cercetare, cât și în practica clinică pentru identificarea alterărilor genetice asociate cu diferite tipuri de cancer. În ce privește cancerul pulmonar, acesta reprezintă o categorie de patologii maligne cu două clase histologice majore diferite – cancerul pulmonar microcelular și cancerul pulmonar non-microcelular, cel din urmă având două subtipuri histologice principale diferite: adenocarcinomul pulmonar și carcinomul pulmonar scuamos. Heterogenitatea acestei patologii rezidă și din numărul ridicat de mutații cauzative ce caracterizează tumora pulmonară, limitând astfel abordarea unei scheme terapeutice cu compuși țintiți în practica clinică.

Acest studiu a urmărit caracterizarea profilului mutațional al tumorilor pulmonare în vederea identificării de gene care pot servi ca potențiale ținte terapeutice în medicina personalizată. Astfel, primul studiu a urmărit caracterizarea profilului mutațional al liniilor celulare pulmonare – tumorale și normale – în vederea stabilirii rețelelor moleculare implicate în oncogeneza tumorilor pulmonare. De asemenea, în acest studiu s-a urmărit identificarea alterărilor genetice prezente în tumorile și în probele de sânge integral de la pacienți cu diferite tipuri de cancer pulmonar în

vederea identificării similarităților sau diferențelor ce caracterizează fiecare tip histologic de cancer pulmonar. În urma identificării celor mai comune mutații cu semnificație clinică (patogene sau asociate cu răspunsul la terapie) s-a utilizat un lot separat de pacienți cu caracteristici clinico-patologice similare pentru validarea prezenței acestora. De asemenea, s-a efectuat analiza datelor de mutații prezente în gene asociate patologiei oncologice în diferite seturi de date ale pacienților cu cancer pulmonar din baza de date TCGA (The Cancer Genome Atlas).

STADIUL ACTUAL AL CUNOAȘTERII

În conformitate cu statisticile GLOBOCAN 2020, cancerul reprezintă o problemă majoră în domeniul sănătății, în fiecare an fiind diagnosticate aproximativ 19.3 milioane de noi cazuri. Analizând situația privind incidența cancerului în GLOBOCAN 2018, unde numărul de noi cazuri înregistrate anual era de aproximativ 18.1 milioane, precum și cea din 2012, unde numărul de noi cazuri înregistrate anual era de 14.1 milioane, se poate observa faptul că patologii oncologice ridică probleme la nivel mondial în ce privește prevenția.

Cancerul pulmonar este împărțit în 2 tipuri histologice majore – cancerul pulmonar cu celule mici (SCLC) și cancerul pulmonar non-microcelular (NSCLC), cel din urmă având două subtipuri principale (adenocarcinomul pulmonar și carcinomul pulmonar scuamos), iar acestea reprezintă aproximativ 85% din totalul cazurilor de cancer pulmonar. După cum era de așteptat, implicațiile terapeutice și prognosticul pacienților sunt diferite pentru cele două grupuri. Tratamentul pacienților cu SCLC nu implică rezecția chirurgicală datorită agresivității tumorii. În cazul pacienților cu NSCLC, tratamentul implică o combinație între rezecția chirurgicală și terapia adjuvantă.

În privința caracteristicilor moleculare ale tumorilor pulmonare, literatura de specialitate menționează că în cazul adenocarcinomului pulmonar, conform unei analize efectuată utilizând seturile de date din TCGA, cele mai frecvent alterate gene sunt *TP53* (46%), *KRAS* (33%), *STK11* (17%), *KEAP1* (17%) și *EGFR* (14%). Lista de gene frecvent mutate fiind completată de gene precum *BRAF*, *NF1*, *PIK3CA*, *RB1* și *CDKN2A*. Același studiu a arătat faptul că în 75% din probele examinate erau întâlnite mutații cauzative în gene implicate în calea de semnalizare *RTK/RAS/RAF*, precum *BRAF*, *EGFR*, *ERBB2*, *HRAS*, *KRAS*, *MAP2K1* și *NRAS*. La pacienții cu SCLC, particularitățile moleculare indicate de studiile de specialitate se referă mai ales la alterări de tipul variațiilor în numărul de copii (CNV) în diferite regiuni, iar acestea includ amplificări în *PDGFRA*, *SOX2*, *FGFR1* și deleții ale genei *CDKN2A*.

Tehnologia NGS poate fi folosită în diferite tipuri de aplicații, inclusiv cercetarea din domeniul cancerului. Abordările din acest domeniu diferă în funcție de ipotezele studiului, dar și în funcție de nevoile consumatorului. Tehnologia NGS include 4 etape

principale: pregătirea probelor biologice, secvențierea propriu-zisă, analiza calitativă, urmată de analiza de date și interpretarea acestora. În domeniul oncologic, medicina de precizie presupune identificarea alterărilor genetice ce pot favoriza alegerea unui tratament eficient. Prin caracterizarea profilului mutațional al unui pacient sau a unei tumori utilizând tehnologia NGS se subliniază importanța alegerii unui tratament țintit pentru terapia pacienților oncologici. Astfel, prin identificarea mutațiilor specifice care duc la alterarea căilor de semnalizare implicate în oncogeneză și progresia tumorală, este facilitată alegerea unui compus terapeutic țintit care să restabilească funcționarea optimă a căilor de semnalizare alterate de mutații specifice. În felul acesta, este posibilă alegerea unui tratament țintit care să asigure răspunsul la terapie și care să contribuie la îmbunătățirea prognosticului pacienților oncologici.

CONTRIBUȚIA PERSONALĂ

Teza de față este împărțită în trei studii principale, iar obiectivele specifice ale acestora sunt următoarele: • evaluarea profilului molecular al liniilor celulare de cancer pulmonar prin utilizarea tehnologiei NGS de secvențiere-țintită pentru identificarea mutațiilor prezente în gene cu implicare cunoscută în cancer; • analiza profilului mutațional al pacienților cu cancer pulmonar prin secvențierea țintită a țesutului tumoral și a probelor de sânge integral de la pacienți cu cancer pulmonar; • validarea celor mai comune mutații identificate prin analiza unei cohorte separate de pacienți cu cancer pulmonar și prin analiza seturilor de date de cancer pulmonar din baza de date TCGA.

Studiul 1. Caracterizarea profilului mutațional al liniilor celulare de plămân – tumorale și normale

Scopul acestui studiu este de a investiga profilul mutațional caracteristic în opt linii celulare de cancer pulmonar (H1155, Calu6, H123, SKMES 1, H522, M22GB, 1792, and A549) și compararea acestora cu mutațiile identificate în 3 linii celulare normale (fibroblaști) (WI-38, IMR90 și HFL). Pentru identificarea mutațiilor a fost efectuat un experiment de secvențiere țintită utilizând un panel de 46 de gene cu implicare în cancer. Astfel, principalul obiectiv al studiului a fost identificarea modificărilor genetice întâlnite la liniile celulare de cancer pulmonar cel mai frecvent utilizate în studiile *in vitro*. Această abordare permite stabilirea unei baze pentru viitoare studii ce urmăresc studierea căilor moleculare implicate în diferite procese ce caracterizează transformarea malignă a celulelor canceroase și ce modulează rezistența la terapie.

În cadrul acestui studiu am evaluat prezența diferitelor mutații în gene la care apar frecvent mutații în cancer în liniile celulare utilizate în laborator comparativ cu caracteristicile mutaționale standard descrise în baza de date ATCC (American Type Culture Collection). O parte din aceste alterări genetice au fost identificate și în liniile celulare secvențiate în cadrul acestui studiu. Aceste descoperiri oferă informații

importante pentru stabilirea celor mai potrivite linii celulare de cancer pulmonar ce urmează a fi folosite pentru viitoare studii și alegerea liniilor celulare ce pot fi utilizate cu rol de control în laborator. De asemenea, această abordare experimentală oferă baze solide în stabilirea altor studii *in vitro* pentru investigarea răspunsului la terapie sau alte teste funcționale pentru observarea comportamentului celulelor în diferite condiții pentru descoperirea de noi opțiuni terapeutice pentru pacienții cu cancer pulmonar.

Studiul 2. Identificarea profilului mutațional al pacienților cu cancer pulmonar prin aplicarea tehnologiei NGS de secvențiere-țintită

Tehnologia NGS permite cercetătorilor secvențierea în paralel a milioane de fragmente de ADN, iar costurile pentru aceasta sunt în continuă scădere. Această situație motivează denumirea acestei perioade ca "era medicinei de precizie", în contextul posibilității de realizare de investigații de tip screening oncologic pentru a detecta prezența de alterări genetice sau epigenetice la acești pacienți.

Studiul a pornit de la ipoteza că clasificarea morfologică (histologică) este reflectată și în profilul molecular/ mutațional al pacienților cu diferite tipuri de cancer pulmonar. În acest sens, am pornit de la premisa că pentru fiecare tip histologic există alterări genetice specifice ce pot fi identificate în urma experimentelor NGS. Astfel, s-a încercat identificarea și caracterizarea mutațiilor apărute în țesutul tumoral pulmonar și probe de sânge integral de la aceiași pacienți prin folosirea ADN-ului extras din aceste probe biologice pentru secvențiere țintită, utilizând un panel de 46 de gene cu implicare în cancer, pentru a identifica genele frecvent alterate la pacienți cu același tip de cancer pulmonar (NSCLC vs. SCLC). De asemenea, am subliniat diferențele în ce privește profilul mutațional la femei vs. bărbați. O altă ipoteză a acestui studiu a fost prezența diferitelor alterări genetice la nivelul unor gene implicate în cancer ce pot afecta prognosticul pacienților. Astfel, în cadrul acestui studiu au fost realizate și analize statistice pentru validarea acestei ipoteze. Analiza de corelație Pearson a urmărit stabilirea unei legături între prezența alterărilor genetice patologice sau posibil patologice și caracteristicile pacienților investigați în prezentul studiu.

Astfel, în cohorta de pacienți analizată, *TP53* a apărut ca fiind gena cel mai frecvent mutată și cu cele mai multe alterări genetice diferite, cu mutații în 93% din pacienții cu SCLC, în timp ce la pacienții cu NSCLC, frecvența alterărilor genetice era de 87% în probele de țesut tumoral. Analiza mutațiilor identificate în probele de sânge a arătat că *TP53* este alterată în 81% din pacienții cu NSCLC și 68% la pacienții cu SCLC. Cele mai multe alterări genetice identificate în gena *TP53* au fost clasificate ca patogene, iar mutația c.215C>G apare ca fiind asociată cu răspunsul la terapie. Ca urmare a faptului că această mutație a apărut în aproximativ 65% din biopsiile tumorale, această mutație a fost selectată pentru validare într-o altă cohortă separată

de pacienți cu cancer pulmonar. De asemenea, analiza statistică de corelație Pearson a indicat o asociere negativă între prezența mutațiilor în gena *PIK3CA* și subtipul histologic de NSCLC, adenocarcinomul pulmonar ($r=-0.50918$; $p=0.0029$).

Astfel, acest studiu a indicat genele cele mai frecvent alterate în probele de țesut tumoral ca fiind *TP53*, *PIK3CA*, *FLT3*, *CSF1R* și *KDR*, pe când în probele de sânge integral de la aceiași pacienți acestea au fost *TP53*, *KDR*, *PIK3CA*, *ERBB4*, *FLT3* și *CSF1R*. De asemenea, în cadrul acestui studiu au putut fi observate diferențe în ce privește profilul mutațional în diferite grupuri (femei vs bărbați și NSCLC vs SCLC), însă un tipar mutațional specific pentru fiecare tip histologic nu a putut fi observat. Toate aceste rezultate indică faptul că profilul mutațional al patologiei oncologice pulmonare este foarte heterogen, iar numărul de mutații cauzative este foarte mare.

Studiul 3. Validarea variațiilor rs3822214 și rs1042522 identificate și caracterizarea profilului mutațional somatic al cancerului pulmonar prin analiza seturilor de date din TCGA.

Studiul acesta a urmărit validarea mutațiilor observate în studiul NGS anterior. Pe baza observațiilor făcute în urma secvențierii țintite realizate utilizând ADN-ul extras din probe de țesut tumoral și sânge integral de la 32 de pacienți cu cancer pulmonar (NSCLC-cu subtipurile adenocarcinom pulmonar și carcinom pulmonar scuamos și SCLC) pentru secvențierea a 46 gene cu implicare în cancer cuprinse în Ion Ampliseq Cancer Panel, a fost descris profilul mutațional al acestor pacienți. Astfel, prezența uneia din cele două mutații specifice investigate (rs3822214) a fost validată într-o cohortă separată de pacienți cu cancer pulmonar. Mutația c.1621A>C (rs3822214) din gena *KIT* fost validată coresponzător, în cea de-a doua cohortă având o frecvență de 19%, similară cu cea observată în prima cohortă (12.5%). În studiul NGS anterior, cele mai multe alterări din gena *TP53* au fost clasificate ca fiind patogene, iar mutația c.215C>G prezentă la nivelul aceleiași gene este descrisă ca fiind asociată cu răspunsul la terapie. Fiind identificată la 65% din probele biologice ale pacienților incluși în prima cohortă de studiu, această variație a fost selectată pentru validare. Prezența variației rs1042522 nu a fost validată în cea de-a doua cohortă de pacienți, aceasta fiind identificată doar în 11% din probele analizate.

De asemenea, în acest studiu am investigat seturile de date de cancer pulmonar din TCGA pentru identificarea de tipare mutaționale care fie se suprapun, fie sunt diferite de observațiile făcute în acest sens în studiul precedent de secvențiere. Studiul a presupus și urmărirea diferențelor observate pe grupuri diferite - NSCLC (LUAD și LUSC) vs. SCLC și femei vs. bărbați pentru observarea diferențelor sau similarităților cu observațiile din studiul NGS anterior.

Astfel, analiza TCGA a indicat gena *TP53* ca având cele mai multe alterări în fiecare dintre cele trei seturi de date analizate (LUAD, LUSC și SCLC), cu diferențe semnificative privind frecvența acestora. De asemenea, genele *LRP1B*, *PIK3CA*,

CDKN2A, *PTPRD*, *RB1*, *PDE4DIP*, *PCLO*, *KRAS*, *FAT1* și *RELN* au fost identificate ca având frecvente alterări, iar aceste rezultate se suprapun parțial cu cele obținute în studiul anterior. În plus, analiza TCGA a permis identificarea unor frecvențe ale alterărilor genetice diferite pentru fiecare tip histologic. Astfel, în setul de date LUAD, cele mai frecvente alterări genetice au apărut în genele *KRAS*, *CDKN2A*, *KEAP*, *STK11*, *EGFR*, *NF1* și *BRAF*. În LUSC, cele mai frecvente alterări genetice se regăsesc la nivelul genelor *PIK3CA*, *CDKN2A*, *PTEN*, *NF1* și *EGFR*; iar în SCLC, acestea se regăsesc în genele *RB1* și *NOTCH1*.

În cadrul acestui studiu s-a efectuat și analiza de supraviețuire (Kaplan-Meier), iar aceasta a indicat faptul că prezența mutațiilor în genele *NRAS*, *CDKN2A* și *IDH2* este corelată cu supraviețuirea pacienților cu LUAD, pe când în LUSC prezența mutațiilor în genele *TP53*, *STK11* și *CDKN2A* este corelată cu prognosticul pacienților, iar în SCLC prezența mutațiilor în gena *STK11* influențează timpul de supraviețuire al acestor pacienți.

Originalitatea și contribuția inovativă a tezei de doctorat

Datele experimentale din prezenta teză de doctorat au indicat că:

- Variațiile genetice observate în urma secvențierii liniilor celulare pulmonare normale și tumorale s-au suprapus în mare măsură cu caracteristicile descrise în baza de date ATCC. Aceste descoperiri oferă informații importante pentru stabilirea celor mai potrivite linii celulare de cancer pulmonar ce urmează a fi folosite pentru viitoare studii și alegerea liniilor celulare ce pot fi utilizate cu rol de control în laborator, pentru teste funcționale care să permită descoperirea de noi opțiuni terapeutice pentru pacenții diagnosticați cu cancer pulmonar.

- Genele *TP53*, *PIK3CA*, *FLT3*, *CSF1R* și *KDR* apar ca având cel mai mare număr de alterări genetice în probele biologice ale pacienților cu cancer pulmonar. Genele *TP53*, *KDR*, *PIK3CA*, *ERBB4*, *FLT3* și *CSF1R* au fost identificate ca fiind alterate într-un număr mai mare de probe biologice de la pacenți cu cancer pulmonar.

- Au fost observate diferențe în ce privește profilul mutațional pe grupuri diferite - NSCLC (LUAD și LUSC) vs. SCLC și femei vs. bărbați, atât în cohorta de pacenți incluși în experimentul de secvențiere țintită, cât și analizând datele mutaționale din seturile de date de cancer pulmonar din TCGA.

- Prezența uneia din cele două mutații specifice investigate (rs3822214) a fost validată într-o cohortă separată de pacenți cu cancer pulmonar în experimentul de SNP Genotyping assay.

- Analiza statistică de corelație Pearson a indicat o asociere negativă între prezența mutațiilor în gena *PIK3CA* și subtipul histologic de NSCLC, adenocarcinomul pulmonar.

- Analiza de supraviețuire efectuată utilizând datele de TCGA a indicat faptul că prezența mutațiilor în genele *NRAS*, *CDKN2A* și *IDH2* este corelată cu

supraviețuirea pacienților cu LUAD, în genele *TP53*, *STK11* și *CDKN2A* sunt corelate cu prognosticul pacienților în LUSC, iar pentru SCLC, mutații în gena *STK11*.

- În urma analizei TCGA, în setul de date LUAD, cele mai frecvente alterări genetice au apărut în genele *KRAS*, *CDKN2A*, *KEAP*, *STK11*, *EGFR*, *NF1* și *BRAF*. În LUSC, cele mai frecvente alterări genetice se regăsesc la nivelul genelor *PIK3CA*, *CDKN2A*, *PTEN*, *NF1* și *EGFR*; iar în SCLC, acestea se regăsesc în genele *RB1* și *NOTCH1*. Toate aceste rezultate indică faptul că profilul mutațional al patologiei oncologice pulmonare este foarte heterogen, iar numărul de mutații cauzative este foarte mare.

SUMMARY OF PhD THESIS

Next Generation Sequencing (NGS) in personalized medicine

PhD Student **Diana-Cecilia Pop (Bica)**

Scientific supervisor **Prof.dr. Ioana Neagoe**



UMF
UNIVERSITATEA DE
MEDICINĂ ȘI FARMACIE
IULIU HAȚIEGANU
CLUJ-NAPOCA

CONTENT

INTRODUCTION	18
State of the art	
1. Cancer landscape	23
1.1. Cancer worldwide and survival rates – a major concern	23
1.2. Cancer genome landscape	23
1.2.1. Cancer-driver and passenger mutations	24
1.2.2. Non-coding RNAs in cancer	25
1.2.3. Non-coding RNA types and their involvement in cancer	26
1.2.4. Precision medicine in oncology – liquid biopsies	30
2. Lung cancer: diagnosis and management	
2.1. Lung cancer statistics	32
2.2. Lung cancer classification and characteristics	33
3. High-throughput technologies and their applications in cancer	36
3.1. NGS technologies	36
3.2. Cancer screening using NGS	38
3.3. Cancer diagnosis and prognosis using NGS	39
3.4. NGS in precision medicine	39
3.4.1. NGS in the molecular profiling of lung cancer	41
PERSONAL CONTRIBUTION	
1. Working hypothesis/objectives	47
2. General methodology	49
3. Study 1. Mutational landscape of lung normal and cancer cell lines.	
3.1. Context of the study	53
3.2. Working hypothesis/objectives	54
3.3. Materials and methods	55
3.4. Results	
3.4.1. Characteristics of the cell lines	56
3.4.2. Mutations identified in the sequencing experiment	57

3.4.3. Relationship between the ATCC profile and the mutations identified in the sequencing experiment	59
3.5. Discussions	60
3.6. Conclusions	61
4. Study 2. Massively parallel tumor multigene sequencing to assess the mutational landscape of lung cancer patients	
4.1. Context of the study	63
4.2. Working hypothesis/objectives	64
4.3. Materials and methods	64
4.4. Results	
4.4.1. Clinicopathological characteristics of the cohort	66
4.4.2. Mutations identified in the study cohort	70
4.4.3. Statistical analysis	79
4.5. Discussions	81
4.6. Conclusions	84
5. Study 3. Validation of rs3822214 and rs1042522 variants on an additional cohort of lung cancer patients and profiling of somatic variants in lung cancer TCGA datasets.	
5.1. Context of the study	85
5.2. Working hypothesis/objectives	86
5.3. Materials and methods	86
5.4. Results	
5.4.1. Clinicopathological characteristics of the cohort	88
5.4.2. Mutation validation results	90
5.4.3. Genetic alterations in the TCGA datasets	90
5.4.4. Statistical analysis	94
5.5. Discussions	95
5.6. Conclusions	96
6. General discussions	99
7. General conclusions	103
8. Novelty and innovative aspects of the thesis	105
References	107

Keywords: NGS, targeted sequencing, mutational landscape, personalized medicine, lung cancer, precision medicine.

INTRODUCTION

Cancer represents a major problem worldwide, its treatment being efficient only in confined cases, due to the heterogeneity reflected in histology, but mostly in the molecular mechanisms underlying their development and progression.

The evolution of normal cells in different tissues into neoplastic transformed cells is described in literature and includes the genome instability driven by mutations at specific sites, undetermined replicative potential, angiogenesis induction at the tumor site, mechanisms of escaping growth suppressors and immune surveillance, the increase of motility to favor invasion and metastasis, and abnormal energetic metabolism. Therefore, extensive studies are performed to determine the molecular changes that drive these disturbances leading to the malignant transformation of the cells in order to identify potential therapeutic targets for personalized medicine that would increase the response rate to treatment in patients, but mostly would reduce side effects that are a major problem in chemo- and radiotherapy. Given the recent technological advances with proper data analysis, it is possible to diagnose the patient at the molecular level in order to choose the precise treatment scheme to ensure a complete remission or at least favorable outcome for cancer patients. Next Generation Sequencing (NGS) represents one of the advanced technologies used nowadays in research laboratories and in the clinical practice to determine genomic alterations associated with several types of cancer. Lung cancer represents a set of malignancies with two main histopathological classes - small cell lung cancer and non-small cell lung cancer, and the latter includes two main histotypes: adenocarcinoma and squamous cell carcinoma. The heterogeneity of the disease is also reflected high mutational burden and an elevated number of driver mutations that limits the targeted approach in the clinical practice. The present study focused on assessing the mutational landscape of lung normal and cancer cell lines according to the molecular networks involved in lung cancer oncogenesis. Another focus of this study was to determine genetic mutations in patients with different types of lung cancer and to point out the similarities/differences in the molecular profile according to the histopathological classification. The final study aimed to validate the molecular alterations identified in the previous study in another cohort of lung cancer patients and by analyzing the datasets in the TCGA (The Cancer Genome Atlas) database for lung cancer.

STATE OF THE ART

According to GLOBOCAN 2020, cancer represents one notable concern in healthcare, with about 19.3 million cases diagnosed each year. Compared with the cancer statistics in 2018 (GLOBOCAN 2018), where the number of new cases recorded was 18.1 million, and with those from 2012 (GLOBOCAN 2012) where the number of

new cases recorded was 14.1 million, one can notice that there are still major flaws in terms of prevention for this group of diseases. Lung cancer is divided in two main histotypes – small-cell carcinoma (SCLC) and non-small cell carcinoma (NSCLC) – with two main subtypes (adenocarcinoma and squamous cell carcinoma) – which accounts for approximately 85% of cases. As expected, the therapeutic implications and prognostic are different for the two lung cancer groups (Figure 3). The treatment for SCLC patients does not include surgical resection due to its aggressiveness. For NSCLC patients the treatment implies a combination of adjuvant therapy and surgery.

In terms of molecular characteristics of NSCLC adenocarcinoma, TCGA analysis on revealed important mutations at the site of the following genes, the top five most mutated genes being *TP53* (46%), *KRAS* (33%), *STK11* (17%), *KEAP1* (17%), and *EGFR* (14%). The list was completed by *BRAF*, *NF1*, *PIK3CA*, *RB1*, and *CDKN2A*. The study revealed that about 75% of the examined samples had driver mutations in the genes involved in the *RTK/RAS/RAF* signaling pathway (*BRAF*, *EGFR*, *ERBB2*, *HRAS*, *KRAS*, *MAP2K1*, and *NRAS*). Molecular particularities of squamous cell carcinoma (in the NSCLC group) according to several studies indicate various regions with CNA, and these include amplifications of *PDGFRA*, *SOX2*, *FGFR1* and deletions of the *CDKN2A*. NGS can be used in different applications, including cancer research. The approaches used differ in line with the questions that are addressed and the needs of the consumer. NGS technology includes four main steps: sample preparation, sequencing, quality analysis, and data analysis and interpretation. In oncology, precision medicine requires the identification of genetic aberrations that would facilitate the selection of an effective treatment. Profiling the mutational status of a patient or a tumor using NGS led to the necessity of using targeted therapies for tumors. In this respect, identifying mutations that lead to altered signaling pathways and identifying a drug that would block/reverse the effect of the mutations would improve the outcome of these patients.

PERSONAL CONTRIBUTION

The present thesis is divided in three main studies, and the specific objectives are: • to evaluate the molecular profile of lung cancer using NGS technology in lung cancer and normal lung cell lines using a commercially available panel of 46 genes with known involvement in cancer; • to analyze the mutational profile of 32 patients diagnosed with non-small cell lung cancer (lung adenocarcinoma, lung squamous cell carcinoma) and small cell lung cancer in order to identify common mutational patterns for patients with the same histotype; • validation of the most common mutations in lung cancer patients in a separate cohort of 32 matched lung cancer patients.

Study 1 - Mutational landscape of lung normal and cancer cell lines.

The purpose of this study was to investigate the mutational landscape of eight lung cancer cell lines (H1155, Calu6, H123, SKMES 1, H522, M22GB, 1792, and A549) in comparison with the mutations identified in three normal cell lines (WI-38, IMR90, and HFL). The mutations were detected using targeted sequencing approach, based on a commercially available cancer panel including 46 genes with known involvement in cancer.

The main objective of the study was to determine the molecular alterations encountered in frequently studied lung cancer cell lines used in *in vitro* studies. This will establish a basis for subsequent studies conducted in the laboratory to establish altered molecular pathways involved in different hallmarks of cancer and mechanisms that modulate resistance to treatment.

We evaluated the mutational landscape of eight lung cancer cell lines and other three lung fibroblast cell lines. The purpose of this study was to assess the presence of particular mutations in genes commonly mutated in cancers in generally present in the lung cancer cell lines we use for studies in the laboratory in comparison to the standard mutational characteristics of the cell lines according to the ATCC database. Part of the molecular alterations characteristic for the cell lines present in the ATCC database were confirmed in our experiment.

This study aimed the characterization of the mutational landscape of several lung cancer cell lines alongside with other normal lung fibroblast cell lines and to compare the results with the standard characteristics of these cell lines in the ATCC. The genetic variations obtained in this targeted sequencing experiment partially overlapped with the characteristics of the cancer cell lines described in ATCC. These findings offer valuable information about how to establish the appropriate lung cancer cell line for different studies and how to select the control cell line for following studies in our laboratory. Also, this assessment would serve as grounds for subsequent *in vitro* studies of response to therapy and other functional assays that would observe the behavior of these cells under different conditions in an attempt to discover new therapeutic options for lung cancer patients.

Study 2 - Massively parallel tumor multigene sequencing to assess the mutational landscape of lung cancer patients

NGS technology allows researchers the parallel sequencing of millions of DNA fragments at costs that continue to decrease, making the present time 'the era of precision medicine' due to the possibility of performing oncogenic screening to detect genetic/epigenetic alterations in a patient.

The study hypothesis was that the morphological (histological) classification is reflected in the molecular/mutational landscape of lung cancer patients. Therefore, we assumed that for each histotype, there are specific alterations that might be depicted in the sequencing experiment performed using a cancer panel. First of all, we aimed to identify and characterize the genetic mutations in tumor biopsy and matched whole blood samples of patients with lung cancer diagnosis using a 46-cancer genes panel and to identify commonly mutated cancer genes in patients with the same lung cancer type (NSCLC vs. SCLC). We also underlined differences among patients in men vs women groups. Another hypothesis was that the presence of different pathogenic alterations in cancer genes might affect patients' outcome and we performed statistical analyses to validate this hypothesis.

In this study, we assessed the presence of different mutations in lung cancer patients using NGS techniques to test their applicability in the clinical context. Therefore, we assessed the genetic alterations encountered in various cancer genes on the training cohort including 32 patients. The study was also focused on investigating differences and similarities in the mutational profile of lung cancer histotypes (NSCLC vs. SCLC) and men vs women. We also assessed the correlation between different damaging or potentially damaging genetic alterations and the clinicopathological characteristics of the patients investigated in the present study.

In the cohort analyzed, *TP53* was altered in about 93% of SCLC samples while in NSCLC samples the frequency was 87%, in tumor tissue samples. In the blood samples analyzed, *TP53* had an alteration frequency of 81% in NSCLC and 68% in SCLC. In *TP53* most of the variations identified were categorized as pathogenic, and c.215C>G mutation is reported as being associated with drug response. Being identified in about 65% of the tumor samples, this mutation was selected for validation in a different cohort of matched lung cancer patients. The fact that numerous abnormalities in *TP53* could predict drug resistance and worse overall survival is a well documented subject. We also performed a Pearson's correlation analysis and we observed a strong negative correlation between adenocarcinoma and mutated *PIK3CA* ($r=-0.50918$; $p=0.0029$).

Our study indicated *TP53*, *PIK3CA*, *FLT3*, *CSF1R*, and *KDR* as the most frequently genes in lung cancer specimens, while *TP53*, *KDR*, *PIK3CA*, *ERBB4*, *FLT3*, and *CSF1R* were found altered in a higher number of samples. In this study, we also managed to identify dissimilarities in the mutational landscape of women vs. men and NSCLC vs. SCLC. A strong positive correlation between altered *PIK3CA* and lung adenocarcinoma was revealed. All these results indicate various genes with different somatic mutations in our cohort, thus confirming the high mutational burden of lung tumors.

Study 3 - Validation of rs3822214 and rs1042522 variants on an additional cohort of lung cancer patients and profiling of somatic variants in lung cancer TCGA datasets.

The present study focused on validating the mutations and mutational patterns observed in the previous NGS study. Based on the observations made upon the NGS experiment, including whole blood and tumor tissue samples from 32 patients diagnosed with either NSCLC (lung adenocarcinoma and lung squamous cell carcinoma) and SCLC using the commercially available Cancer Panel from Ion Ampliseq that assessed the mutational profile using 46 cancer genes. Therefore, we validated a part of the identified mutations using tumor tissues from a separate cohort of 32 matched lung cancer patients. Moreover, we investigated the lung cancer datasets in the TCGA database to uncover mutational patterns that overlap or are different in the public databases compared to the cohort in the sequencing experiment. We also investigated differences between NSCLC (LUAD and LUSC) and SCLC datasets, and men vs. women to depict any differences or similarities that might overlap with the results observed in the previous study.

The TCGA analysis indicated *TP53* as the most altered gene in all the three datasets, with differences among the datasets. *LRP1B*, *PIK3CA*, *CDKN2A*, *PTPRD*, *RB1*, *PDE4DIP*, *PCLO*, *KRAS*, *FAT1*, and *RELN* were also found as harboring frequent alterations. These results overlap partially with those in the sequencing experiment. In the TCGA datasets, we were able to depict alteration frequencies different for each histotype: LUAD – frequent alterations in *KRAS*, *CDKN2A*, *KEAP*, *STK11*, *EGFR*, *NF1*, and *BRAF*; LUSC – in *PIK3CA*, *CDKN2A*, *PTEN*, *NF1*, and *EGFR*; SCLC – in *RB1* and *NOTCH1* (Figure 3). The TCGA analysis allowed the identification of specific mutational patterns for each dataset – LUAD, LUSC, and SCLC. In the training cohort, we included comparable proportions of SCLC (50%) and NSCLC (50%) patients, unlike the TCGA cohorts where the NSCLC group represented a majority (1097 patients out of 1217).

In *TP53* most of the variations identified were categorized as pathogenic, and c.215C>G mutation is reported as being associated with drug response. Being identified in about 65% of the tumor samples, this mutation was selected for validation in a different cohort of matched lung cancer patients. The presence of this variation in the second cohort was not properly validated, being detected in 11% of the samples. On the other hand, the experiment properly validated c.1621A>C variant in *KIT*, as 6 out of 31 (19%) patients had a positive identification and this proportion was similar to the one observed in the sequencing cohort (12.5%). Additionally, the survival analysis revealed that mutated *NRAS*, *CDKN2A*, and *IDH2* are correlated with survival of patients with LUAD, while in LUSC mutated *TP53*, *STK11*, and *CDKN2A* are correlated with survival time and in SCLC, one gene with alterations (*STK11*) was correlated with survival of these patients.

General conclusions and novelty

Our experimental data reveal that:

- The genetic variations obtained in this targeted sequencing experiment partially overlapped with the characteristics of the cancer cell lines described in ATCC - this assessment would serve as grounds for subsequent *in vitro* studies of response to therapy and other functional assays that would observe the behavior of these cells under different conditions in an attempt to discover new therapeutic options for lung cancer patients.

- *TP53*, *PIK3CA*, *FLT3*, *CSF1R*, and *KDR* as the most frequently altered genes in lung cancer specimens. *TP53*, *KDR*, *PIK3CA*, *ERBB4*, *FLT3*, and *CSF1R* were found altered in a higher number of samples.

- We managed to identify dissimilarities in the mutational landscape of women vs. men and NSCLC vs. SCLC in the training cohort and also in the lung cancer TCGA datasets.

- The SNP Genotyping assay properly validated only one (c.1621A>C in *KIT*) out of the two mutations selected in an additional cohort of lung cancer patients.

- A strong positive correlation between altered *PIK3CA* and lung adenocarcinoma was revealed.

- The survival analysis performed on the TCGA datasets revealed that mutated *NRAS*, *CDKN2A*, and *IDH2* are correlated with survival of patients with LUAD, while in LUSC mutated *TP53*, *STK11*, and *CDKN2A* are correlated with survival time and in SCLC.

- The TCGA study revealed differences in terms of mutation frequencies of different genes for each of the three histotypes, with LUAD harboring frequent mutations in *KRAS*, *CDKN2A*, *KEAP*, *STK11*, *EGFR*, *NF1*, and *BRAF*; LUSC harboring frequent mutations in *PIK3CA*, *CDKN2A*, *PTEN*, *NF1*, and *EGFR*, while in SCLC dataset, frequent mutations appeared in *RB1* and *NOTCH1*.

- All these results indicate various genes with different somatic mutations in our cohort, thus confirming the high mutational burden of lung tumors.