

CERCETĂRI FUNDAMENTALE

DETECTAREA ŞI ANALIZA MUTAŢIILOR SOMATICE ÎN
UNELE PROBE DE ADENOCARCINOM PULMONAR PRIN
TEHNOLOGIA DE SECVENŢIERE *ION TORRENT*

Valentina STRATAN, dr. şt. biol., conf. cerc., Victor SÎTNIC, cerc. şt.,
Valeri ȚUȚUIANU, cerc. şt., Valeriu BÎLBA, dr. şt. med., conf. cerc.,
Sergiu BRENÎȘTER, dr. şt. med., conf. cerc., Vladimir ȘUTKIN, dr. hab. şt. med.,
Ana LOZOVANU, cerc. şt., doctorand, Viorel MUNTEANU, cerc. şt.,
Cristina POPA, cerc. şt. stag.

IMSP Institutul Oncologic

Rezumat

Cancerul pulmonar este cea mai comună cauză a mortalității prin cancer și cel mai comun cancer din lume cu peste 1,5 milioane de decese anual. Explorarea aspectelor genomice, studiul profilului molecular și selectarea terapiilor țintite pentru aceste tumori sunt tendințe actuale și importante în oncologia de precizie. Lucrarea de față descrie profilul molecular și adnotarea clinică a datelor genomice obținute în urma secvențierii prin tehnologia *Ion Torrent* a 8 probe de adenocarcinom pulmonar. Secvențierea a fost efectuată în cadrul IMSP Institutul Oncologic în perioada 2018-2019 cu utilizarea panelului *Ion AmpliSeq Colon and Lung Cancer Research Panel v2* care include 22 gene.

Cuvinte-cheie: secvențiere, NGS, cancer, adenocarcinom pulmonar, genom tumoral, adnotare clinică, terapie țintită

Summary

Lung cancer is the most common cause of cancer mortality and the most common cancer in the world with over 1.5 million deaths annually. The exploration of genomic aspects, the study of molecular profile and the selection of targeted therapies for these tumors are current and important tasks in precision oncology. The present paper describes the molecular profile and clinical annotation of the genomic data obtained from *Ion Torrent* sequencing of 8 lung adenocarcinoma samples. The sequencing was performed within the Oncological Institute in the period 2018-2019 using the *Ion AmpliSeq Colon and Lung Cancer Research Panel v2* which includes 22 genes.

Key-words: sequencing, NGS, cancer, lung adenocarcinoma, tumor genome, clinical annotation, targeted therapy

Резюме

Рак легких является самой распространенной причиной смертности от рака и самой распространенной формой рака в мире с более чем 1,5 миллиона смертей в год. Изучение геномных аспектов, молекулярного профиля и выбор целевой терапии для этих опухолей являются актуальными и важными задачами в области точной онкологии. В статье описывается молекулярный профиль и клиническая аннотация геномных данных, полученных из секвенирования 8 образцов аденокарциномы легкого с применением технологий *Ion Torrent*. Секвенирование проводилось в Онкологическом Институте в 2018-2019 годах с использованием панели *Ion AmpliSeq Colon and Lung Cancer Research Panel v2*, включающий 22 ген.

Ключевые слова: секвенирование, NGS, рак, аденокарцином легкого, раковый геном, клиническая аннотация, таргетная терапия

Introducere. Profilul molecular al cancerului a devenit esențial pentru selectarea și aplicarea terapiilor țintite, iar Secvențierea de Noua Generație (Next Generation Sequencing - NGS) a îmbunătățit considerabil cercetarea din domeniul biomedical și a oferit noi provocări domeniului bioinformatic de analiză efektivă a datelor de secvențiere. Pentru Republica Moldova explorarea genomului tumoral cu utilizarea tehnologiei NGS reprezintă o premieră iar detectarea rapidă și simultană a alterațiilor genetice în 22 de gene implicate în tumorigeneză prezintă relevanță terapeutică pentru cancerul pulmonar. Adenocarcino-

mul pulmonar este recunoscut ca o malignitate complexă, iar diversitatea moleculară și genomică solicită tratamentul centrat pe pacient. Variantele genetice patogene ar putea fi utilizate ca biomarkeri de selecție a tratamentului iar abordarea terapeutică - realizată atât prin medicamente aprobate și validate clinic, precum și prin medicamente care sunt incluse în prezent în studiile clinice. Efectuarea secvențierii unui număr mare de gene la o singură lansare permite descifrarea peisajului genomic al tumorilor, obținerea rezultatelor în timp rapid și adnotarea clinică a țintelor moleculare. Rezultatele obținute în cadrul cercetărilor pot

Tabelul 1.

Ion AmpliSeq Colon and Lung Cancer Research Panel v2	
Tipul probei	Țesut tumoral parafinizat
Aplicare	Detectarea mutațiilor somatice
Gene	KRAS, EGFR, BRAF, PIK3CA, AKT1, ERBB2, PTEN, NRAS, STK11, MAP2K1, ALK, DDR2, CTNNB1, MET, TP53, SMAD4, FBX7, FGFR3, NOTCH1, ERBB4, FGFR1, FGFR2
Perechi de primeri, lungime ampliconi	92 perechi de primeri, 92 ampliconi cu o lungime medie de 162 pb
Cantitatea de ADN utilizată	10 ng
Performanță	Numărul total de read-uri: > 100 000 Acoperirea medie a bazelor: > 500x

servi ca suport pentru aplicarea tehnologiilor NGS în unele cazuri clinice atât în ceea ce vizează oncologia de precizie cât și alte specialități medicale.

Materiale și metode

Pentru realizarea secvențierilor celor 8 probe de adenocarcinom pulmonar au fost utilizate probe de țesut tumoral parafinizat, ADN-ul genomic fiind izolat cu kit-ul *GeneJET FFPE DNA Purification Kit*. A fost aplicată tehnologia *Ion Torrent* pentru panelul *Ion AmpliSeq Colon and Lung Cancer Research Panel v2* care include 22 gene (tabelul 1).

Performanța investigațiilor genomice a fost evaluată apreciindu-se numărul de read-uri și acoperirea medie a bazelor (Coverage) în regiunile de interes. Pentru toate probele a fost obținută o performanță maximă numărul de read-uri depășind 100000 iar acoperirea medie a bazelor fiind peste 500x. Densitatea medie de încărcare a chipului precum și alte detalii ale experimentului sunt prezentate în *figura 1*.

Analiza datelor genomice brute și filtrarea variantelor genetice a fost realizată cu ajutorul platformei *Ion Reporter* și a plugin-ului *Torrent Variant Caller*. Adnotarea și interpretarea clinică a genelor mutante și

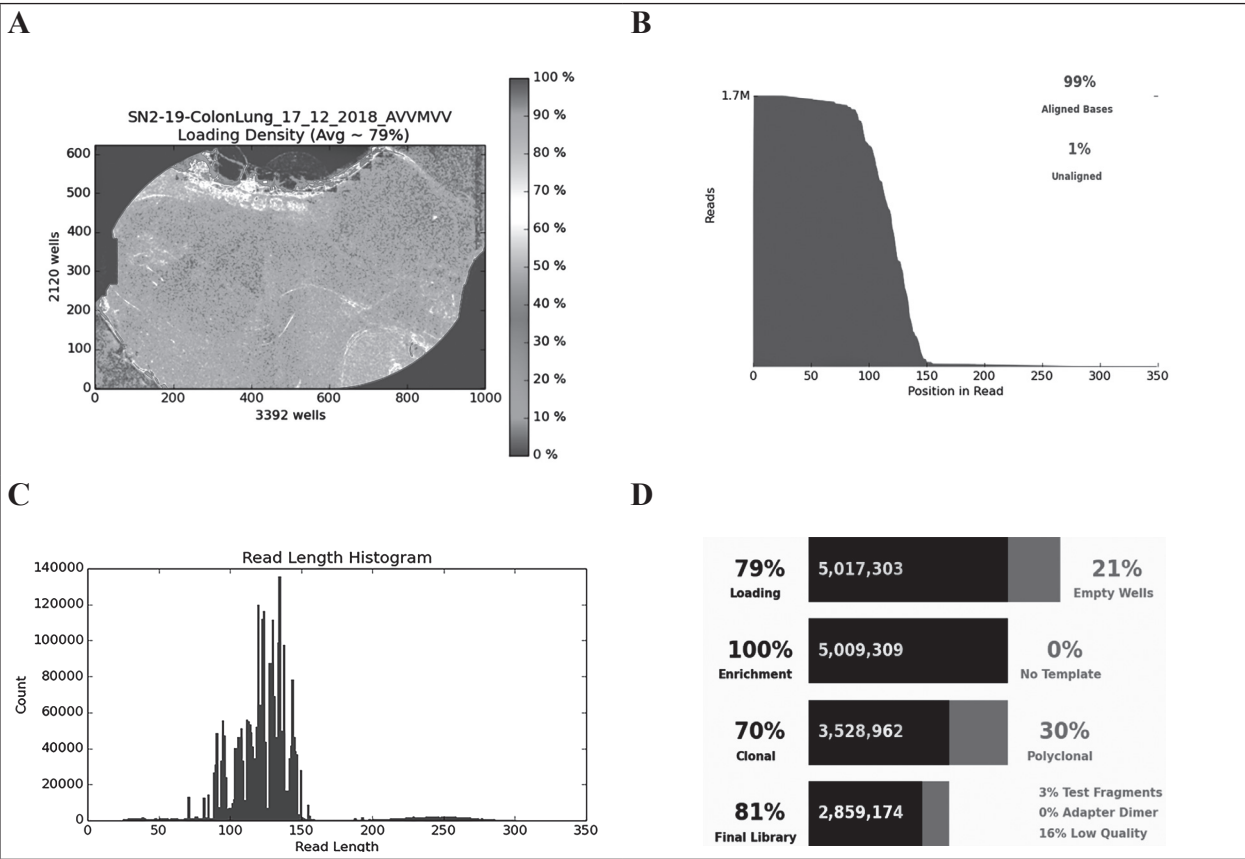


Figura 1. Detaliile experimentului de secvențiere
(A. Densitatea medie de încărcare a chipului; B. Alinierea read-urilor;
C. Lungimea read-urilor; D. Detalii ISP (Ion Sphere Particles))

Tabelul 2.

Mutațiile identificate per probă						
Proba tu- morală	Gena și varianta gene- tică detectată	Patogenitate	Tipul	Efectul	Proteina	Coverage
AP01	MET c.534C>T	Benign	SNV Exonic	Synonymous	p.(=)	3994
	MET c.1124A>G	Benign	SNV Exonic	Missense	p.Asn375Ser	2611
	ERBB2 c.2324_2325 ins ATACGTGATGGC	Pathogenic/ Likely pathogenic	INDEL Exonic	Non Frameshift Insertion	p.Glu770_ Ala771 insAlaTyrValMet	3971
AP02	EGFR c.1498+22A>T	Not reported	SNV Intronic	Unknown	p.?	3995
	TP53 c.880G>T	Pathogenic	SNV Exonic	Nonsense	p.Glu294Ter	3921
	TP53 c.215C>G	Benign	SNV Exonic	Missense	p.Pro72Arg	3999
AP03	DDR2 c.1505-20C>T	Benign	SNV Intronic	Unknown	p.?	4000
	ERBB4 c.421+58A>G	Not reported	SNV Intronic	Unknown	p.?	4000
	KRAS c.35G>C	Pathogenic/ Likely pathogenic	SNV Exonic	Missense	p.Gly12Ala	3967
AP04	FGFR3 c.1953G>A	Not reported	SNV Exonic	Synonymous	p.(=)	
	EGFR c.1498+22A>T	Not reported	SNV Intronic	Unknown	p.?	3993
	TP53 c.215C>G	Benign	SNV Exonic	Missense	p.Pro72Arg	3998
AP05	NOTCH1 c.4732_4734delGTG	Not reported	INDEL Exonic	Non Frameshift Deletion	p.Val1578del	2070
AP06	MAP2K1 c.167A>C	Pathogenic/ Likely pathogenic	SNV Exonic	Missense	p.Gln56Pro	2608
	TP53 c.892G>T	Pathogenic	SNV Exonic	Nonsense	p.Glu298Ter	2627
AP07	ERBB4 c.421+58A>G	Not reported	SNV Intronic	Unknown	p.?	3440
	CTNNB1 c.98C>G	Likely pathogenic	SNV Exonic	Missense	p.Ser33Cys	2662
	AKT1 c.49G>A	Pathogenic/ Likely pathogenic	SNV Exonic	Missense	p.Glu17Lys	1461
AP08	EGFR c.1498+22A>T	Not reported	SNV Intronic	Unknown	p.?	3991
	TP53 c.1024C>T	Pathogenic	SNV Exonic	Nonsense	p.Arg342Ter	3944

identificarea interacțiunilor acestora cu diferiți agenți medicamentoși a fost realizată cu utilizarea *limbajului R* și a librăriei *rDGIdb* (*R Drug Gene Interaction Database*) efectuându-se explorarea mai multor baze de date de profil. Pachetul *rDGIdb* permite efectuarea

căutărilor în diferite surse și identificarea asocierilor dintre producții genici (ținte) și agenții medicamentoși (liganzi). Multe din aceste asocieri includ informații despre tipul interacțiunii dintre ligand și țintă (ex. inhibitor, activator, chaperone, modulator alosteric,

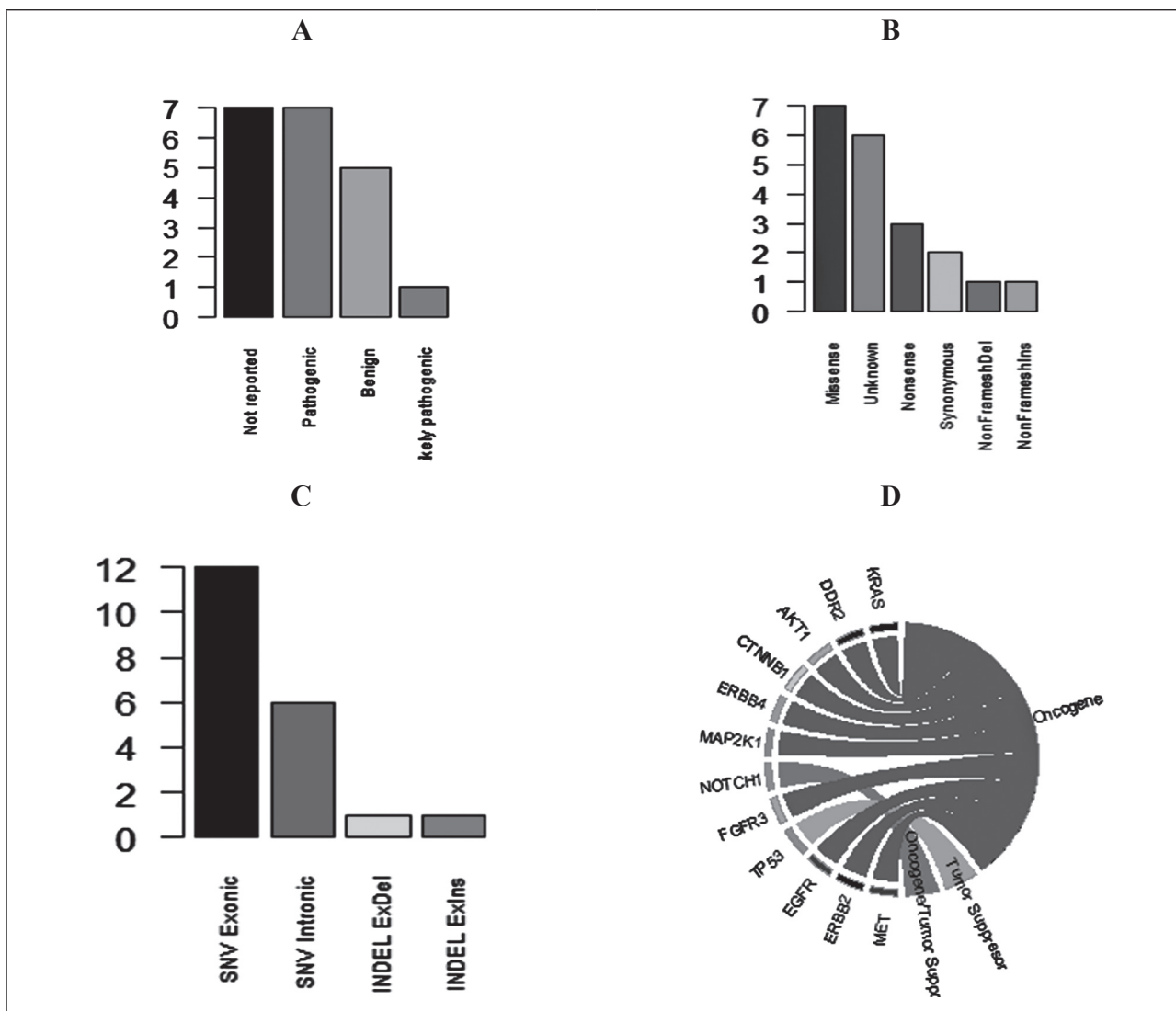


Figura 4. A,B,C – Patogenitatea, efectul translațional și tipul variantelor genetice detectate; D – Funcția genelor mutante

etc.) și despre categoriile genelor de interes (ex. gene din subșetel *Druggable Genome*).

Rezultate și discuții

Secvențierea celor 8 probe de adenocarcinom pulmonar a permis identificarea a 20 mutații în 12 gene: EGFR, ERBB4, TP53, AKT1, CTNNB1, DDR2, ERBB2, FGFR3, KRAS, MAP2K1, MET, NOTCH1. Varianta EGFR c.1498+22A>T a fost identificată în trei probe tumorale, ERBB4 c.421+58A>G și TP53c.215C>G în două probe, în celelalte cazuri detectându-se câte o variantă per probă. În tabelul 2 sunt prezentate mutațiile identificate și caracteristicile acestora.

În figura 4 este prezentată statistica datelor obținute conform următoarelor criterii: patogenitatea, efectul translațional, tipul variantelor genetice și funcția genelor mutante (oncogene, gene supresoare

de tumori, gene cu funcție dublă). Din 8 probe, 6 prezintă mutații patogene sau potențial patogene, celelalte 2 prezintă mutații benigne sau cu patogenitate incertă.

Adnotarea clinică cu utilizarea instrumentului bioinformatic *rDGIdb* a permis identificarea a 796 interacțiuni între cele 12 gene mutante și diferiți agenți medicamentoși. Interacțiunile cu un scor mai mare de 40 care sunt 11 la număr și numărul total de asocieri per sursă pot fi vizualizate în figura.5.

Concluzii.

Cercetările efectuate au permis detectarea unor biomarkeri genetici într-un panel de 22 gene driver ale cancerului și asocierea acestora cu diferiți agenți medicamentoși aprobați clinic sau aflați în faza de studii clinice. Avansarea în secvențierea genomului tumoral și identificarea țintelor moleculare acționa-

A			
Nr	Gena	Medicament	Scor
1	EGFR	Erlotinib	167
2	EGFR	Afatinib	141
3	EGFR	Gefitinib	125
4	EGFR	Cetuximab	110
5	ERBB2	Trastuzumab	99
6	ERBB2	Lapatinib	91
7	KRAS	Cetuximab	73
8	KRAS	Panitumumab	56
9	MET	Crizotinib	52
10	KRAS	Selumetinib	44
11	KRAS	Trametinib	40

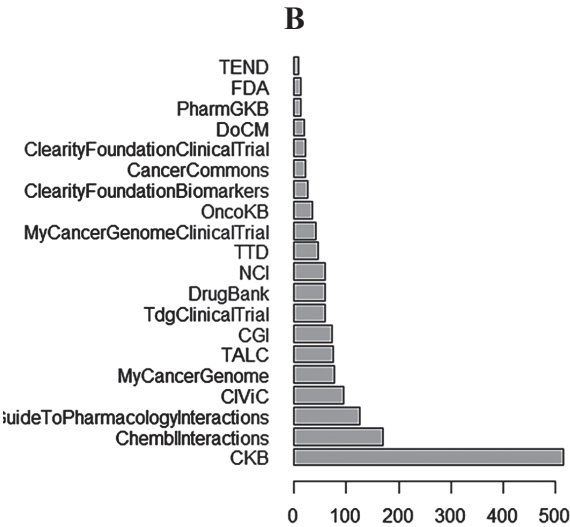


Figura 5. A. Interacțiunile genă-medicament cu scor mai mare de 40;
B. Numărul de interacțiuni genă-medicament identificate în diferite baze de date

bile favorizează selectarea celor mai eficiente terapii țintite în tratamentul cancerului, totuși există unele bariere în ce privește translarea pe larg a acestor tehnologii în practica clinică cum ar fi calitatea probelor, alegerea unei strategii de testare efectivă din punct de vedere al costului și integrarea complexă a profilului molecular cu cel clinic. Pentru a obține beneficii maxime de pe urma explorării genomului tumoral sunt necesare probe biologice calitative, optimizarea strategiilor și utilizarea algoritmilor bioinformatici performanți astfel încât datele moleculare să poată fi eficient interpretate, integrate cu profilul clinic și reflectate în rapoarte clinice accesibile pentru clinicieni.

Bibliografie

1. The Cancer Genome Atlas Research Network, Comprehensive genomic characterization of squamous cell lung cancers, Nature volume 489, pages 519–525 (2012), DOI: <https://doi.org/10.1038/nature11404>;

2. Matan Hofree, John P. Shen, Hannah Carter, Andrew Gross & Trey Ideker, Network-based stratification of tumor mutations, Nature Methods volume 10, pages 1108–1115 (2013), DOI: <https://doi.org/10.1038/nmeth.2651>;

3. The Cancer Genome Atlas Research Network et

al., The Cancer Genome Atlas Pan-Cancer analysis project, Nature Genetics volume 45, pages 1113–1120 (2013), DOI: <https://doi.org/10.1038/ng.2764>;

4. Matthew H. Bailey et al., Comprehensive Characterization of Cancer Driver Genes and Mutations, CELL, Volume 173, ISSUE 2, P371-385.e18, April 05, 2018, DOI:<https://doi.org/10.1016/j.cell.2018.02.060>;

5. Singer J. et al., Bioinformatics for precision oncology, Brief Bioinform. 2019 May 21;20(3):778-788, DOI: 10.1093/bib/bbx143.

6. R. Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

7. Thurnherr T., Singer F., Stekhoven D.J. and Beerenwinkel N. Genomic variant annotation workflow for clinical applications [version 2; referees: 2 approved]. F1000Research 2016, 5:1963. doi:10.12688/f1000research.9357.2

8. Wagner A.H., Coffman A.C., Ainscough B.J., Spies N.C., Skidmore Z.L., Campbell K.M., Krysiak K., Pan D., McMichael J.F., Eldred J.M., Walker J.R., Wilson R.K., Mardis E.R., Griffith M., Griffith O.L. DGIdb 2.0: mining clinically relevant drug-gene interactions. Nucleic Acids Research. 2016 Jan 4;44(D1):D1036-44. doi:10.1093/nar/gkv1165.