

## RECEPȚIONAT

Agenția Națională pentru Cercetare și Dezvoltare

La data: \_\_\_\_\_

## AVIZAT

Secția AȘM \_\_\_\_\_

**RAPORT ȘTIINȚIFIC FINAL**  
**privind executarea proiectului de cercetări științifice aplicative din cadrul**  
**Programului de Stat: *Medicina de precizie în prevenirea, diagnosticul și tratamentul patologiilor***  
**pentru anii 2018-2019**

Proiectul (titlul) **CREAREA SUPORTULUI DECISIONAL ÎN RAPOARTELE DE  
SECVENȚIERE DE URMĂTOAREA GENARȚIE PENTRU VARIANTELE  
SOMATICE A CANCERULUI**

Cifrul Proiectului **18.80.07.16A/PS**

Direcția Strategică ”Sănătate și biomedicină” **80.07**

termen de executare: 31 decembrie 2019

Coordonatorul programului	<b>dr. STRATAN Valentina</b>	_____
Conducătorul proiectului	<b>acad. DUCA Maria</b>	_____
Rectorul Universității de Stat ”Dimitrie Cantemir	<b>dr. hab. HANGANU Aurelia</b>	_____
Președintele Senatului USDC	<b>dr. hab. HANGANU Aurelia</b>	_____

**L.Ș.**

**CHIȘINĂU 2019**

## CUPRINS

1.	<b>Scopul și obiectivele propuse spre realizare în cadrul proiectului</b>	<b>3</b>
2.	<b>Rezultatele științifice obținute în cadrul proiectului</b>	<b>4</b>
2.1.	<i>Tehnologii de secvențiere</i>	<b>4</b>
2.2.	<i>Etape de secvențierii NGS</i>	<b>7</b>
2.3.	<i>Analiza datelor de secvențiere și clasificarea variantelor</i>	<b>8</b>
2.4.	<i>Obiectul și metodele de cercetare</i>	<b>14</b>
2.5.	<i>Evaluarea statutului mutațional și a căilor de semnalizare afectate</i>	<b>15</b>
2.6.	<i>Analiza bioinformatică a rețelelor de semnalizare</i>	<b>16</b>
3.	<b>Cele mai relevante realizări obținute în cadrul proiectului</b>	<b>17</b>
4.	<b>Participarea în programe și proiecte</b>	<b>18</b>
5.	<b>Colaborări științifice internaționale/naționale</b>	<b>18</b>
6.	<b>Vizite ale cercetătorilor științifici din străinătate</b>	<b>18</b>
7.	<b>Teze de doctorat/postdoctorat susținute pe parcursul realizării proiectului</b>	<b>18</b>
8.	<b>Manifestări științifice organizate la nivel național/internațional</b>	<b>19</b>
9.	<b>Aprecierea activității științifice promovate la executarea proiectului</b>	<b>19</b>
10.	<b>Rezumatul raportului cu evidențierea rezultatului, impactului, implementărilor, recomandărilor</b>	<b>20</b>
11.	<b>Concluzii</b>	<b>21</b>
12.	<b>Bugetul proiectului, lista executorilor, lista tinerilor cercetători, doctoranzilor</b>	<b>22</b>
13.	<b>Lista publicațiilor științifice ce țin de rezultatele obținute în cadrul proiectului</b>	<b>23</b>
14.	<b>Participări la manifestări științifice naționale/internaționale</b>	<b>24</b>

## 1. SCOPUL ȘI OBIECTIVELE PROPUSE SPRE REALIZARE ÎN CADRUL PROIECTULUI

În ultimii ani, îmbunătățirile imense a tehnologiilor de secvențiere și a metodelor computaționale au condus la apariția unor platforme de secvențiere de generație următoare (NGS) care au scăzut drastic timpul și costul asociat cu analiza exhaustivă a genomului. NGS permit secvențierea întregului genom și a întregului exom cât și oferă o flexibilitate mare în selectare zonelor de interes pentru a fi analizate, astfel pot fi selectate regiunile genomice care implică variante genetice ce servesc ca ținte terapeutice. Tehnologiile NGS facilitează screening-ul mai multor gene cu probă inițială limitată, un avantaj semnificativ față de platformele convenționale de secvențiere care necesită cantități relativ mai mari de ADN.

Profilul molecular al cancerului a devenit esențial pentru a prezice răspunsul terapeutic la terapiile țintite. Pentru Republica Moldova reprezintă o premieră aplicarea clinică și validarea analizei testului de secvențiere (NGS) integrat, pentru detectarea rapidă și simultană a variantelor unui singur nucleotid, inserțiilor și delețiilor mici și variații ale numărului de copii în 23 de gene implicate în apariția cancerului și care prezintă relevanță terapeutică pentru cancerul pulmonar. Aceste variante sunt abordate terapeutic prin medicamente aprobate de *U.S. Food and Drug Administration*, precum și medicamente care fac parte din ghidurile Rețelei Naționale pentru Cancer SUA, sau sunt incluse în prezent în studii clinice. Studiul respectiv permite obținerea datelor relative despre incidența acestor mutații în populația Republicii Moldova. Efectuarea secvențierii unui număr mare de gene la o singură lansare permite reducerea costului per analiză, obținerea rezultatelor în timp rapid și aplicarea tratamentului personalizat la câteva zile după diagnosticul cancerului. Rezultatele obținute în timpul realizării acestui proiect pot servi ca pilon de bază pentru aplicarea tehnologiilor NGS în rutina clinică atât în ceea ce vizează oncologia personalizată cât și alte specialități medicale.

Dezvoltarea tehnicilor menționate anterior a făcut posibilă acumularea de cunoștințe de către cercetători din domenii diferite, cunoștințe despre modul în care alterările de la diferite niveluri pot influența fiziopatologia, cu importanță semnificativă în domeniul oncologic. În timp ce PCR și alte tehnici comune de biologie moleculară oferă rezultate rapide și ușor de interpretat, metodele de tip “*high throughput*” generează cantități mari de date ce trebuie să treacă prin numeroase stadii de pre-procesare, analiză și integrare. Așadar este necesară folosirea de metode de bioinformatică și algoritmi pentru a manipula aceste date, de la stocare până la interpretare. Acest domeniu multidisciplinar combină cunoștințe din biologia moleculară, genetică, biochimie, biologie sistemică și informatică, devenind obligatoriu pentru descifrarea “*big data*”.

Cu acest scop s-a propus realizarea următoarelor obiective în cadrul proiectului:

- *Crearea și testarea mediului software pentru analiza datelor de secvențiere*
- *Identificarea variantelor genetice din rezultate brute de secvențiere*
- *Modele in silico și machine – learning pentru prezicerea funcțională a vriantelor genetice*
- *Caracterizarea și adnotarea variantelor somatice.*
- *Elaborarea rapoartelor cu utilitate clinică. Generarea statisticilor privind profilele molecular-genetice a tumorilor solide.*

## 2. REZULTATELE ȘTIINȚIFICE OBTINUTE ÎN CADRUL PROIECTULUI

Realizările științifice din ultimele decenii privind structura, expresia și editarea genomului, au determinat o evoluție rapidă a științelor biologice, schimbând radical imaginea și concepțiile noastre despre procesele fundamentale ale materiei vii. Se constată că în momentul în care cercetarea se întâlnește cu tehnologiile de înaltă performanță, au loc progrese majore, ca dovadă fiind informațiile generate de progresele tehnologice recente din domeniul biologiei moleculare, mai precis în genetica funcțională și structurală.

### 2.1. TEHNOLOGII DE SECVENȚIERE

Secvențierea se referă la metodele ce determină ordinea sau succesiunea nucleotidelor în moleculele de ADN și ARN izolate de la animale, plante, bacterii, archaea sau orice altă formă de viață pe Terra. Apariția tehnologiilor de secvențiere a jucat un rol important în analiza secvențelor genomice ale organismelor. Un secvențiator produce fișiere care conțin secvențe de ADN. Aceste secvențe reprezintă șiruri, numite citiri, pe un alfabet format din cinci litere {A, T, C, G, N}. Simbolul N este utilizat pentru a reprezenta o ambiguitate. Secvențierea genetică a dus la dezvoltarea de noi concepte și la descoperirea genelor implicate în afecțiuni considerate anterior idiopatice. Secvențierea ADN poate fi folosită pentru determinarea secvențelor ADN individuale ale genelor, ale regiunilor mari genetice, grupuri de gene sau operoni, cromozomi întregi sau chiar genomuri. Astfel, tehnicile de secvențiere ale ADN și ARN sunt instrumente cheie în multe domenii, inclusiv în cercetarea biomedicală, în medicină legală, arheologie, antropologie etc.

Primele metode de secvențiere și de determinare a secvenței nucleotidice a unor fragmente de ADN au fost elaborate independent de **Allan MAXAM** și **Walter GILBERT** de la Universitatea Harvard în 1976–1977 (calea chimică în care se folosesc reacțiile chimice de clivare a ADN-ului în baze individuale) și **Frederick SANGER** de la Universitatea Cambridge în 1970 (calea enzimatică în care ADN-ul este sintetizat *in vitro* pe baza matriței ADN studiat, în așa fel încât reacția se termină specific în poziția care corespunde unei baze anumite).

Pentru a determina o secvență de nucleotide pe una din căile menționate, ADN-ul este supus seriei de patru reacții separate, fiecare reacție fiind specifică pentru una dintre cele patru baze azotate.

Proc. Natl. Acad. Sci. USA  
Vol. 74, No. 12, pp. 5463-5467, December 1977  
Biochemistry

**DNA sequencing with chain-terminating inhibitors**  
(DNA polymerase/nucleotide sequences/bacteriophage  $\phi$ X174)

F. SANGER, S. NICKLEN, AND A. R. COULSON  
Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 2QH, England  
Contributed by F. Sanger, October 3, 1977

**ABSTRACT** A new method for determining nucleotide sequences in DNA is described. It is similar to the "plus and minus" method [Sanger, F. & Coulson, A. R. (1975) *J. Mol. Biol.* 100, 392-413] but uses a different set of reagents. The method is based on the use of dideoxynucleotides (ddNTPs) which are a stereoisomer of ribose in which the 3'-hydroxyl group is oriented in *trans* position with respect to the 2'-hydroxyl group. The dideoxynucleotides act as chain-terminating in-



Frederick  
SANGER  
(1918 - 2013)

Prin electroforeză produșii de reacție vor migra în patru curse paralele, pe același gel. Urmărind bandă cu bandă, poate fi identificată ordinea nucleotidelor în ADN. Tehnologiile de secvențiere Maxam-Gilbert și Sanger sunt clasificate ca **Tehnologii de Secvențiere de primă Generație**, care permit determinarea secvenței nucleotidice în molecula de ADN și care au fost puse în aplicare încă din anul 1977 după publicarea acestora. Aceste descoperiri au deschis ușa spre studiile codului genetic al ființelor vii și le-a adus inspirația cercetătorilor la dezvoltarea tehnologiilor de secvențiere mai rapide și mai eficiente.

Pentru eficiența înaltă, tehnica propusă de F. Sanger a devenit tehnologia de secvențiere cea mai aplicată și a fost comercializată sub denumirea de „Tehnologia de secvențiere Sanger”, fiind utilizată la scară largă în varianta automatizată. Metoda a dominat industria timp de aproape două decenii și a condus la o serie de realizări monumentale din istoria științei, inclusiv finalizarea secvențierii genomului uman. Pentru aceste performanțe și realizări savanții au primit *Premiul Nobel pentru chimie*, în 1980.

Secvențierea semi-automată Sanger a fost și este folosită cu succes în cercetare și testarea clinică constituind standardul de aur în secvențiere. Cu toate acestea, limitările sale includ randamentul redus și costurile înalte, făcând secvențierea panelurilor cu mai multe gene foarte laborioase și costisitoare. Astfel, în ultimii ani, a existat o trecere de la aplicarea secvențializării automate pentru analiza genomului la metode și tehnici mai avansate, mai eficiente și mai fundamentale.

Necesitatea de a reduce costurile și de a scurta timpul de secvențiere a dus la dezvoltarea tehnicilor de nouă generație, cunoscute ca **Secvențiere de Următoarea Generație (next generation sequencing - NGS)**, care a deschis noi perspective pentru analiza și explorarea genomurilor.

NGS utilizează o singură matrice moleculară, clonal amplificată, care este apoi secvențiată într-un mod masiv paralel. Aceasta a condus la creșterea randamentului cu mai multe ordine de mărime. Prin NGS pot fi efectuate trei nivele de analiză, cu un grad de complexitate în avansare:

1. paneluri de gene de interes;
2. secvențierea exomului (SE)
3. secvențierea genomului (SG).

Toate prezintă avantaje în ceea ce privește secvențierea Sanger prin capacitatea lor de a secvenția cantități masive de secvențe de ADN.

De la introducerea acestor tehnologii, numărul de aplicații și metode care favorizează puterea secvențierii la scară genomică a crescut într-un ritm exponențial. Această revizuire evidențiază concepte, tehnologii și metode recente de secvențiere de generație următoare pentru a ilustra lățimea și profunzimea aplicațiilor și a domeniilor de cercetare care determină progresul în genomică.

Platformele principale utilizează șabloane amplificate clonale, care nu sunt afectate de pierderile arbitrare ale secvențelor genomice și care sunt inerente în metodele de clonare bacteriană.

Există patru etape primare în procesele chimice ale NGS: terminarea ciclică reversibilă, secvențierea prin ligare, pirosecvențiere și secvențierea în timp real, care sunt descrise în această revizuire.

Pentru a apela variante de secvențe în genom, citirile NGS sunt aliniate la o secvență de referință utilizând diferite instrumente de cartare bioinformatică. Progresul major oferit de NGS este capacitatea de a produce un volum enorm de date cu costuri reduse - în unele cazuri peste un miliard de citiri scurte pe instrument. Această caracteristică extinde domeniul experimentării dincolo de determinarea ordinii bazelor. Aplicarea cea mai largă a NGS poate fi *re-secvențierea* genomilor umane pentru a spori înțelegerea modului în care diferențele genetice afectează sănătatea și bolile. Varietatea caracteristicilor NGS face posibil ca mai multe platforme să coexiste pe piață.

Creșterea uimitoare a producției de date și scăderea costului pe bază secvențiată a fost determinată în principal de creșterile paralelizării tehnologiilor de secvențiere cu citire scurtă, cum ar fi platformele Illumina și Ion Torrent.



**Platforma Illumina** asigură secvențierea în pereche, în care fiecare capăt al aceleiași molecule de ADN este secvențiat la lungimea completă de citire. Deoarece dimensiunea aproximativă a inserției este cunoscută, informațiile din perechi îmbunătățesc foarte mult ratele de aliniere unice, comparativ cu citirile singulare. Dominanța platformei Illumina atât în literatura de specialitate, cât și în cantitatea de date de pe platforma depusă la Arhiva de citire a secvențelor demonstrează eficiența și puterea acesteia.

Illumina s-a dovedit a fi o tehnologie puternică atât pentru abordări de secvențiere (cum ar fi secvențiere între genomul întreg, cât și pentru exom-uri întregi), expresia genelor (cum ar fi secvențiere ARN (RNA-seq) și secvențiere de imunoprecipitare cromatină (ChIP-seq).

**Platforma Ion Torrent**, distribuită de Life Technologies din 2011, aduce o abordare unică în ceea ce privește secvențierea de nouă generație. Cu toate că tehnologia se bazează în continuare pe secvențierea prin sinteză, aceasta nu necesită dNTP marcate fluorescent sau chemiluminiscente și o detecție bazată pe imagine a nucleotidelor incorporate. În loc de depistarea pirofosfatului eliberat în timpul incorporării nucleotidelor, chip-ul Ion Torrent funcționează ca un pH-metru, sesizând schimbările subtile de pH ce pot să apară la formarea unei legături fosfodiesterice în timpul elongării catenei de secvențiere.

Fluxul de lucru pornește odată cu efectuarea librăriiilor de ADN. ADN genomic este în primă fază fragmentat în catene de 200-400 pb, capetele sunt reparate, iar librăria este clonal amplificată și purificată. Apoi se atașează adaptorii specifici la capetele fragmentelor, iar golurile sunt completate.

Tehnologiile de secvențiere continuă să evolueze și să se îmbunătățească, numărul acestora crescând în ultimii cinci ani. Astfel, la moment distingem o răspândire largă a tehnologiilor de secvențiere de generația a doua caracterizate prin necesitatea de pregătire a băncilor de secvențiere amplificate, înainte de a începe secvențierea clonelor amplificate de ADN.

Astfel, eforturile din ultimul deceniu, au catalizat dezvoltarea tehnologiilor NGS, care permit determinarea secvenței nucleotidice a unei molecule de ADN sau ARN, precum și a secvenței de aminoacizi a proteinei codificată de genă (în cazul în care ea nu a fost identificată). NGS prezintă un potențial major privind îmbunătățirea posibilităților clinice de diagnostic și de terapie țintită. În rezultat, studierea profilului molecular, privind numeroase patologii, a devenit mai accesibilă și permite descoperirea unor terapii eficiente, dar și a biomarkerilor pentru diagnostic timpuriu, stratificare și prognostic.

Cu părere de rău, tehnologiile NGS nu sunt capabile să citească secvența completă a ADN-ului genomic, fiind limitate la secvențe scurte de ADN. Această limitare rămâne un punct slab, în special în proiectele de asamblare a genomurilor, care necesită și resurse computaționale înalte. Drept urmare, asistăm la apariția tehnologiilor de **Secvențiere de a Treia Generație**, clasificate ca tehnologii de

secvențiere de o singură moleculă, deoarece permit secvențierea unei singure molecule de ADN sau ARN, fără crearea unei librării amplificate, precum și generarea unor citiri mult mai lungi, într-un timp mai scurt.

Prin **Secvențierea de Generație a Treia** numită și secvențierea "cu citiri lungi" (*long read sequencing*), într-un timp de câteva ore pot fi secvențiate, cu mare acuratețe, molecule întregi de ADN, chiar până la 100 mii de nucleotide, dar costurile acestor noi tehnologii sunt încă foarte înalte. Cele mai cunoscute tehnologii de secvențiere de generația a treia au fost comercializate de compania **PacBio** (*single-molecule real-time (SMRT) sequencing*) și de compania **Oxford Nanopore Technologies** (ONT) (secvențierea nanopore).

Aceste două tehnologii de secvențiere disponibile produc citiri de mii de baze pe citire.



Ambele utilizează secvențialarea cu o singură moleculă, deși cu metode de detectare foarte diferite. Astfel, Oxford Nanopore folosește nanopori pentru detectare, în timp ce Pacific Biosciences utilizează detectarea optică a unei reacții de secvențiere prin sinteză, care are loc în interiorul unui ghid de undă în mod zero.

## 2.2. ETAPE DE SECVENȚIEREI NGS

Timpul necesar pentru a realiza un test NGS variază foarte mult. Acesta depinde de măsura de interogare a genomului (de ex. panel de gene, exom sau genom), nivelul de automatizare și procedurile utilizate în prepararea librărilor, metodei de secvențiere, aliniere, apelare a variantelor, filtre și/sau evaluarea clinică.

În general tehnologiile de secvențiere NGS implică mai multe etape majore cum ar fi pregătirea probei, secvențierea și analiza datelor (Figura 1). Produsul de intrare pentru majoritatea platformelor NGS este reprezentat de un set de fragmente scurte de ADN (100-500 perechi de baze), flancate de adaptorii specifici platformelor. Astfel, procesul începe cu extragerea ADN-ului genomic, iar în cazul unor abordări specifice, cum ar fi panourile țintă și secvențierea exomului (SE), se includ și strategii de îmbogățire pentru focusarea asupra unui subset de ținte genomice. Pentru a converti proba de ADN în formatul corespunzător pentru secvențiere sunt necesare o serie de etape de procesare.

NGS poate fi efectuată pentru orice probă biologică care conține ADN, atâta timp cât calitatea și cantitatea ADN-ului sunt suficiente. De obicei, laboratoarele de secvențiere, în dependență de platforma utilizată, specifică tipul și calitatea probelor pentru care au validat metodele de secvențiere.

Generarea librărilor de ADN este procesul de creare a fragmentelor randomizate de ADN, de o anumită mărime, ce vor conține secvențe adaptorii la ambele capete. Adaptorii sunt complementari PCR-ului specific, platformei și amorselor de secvențiere. Fragmentarea ADN-ului poate fi realizată prin mai multe metode, fiecare prezentând atât părți tari cât și slabe. Pentru majoritatea platformelor, anterior secvențierii, este necesară amplificarea PCR a librării.

Barcodarea se referă la etichetarea moleculară a probelor cu coduri de secvență unice, de obicei constând din trei sau mai multe perechi de baze. Aceasta permite amestecarea probelor biologice, ceea ce contribuie la reducerea costului de procesare a unei probe.



Figura 2. Etape principale în procesul de secvențierea de următoare generație

Numărul de probe care pot fi amestecate vor depinde de acoperirea regiunii de interes care trebuie secvențiată. Barcodurile pot fi parte din adaptori sau pot fi adăugate în etapa de îmbogățire PCR inclusă în majoritatea protocoalelor de secvențiere.

Dacă nu se face secvențierea genomului, genele sau regiunile de interes trebuie izolate înainte de secvențiere. Țintele pot fi atât câteva gene de interes (de ex. câteva gene asociate cu o patologie specifică), până la întregul exom - toți exonii codificatori de proteine cunoscuți. Îmbogățirile țintite, cel mai des, sunt bazate pe metode PCR și de hibridizare a oligonucleotidelor.

Platformele comerciale au dezvoltat tehnologii care au o capacitate de secvențiere de milioane de fragmente de ADN în paralel. Platformele comerciale disponibile la moment sunt bazate pe abilitatea de a efectua numeroase reacții chimice în paralel, astfel încât să permită în final analiza produselor individuale. Fiecare platformă se caracterizează prin anumite particularități specifice privind mecanismele și procesele chimice, care stau la baza secvențierii și determină capacitatea totală a secvențelor, lungimea citirilor de secvență, timpul de rulare pentru secvență și în final, calitatea și acuratețea datelor rezultate. Procesele chimice includ secvențierea prin sinteză, secvențierea prin ligare cu terminatori reversibili sau detectarea ionilor. Toate platformele posedă parametri valoroși pentru interesele laboratorului, inclusiv tipul de probe analizate, mărimea instrumentului, costul, timpul de rulare, lungimea citirilor și costul per probă. Alegerea platformei și a lor specifice depinde de scopurile propuse. La finalizarea secvențierii, citirile secvențelor rezultate sunt procesate prin metode computaționale pentru detectarea variantelor ADN.

### 2.3. ANALIZA DATELOR DE SECVENȚIERE ȘI CLASIFICAREA VARIANTELOR

Avantajul principal al tehnologiilor de următoarea generație față de metodele convenționale reprezintă generarea volumelor mari de secvențe. Pentru a transforma volumele masive de date de



înalță performată în date accesibile este necesară organizarea și descifrarea lor cu ajutorul metodelor bioinformatic.

Analiza datelor NGS este deseori împărțită în trei faze distincte: analiza primară, secundară și terțiară. *Designul* și optimizarea analizei bioinformatic mai întâi implică analiza primară și secundară a datelor derivate de platforma de secvențiere.

Faza primară de analiză este dezvoltată de producătorul tehnologiei și este specifică platformei. Analiza primară include secvențierea care produce un set de citiri cu scoruri de calitate, depozitate în fișiere electronice.

În timpul analizei secundare, citirile amestecate sunt separate și asociate cu proba pacientului respectiv. Ulterior, citirile sunt cartate și aliniate la genomul de referință. Analizele primară și secundară, tipic, sunt automatizate.

Analiza terțiară implică adnotarea și evaluarea clinică a variantelor secvențelor pentru identificarea secvențelor relevante pentru pacient și indicațiile clinice pentru testele ulterioare. Analiza terțiară include atât procese automatizate cât și manuale.

*Analiza primară* are loc la instrumentul de secvențiere și tipic implică procesul de “apelare a bazelor” pentru generarea unui fișier ce conține un set de secvențe numite “citiri” asociate cu scoruri de calitate a bazelor. În acest scop sunt utilizate două formate de bază: (1) formatul FASTQ care conține citiri cu scoruri de calitate per bază și (2) formatul BAM care reprezintă un format de fișiere de aliniere binare și poate să conțină citiri cartate sau ne-cartate la o asamblare genomică de referință. Ulterior, datele trec o etapă de control a calității în care citirile sunt filtrate pentru a le elimina pe cele care nu îndeplinesc criteriile producătorului platformei sau criteriile stabilite de laborator înainte de aliniere la genomul de referință. Pe parcursul acestui proces, capetele 5' și/sau 3' ale citirilor sunt tăiate automatizat de către algoritmul software a instrumentului de secvențiere sau sunt manual ajustate în acord cu criteriile stabilite de laborator, deoarece scorurile caracteristice pentru apelarea bazelor sunt deseori mai joase la capetele terminale a citirilor.

*Analiza secundară* reprezintă procesul de aliniere a citirilor la secvența de referință și generarea variantelor. Dacă mai multe probe a pacienților sunt amestecate (fiecare probă individuală este etichetată pentru identificare), sau sunt multiplexate pe parcursul procesului de secvențiere, este nevoie ca datele rezultate să fie de-multiplexate pentru separarea secvențelor pacientului anterior analizei. Citirile pot să se alinieze în mai multe regiuni a genomului de referință din mai multe raționamente. De exemplu, citirile pot să se alinieze identic în loci multipli dacă aceștia derivă din regiuni înalt omologe a genomului. Alternativ, acestea pot să se alinieze la una sau mai multe locații nespecifice conducând la rezultate fals-pozitive. Cu toate acestea devine clar că cu cât asamblările de referință sunt mai complete, cu atât mai eficientă devine alinierea. Identificarea robustă a inserțiilor, delețiilor (INDELS) este de asemenea destul de dificilă. Alinierea inconsecventă pot conduce la rezultate VSN (variante de o singură nucleotidă) sau INDEL false sau greșite în procesul de aliniere inițială a citirilor. Instrumentele recomandate, de ex. **Genome Analysis Toolkit** efectuează o etapă de post-procesare care constă în realinierea regiunilor care potențial conțin INDEL.

Diferențele dintre asamblarea de referință și citirile pacientului sunt identificate în timpul alinierii. Mai multe programe software care utilizează o varietate de algoritmi sunt disponibile pentru a evalua probabilitatea dacă varianta este prezentă sau absentă. Acești algoritmi de obicei utilizează mai multe metode care pot să includă enumerarea numărului de citiri asociate cu fiecare alelă după setarea corespunzătoare a pragurilor corespunzătoare și a calității de cartare, informații anterioare despre variante, frecvențele alelelor, proprietățile platformei de secvențiere și datele de dezechilibru a *linkage-*

ului. Aceste abordări au avantaje pentru secvențele diploide a probelor, însă utilizarea acestora pentru cromozomii sexuali într-o singură copie pot fi mai problematice.

Lista variantelor individuale cel mai des este reprezentată în formatul unui fișier digital. Formatul VCF este cel mai utilizat în comunitatea clinică pentru reprezentarea variantelor alelelor, însă implementarea și informația de adnotare pe care o conține variază mult între laboratoare. Formatul VCF a fost dezvoltat în proiectul *1000 Genomuri*, reprezentând formatul generic pentru stocarea variantelor secvențelor de ADN, combinând împreună variantele de o singură nucleotidă (VSN), INDEL-urile și variantele structurale. Acest format conține o secțiune pentru antet și alta pentru datele variantelor secvențelor. Antetul permite adaptarea informației standardizate la datele secvențelor. Alte formate, precum GVF (Genome Variant Format) sunt la fel utilizate. Formatele VCF și GVF permit reprezentarea genotipului și oferă o flexibilitate adițională pentru descrierea adițională a atributelor, esențială pentru analiza în aval, privind variantele.

Este important de notat că pentru o specificare dată a fișierului (de ex. VCF), reprezentarea conținutului precum și utilizarea diferitor metode de aliniere a și apelare a variantelor, sau metode de descriere a variantelor în fișier pot să varieze semnificativ la diferite laboratoare. Lipsa unor cereri standardizate pentru fișierele de tip VCF reprezintă un obstacol semnificativ în compararea și schimbul de fișiere de tip VCF între laboratoare.

*Analiza terțiară* utilizează rezultatele analizei secundare pentru filtrarea, prioritizarea, identificarea și clasificarea variantelor semnificative. Aceste procese încep cu adnotarea, procedul de colectare și asociere a informațiilor disponibile privind variantele particulare. Variantele adnotate pot fi apoi sortate, filtrate și prioritizate utilizând reguli personalizate (elaborate în interiorul laboratorului) pentru determinarea variantelor relevante la indicațiile de testare. Cu toate că unele adnotări sunt atribuite direct în faza de analiză secundară (de ex. calitatea variantelor și *depth coverage*), cea mai mare parte de date derivă din sursele informaționale externe, informații care nu apar direct din secvența derivată de procesul de secvențiere (de ex. consecințele funcționale cunoscute sau prezise, frecvența în populație și/sau relația genei cu patologia sau fenotipul).

Aceste adnotări pot fi obținute deseori printr-un proces automatizat în urma utilizării instrumentelor bioinformatiche comerciale sau *open source*.

Lista prioritizată de variante genomice, genele asociate cu acestea și efectele funcționale prezise sunt supuse unei evaluări clinice pentru identificarea variantelor relevante și includerea acestora în raportul final al laboratorului. În general aceste etape necesită o analiză și revizuire manuală a variantelor candidate de către un expert capabil de a oferi o apreciere profesională privind relevanța acestora pentru fenotipul pacientului sau a patologiei și să facă o decizie finală privind relevanța dar și limitările și interpretările testului.

Timpul necesar pentru a evalua o variantă candidată poate să varieze semnificativ. O variantă bine caracterizată în literatură, asociată și relevantă pentru testare, poate fi rapid și ușor interpretată în contextul fenotipului pacientului. Variantele cu reprezentarea limitată în literatură însă cu rezultate și date consistente de asociere clinică (de ex. datele de segregare, corelări structurale și/sau funcționale, prezența sau absența acestora în gene asociate cu patologii) pot lua mai mult timp pentru evaluare și interpretare.

Filtrarea și prioritizarea variantelor poate fi automatizată în viitor prin utilizarea algoritmilor *machine learning*, dar și alți algoritmi, însă analiza profundă a variantelor și genelor cu includerea evaluărilor literaturii rămâne a fi efectuată manual pentru ceva timp. În efectuarea unui diagnostic, clasificarea variantelor formează baza deciziilor clinice, unde calitatea clasificării variantelor este critică

pentru bună starea și rezultatele tratamentului. Fără interpretarea minuțioasă și evaluarea evidențelor, rezultatele secvențierii reprezintă date fără interes bio-medical.

De aceea atunci când vine vorba de clasificarea variantelor este crucial ca clinicienii să fie încrezuți și să înțeleagă rezultatele și evidențele oferite de laboratorul genetic de diagnostic. Variantele din gene cu o asociație față de patologie stabilită sunt prioritizate în baza patogenității prezise în contextul prezentării clinice a pacientului în timpul evaluării manuale și a clasificării.

Numeroase laboratoare clinice clasifică variantele în cinci categorii discrete: (i) variante benigne; (ii) variante probabil benigne; (iii) variante cu semnificație incertă, (iv) variante probabil patogene și (v) variante patogene. Este important de notat că clasificarea clinică nu este separată complet de etapele de filtrare și prioritizare.

De exemplu, setul de variante poate fi filtrat devreme în timpul analizei din cauza cunoașterii lipsei asocierii dintre genă și condițiile medicale. Există o practică standard în întreaga industrie de diagnostic genetic pentru fiecare companie sau laborator de cercetare, să-și dezvolte și să folosească propriul sistem intern de clasificare a variantelor însă acest lucru poate deveni destul de confuz, mai ales când clasificările rezultă de la mai multe companii sau laboratoare.

În acest context Colegiul American de Genetică și Genomică Medicală (ACMG) vine cu un set de standarde și instrucțiuni pentru interpretarea variantelor secvențelor, folosind dovezile științifice din populațiile analizate și bazele de date specifice genelor/patologiilor, utilizarea instrumentelor de predicție *in silico* și literatura științifică corespunzătoare. Pentru a clasifica variantele genetice obținute a fost elaborată o schemă cu cinci nivele: (1) variantă patogenică; (2) variantă probabil patogenică; (3) variantă cu semnificație incertă; (4) variantă probabil benignă și (5) variantă benignă. Ca fundament de notare, este folosit un sistem de punctaj, care descrie cantitatea și calitatea dovezilor necesare pentru a clasifica o variantă genetică.

Variante patogene și probabil patogene sunt considerate cauze reale ale bolii pacientului (Tabelul 1).

Tabelul 1. **Variante patogene**

<b>Puncte necesare:</b>	
<b>1 punct</b>	
Mutație bine studiată printr-un larg consensus din domeniu asupra patogenității mutației. Tipic este bine stabilită segregarea familială și mai multe publicații suportă patogenitatea	1 p.
<b>sau cel puțin 5 puncte</b>	<b>Obligatoriu a sau b:</b>
Segregare pozitivă cu patologia ( $\geq 2$ familii) sau cel puțin 5 pacienți ne-înrușiți cu aceeași variantă și fenotip	2 p.
$\geq 5$ cazuri cu aceeași variantă și fenotip raportate comunității științifice medicale	1 p.
<b>Puncte adiționale:</b>	
Varianta este nouă sau foarte rară în populațiile de control	1 p.
S-a stabilit că pierderea funcției genei provoacă un mecanism de patogenitate; există evidențe pentru asociațiile genotip-fenotip	1 p.
Varianta missense este precisă a fi dăunătoare de majoritatea instrumentelor de prezicere <i>in silico</i> și/sau sunt bine stabilite mutații paraloage	1 p.
Varianta este considerată dăunătoare (variantă de tip substituție sau indel în splice-site (+/-1, 2), nonsense sau frameshift)	1 p.
Funcție deficientă a proteinei demonstrată pe modelele animale cu mutația echivalentă	1 p.
Evidențe bine stabilite privind existența mutațiilor paraloage	1 p.
O altă mutație bine stabilită în același codon	1 p.
Alte date stricte care suportă clasificarea patogenică	1 p.

Varianta este bine stabilită drept cauzală a bolii din bazele de date și în literatura științifică, precum și există un consensus larg asupra patogenității variantei. În aceste cazuri se verifică segregarea în familie și publicațiile care suportă patogenitatea. Aceste informații genetice pot fi utilizate independent în deciziile clinice și în evaluarea riscului pentru membrii familiei.

Variante probabil patogene. Varianta identificată este considerată cauza probabilă a bolii pacientului. Aceste informații ar trebui utilizate cu precauție pentru luarea deciziilor clinice, deoarece există un anumit grad de incertitudine (Tabelul 2).

Pentru aceste variante există o corelație clară dintre genotip și fenotip și este esențial de colectat informații despre fenotipul pacientului, ceea ce ajută la determinarea patogenității probabile. Varianta rezultă în trunchierea prematură (produs proteic incomplet) a produsului genei, cu stabilirea pierderii funcției ca mecanism de patogenitate pentru boala suspectată a pacientului.

Alternativ, varianta poate prezenta o substituție de aminoacizi (tip *missense*), care este prezisă a fi dăunătoare de majoritatea instrumentelor *in silico*. În plus, varianta este nouă sau foarte rară în populația de control.

*Variante cu semnificație incertă.* Aceste variante au caracteristici de mutații independente cauzative a patologiei, însă evidențele sunt insuficiente sau sunt contradictorii. Varianta tipic este foarte rară și este prezisă a fi dăunătoare, iar gena are o asociere cu fenotipul bolii pacientului.

Tabelul 2. **Variante probabil patogene**

<b>Puncte necesare:</b>	
<b>2 puncte</b>	
Alterațiile rezultă în trunchierea prematură (de ex. <i>frameshift</i> , <i>nonsense</i> , <i>splice site (+/-1,2)</i> ) și funcția genei a fost stabilită a fi implicată în mecanismele de patogenitate pentru boala pacienților.	1 p.
Varianta este nouă sau foarte rară în populațiile de control	1 p.
<b>sau cel puțin 4 puncte</b>	
Există o corelație clară genotip-fenotip	1 p.
Varianta este nouă sau foarte rară în populațiile de control	1 p.
Varianta de tip <i>missense</i> este prezisă a fi dăunătoare de majoritatea instrumentelor de prezicere <i>in silico</i>	1 p.
Varianta a fost identificată în $\geq 2$ indivizi cu aceeași manifestație a bolii	1 p.
Evidențe bine stabilite privind existența mutațiilor parologe	1 p.
Alterație <i>de novo</i> pentru o nouă boală din cadrul familiei	1 p.
Varianta este considerată dăunătoare (variantă de tip substituție sau indel în <i>splice site (+/1,2)</i> , <i>nonsense</i> , <i>frameshift</i> ) identificată în gene cu evidențe slabe pentru cauzalitatea patologiei	1 p.
Funcție deficientă a proteinei demonstrată pe modelele animale cu mutația echivalentă	1 p.
O altă mutație bine stabilită în același codon	1 p.
Alte date stricte care suportă clasificarea probabil patogenică	1 p.

*Variante benigne și probabil benigne.* Variantele de acest tip nu sunt considerate a fi cauza patologiei testate/studiate. Evidențele arată că varianta nu segregă cu patologia familiei în care doi sau mai mulți indivizi sunt afectați (Tabelul 3).

Tabelul 3. Variante benigne

<b>Puncte necesare:</b>	
<b>2 puncte</b>	
Nu are loc segregarea cu boala familiei, sau familiile cu doi sau mai mulți membri afectați	1 p.
orice alt criteriu descris mai jos	1 p.
<b>sau cel puțin 4 puncte</b>	
Frecvența alelei minore în populația de control (1000G și ExAC) este considerabilă. Prevalența bolii trebuie luată în considerație.	1 p.
Variantele homozigote a genei nu au asocieri cu patologia	1 p.
Co-ocurența cu o mutație patogenică în aceeași genă sau în altă genă, care descriu clar fenotipul probandului	1 p.
Lipsa asocierilor în studiile mici de tip caz-control	1 p.
Majoritatea instrumentelor <i>in silico</i> prezic că substituția este benignă	1 p.
Alte date care suportă clasificarea benignă	1 p.

**Variante probabil benigne.** Aceste variante nu sunt probabil cauza bolii testate. Luând în considerație prevalența bolii și penetranța, frecvența alelei minore (MAF) în populația de control este considerabilă (MAF<0.001) (Tabelul 4.).

Un factor important reprezintă numărul de analiști experimentați în analiza variantelor genomice precum și capacitatea computațională a laboratorului. Evaluarea clinică, cel mai des, reprezintă componenta care consumă cel mai mult timp în procesul de analiză a rezultatelor. Cele mai multe cazuri conțin un număr de variante care nu sunt bine descrise în literatură de a fi asociate cu patologii, însă pot fi asociate cu fenotipul pacientului. Prin urmare, timpul necesar pentru evaluarea clinică a acestor variante poate varia între 30 minute și ore pentru fiecare variantă genetică care necesită o evaluare manuală pentru a efectua o afirmație de clasificare (benignă, probabil benignă, variantă cu semnificație necunoscută, probabil patogenică, patogenică).

Pentru evaluarea condițiilor ereditare, uneori sunt necesare datele clinice și testarea genetică adițională a membrilor de familie. La moment, timpul necesar pentru a evalua un test NGS poate varia între ore și săptămâni.

Tabelul 4. Variante probabil benigne

<b>Puncte necesare:</b>	
<b>1 punct</b>	
Frecvența alelei minore în populația de control (1000G și ExAC) este considerabilă, MAF<0.001. Prevalența bolii trebuie luată în considerație.	1 p.
<b>sau cel puțin 2 puncte</b>	
MAF<0.001 în populația de control, însă varianta este detectată în controalele sănătoase, fără asocieri cu patologia în studiile caz-control.	1 p.
Variantele homozigote a genei nu au asocieri cu patologia	1 p.
Co-ocurența cu o mutație patogenică în aceeași genă sau în altă genă, care descriu clar fenotipul probandului	1 p.
Majoritatea instrumentelor <i>in silico</i> prezic că substituția este benignă	1 p.
Alte date care suportă clasificarea benignă	1 p.

Etapa finală a analizei în constituie elaborarea raportului clinic rezultat. Raportul clinic cu rezultatele NGS trebuie să conțină informații suficiente pentru a comunica rezultatele testului și limitările acestuia. Provocarea în raportarea rezultatelor NGS o constituie balanța dintre informația necesară, utilă pentru medic, cu un nivel corespunzător de detalii astfel încât utilizarea și limitările

analizei să fie clar înțelese și/sau medicul sau alt utilizator să fie ușor direcționat spre alte resurse sau asistență utilă.

Criteriile și recomandările menționate sunt destinate pentru interpretarea patologieilor moștenite (primar Mendeliene) și nu poate fi utilizată pentru variațiile somatice, variantelor farmacogenomice (PGx), sau a varinatelor în genele asociate cu patologieile complexe multi-genice non-Mendeliene.

## **2.4. OBIECTUL ȘI METODELE DE CERCETARE**

Cercetările în scopul stabilirii profilului mutațional și căilor de semnalizare afectate în cancerul de sân au fost efectuate în colaborare cu *Laboratorul Imunogenetic și Morfologia Tumorilor*, IMSP Institutu Oncologic din Republica Moldova.

Pentru generarea datelor NGS, ADN-ul a fost izolat din 19 probe de țesut tumoral proaspăt obținut chirurgical. Selectarea cazurilor s-a bazat pe diagnosticul confirmat clinic.

Izolarea ADN-ului s-a efectuat cu kit-ul GeneJET Genomic DNA Purification Kit (Thermo Scientific) utilizând protocolul standard. Cantitatea și puritatea probelor a fost estimată prin măsurări spectrofotometrice, utilizând spectrofotometrul NanoDropLite. Raportul  $\lambda 260/\lambda 280$  a fost în limita valorilor 1,6 – 2,0, iar cantitatea de ADN 320000 – 780000 ng.

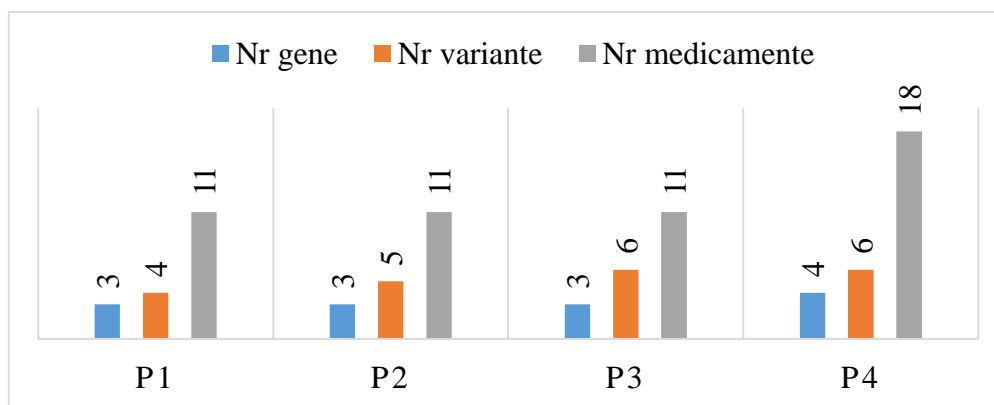
Integritatea ADN-ului genomic a fost estimată electroforetic utilizând sistemul de electroforeză în timp real E-Gel iBase (Invitrogen). ADN-ul genomic au fost nefragmentat prezentându-se o bandă solidă în apropierea zonei de start a gelului de agaroză. Concentrația a fost determinată kit-ul Qubit dsDNA BR Assay Kit (Invitrogen) utilizând fluorometrul Qubit 3.0 (Invitrogen). Concentrația ADN-ului a variat între 230 – 570 ng/uL. Panelul Ion AmpliSeq Cancer Hotspot Panel v2 utilizat, constă dintr-un ansamblu de 207 perechi de primer într-un singur tub, fiecare pereche proiectată pentru a amplifica o regiune genomică specifică care împreună acoperă 50 oncogene și gene supresoare de tumoare: ABL1, EGFR, GNAS, KRAS, PTPN11, AKT1, ERBB2, GNAQ, MET, RB1, ALK, ERBB4, HNF1A, MLH1, RET, APC, EZH2, HRAS, MPL, SMAD4, ATM, FBXW7, IDH1, NOTCH1, SMARCB1, BRAF, FGFR1, JAK2, NPM1, SMO, CDH1, FGFR2, JAK3, NRAS, SRC, CDKN2A, FGFR3, IDH2, PDGFRA, STK11, CSF1R, FLT3, KDR, PIK3CA, TP53, CTNNB1, GNA11, KIT, PTEN, VHL.

Parametrii de bază de la producător: lungimea ampliconilor variază între 111 și 187 bp, cu o medie de 154 bp. Procentul bazelor acoperite la  $\geq 20\%$  din acoperirea medie (uniformitatea acoperirii) este  $>95\%$  iar procentul de read-uri care se aliniază în regiuni țintă, din totalul de read-uri aliniat per lansare (read-uri per țintă) este  $>90\%$ . Adâncimea medie de acoperire este  $>2000\times$  pe chip-ul 314.

Bibliotecile genomice au fost pregătite conform protocolului standard AmpliSeq utilizând kit-ul Ion AmpliSeq Library Kit 2.0 (Thermo Fisher Scientific) și panelul de primeri Ion AmpliSeq Cancer Hotspot Panel v2 (Thermo Fisher Scientific). Fiecare bibliotecă a fost barcodată cu Ion Xpress Barcode Adapters 1-16 Kit (Thermo Fisher Scientific) apoi cuantificate utilizând Qubit dsDNA HS (High Sensitivity) Assay Kit și fluorometrul Qubit 3.0 (ambele Thermo Fisher Scientific) și diluată în apă fără nuclează pentru a obține o concentrație finală de 100 M. Bibliotecile pregătite au fost amplificate clonal pe particule Ion Sphere (ISP) folosind PCR de emulsie cu sistemul Ion OneTouch 2 System (Thermo Fisher Scientific) conform instrucțiunilor producătorului. Concentrația bibliotecilor a fost crescută de la 2  $\mu\text{L}$  sugerată de producător la 6  $\mu\text{L}$  pentru a obține un număr mai mare de particule ISP încărcate, fără a avea o rată inacceptabilă de read-uri policlonale. IPS au fost îmbogățite prin metoda bazată pe bile magnetice utilizând sistemul Ion OneTouch ES Enrichment System (Thermo Fisher Scientific). ISP-urile îmbogățite au fost încărcate pe chip-uri 314 și 318.

## 2.5. EVALUAREA STATUTULUI MUTAȚIONAL ȘI A CĂILOR DE SEMNALIZARE AFECTATE

*Analiza țesuturilor afectate de cancer pulmonar.* În urma procesării și secvențierii celor 4 probe de adenocarcinom pulmonar au fost depistate 10 mutații. Datele NGS analizate pentru variantele genetice și potențiale interacțiuni medicamentoase prezintă mai multe sugestii specifice de tratament incluse în acest raport (Figura 1), care, împreună cu datele clinice ale pacientului, cum ar fi istoricul tratamentului, comorbiditățile și radiografiile, ar putea forma baza pentru luarea deciziilor terapeutice într-o manieră interdisciplinară.



**Fig. 1. Mobilitatea strategiilor terapeutice per pacient**

Numărul posibilelor opțiuni terapeutice depinde nu doar de numărul mutațiilor per tumoare ci și de tipul genei mutante. Conform Figurei 1, pacientul nr 4 care are mutații în gena MET și ERBB2, nr de opțiuni terapeutice este mult mai mare decât la pacienții cu genele respective normale. O serie de preparate pot fi utilizate pentru toți pacienții, conform poștelor moleculare individuale, iar altele pot fi utilizate specific doar pentru un pacient (Figura. 2).

Opțiunile terapeutice depind de căile de semnalizare afectate și posibilitatea de a acționa asupra mai multor poziții într-o cale de semnalizare specifică. Putem conecta modificările genetice în celulele canceroase cu căi de semnalizare care controlează procesele asociate cu tumorigeneza și le plasează în contextul distorsiunilor rețelelor de semnalizare mai largi care alimentează progresia cancerului.

În fiecare caz, rezultatul este semnalizarea neregulată care nu este supusă mecanismelor normale de control. Variantele genetice identificate în urma analizei datelor contribuie în mare măsură la producerea de proteine mutante a căror activitate este defectă (de exemplu mutația în frame a genei ERBB2). Cu toate acestea, una dintre cele mai frecvent mutate gene în cancer este supresorul tumoral p53. p53 este un hub critic care controlează proliferarea celulelor și semnalele de stres, cum ar fi apoptoza și răspunsurile la deteriorarea ADN. Având în vedere că multe dintre aceste procese celulare importante sunt cooptate în mod obișnuit în cancerul pulmonar, nu este surprinzător faptul că căile efectoare care stau la bază sunt adesea defecate în celulele tumorale. Unele gene cu mutații au activitate kinazică (EGFR, ERBB2, MET) și, ca rezultat, sunt cele mai eficiente pentru terapiile țintite.

*Analiza țesuturilor afectate de cancer de sân* a permis identificarea unui număr total de 74 mutații. Cele mai multe mutații au fost identificate în genele PIK3CA, TP53, KDR și PTEN, între 8 și 5 mutații per genă, iar în 16 gene a apărut câte o singură mutație. Cele mai multe mutații s-au dovedit a fi de tip missens, iar 18 - fiind necunoscute. Mutațiile la nivelul oncogenilor s-au întâlnit cu o frecvență mai mare, comparativ cu genele supresoare de tumoare, coraportul fiind de 20 la 13. Frecvența generală cu

care au mutat genele per numărul total de probe variază între 30 și 1, cea mai mare frecvență fiind în genele RET, PDGFRA, TP53, FLT3, KDR, FGFR3 și PIK3CA, iar în 10 gene, cum ar fi AKT1, ALK, CDH1 și altele s-a întâlnit doar câte o singură mutație. Fiecare dintre cele 20 de gene de cancer în care au fost identificate substituții sau indeluri importante sunt reprezentate grafic. Cinci dintre cele 20 gene (EGFR (89,47), KIT (36.84), PIK3CA (36.84), PTEN (26,31) și NOTCH1 (26,31)) au fost mutate preponderent. Toate mutațiile frameshift au apărut în genele supresoare de tumoare

## ***2.6. ANALIZA BIOINFORMATICĂ A REȚELOR DE SEMNALIZARE***

În mod tradițional, cercetătorii și-au concentrat eforturile pe fenomene biologice unice (de exemplu, o mutație a unei singure gene) sau pe o cale de semnalizare specifică. Acum, epoca „omics” a datelor mari a adus metode de prelucrare de ultimă oră pentru interpretarea mega datelor biologice, care acum au fost adoptate în mod universal. Pe baza unor astfel de mega date (așa-numitele „date mari”), cercetătorii urmăresc să înțeleagă modificările fenotipului bazate pe niveluri de sistem, evaluând căi / rețele întregi și nu doar o singură entitate.

Eforturile bioinformaticii au la bază elaborarea strategiilor de tratament care vizează pozițiile specifice ale moleculelor de semnalizare sau ale efectorilor lor din aval. Acest tip de terapie are un mare potențial, deoarece se bazează, mai degrabă, pe blocarea moleculelor specifice, decât pe calea chimioterapiei sau radioterapiei tradiționale.

Analiza acoperirii mediane pentru fiecare nucleotidă în 92 ampliconi din cele 4 probe a arătat o adâncime (coverage) de acoperire mediană de >1000 pentru toate probele, ceea ce indică acoperirea secvențială în ansamblu foarte eficientă.

Au fost identificate variante genetice în genele implicate în 11 căi de semnalizare, combinația de gene aberante variind de la brobă la probă. După filtrarea mutațiilor în dependență de efectul acestora asupra funcției proteinei și ca urmare, efectul asupra căilor de semnalizare de bază (baza de date COSMIC), numărul maxim de mutații într-u tumoare individual a fost de 6, variind între 2 și 5, dar într-un caz s-a identificat o singură mutație în genă driver. Mutațiile cu efect asupra funcției proteinei sau expresiei genei repartizându-se egal între tipul de genă (10 oncogene și 10 TSG).

Sunt potențial afectate 8 căi de semnalizare din cele 12 căi de semnalizare de bază implicate în dezvoltarea cancerului, 5 căi de semnalizare implicate în procesul de supravețuire celulară, 2 implicate procesul de maturizare celulară și 1 implicată în mentenanța genomului.

Toate tumorile prezintă hiperactivarea căii de semnalizare PI3K care și este conform Datele Cancer Genome Atlas una din cele mai relevante căi țintă în cancerul de sân. Calea PI3K este dereglată printr-o varietate de mecanisme, incluzând pierderea sau inactivarea PTEN-ului (genă supresoare de tumoare), mutația sau amplificarea PIK3CA, precum și activarea receptorilor tirozin kinazici sau oncogenelor în amonte de PI3K.

Cinci probe tumorale prezintă și mutații ale genei PTEN în calea de semnalizare PI3K. PTEN-u având efect negativ în ceea ce privește eficacitatea inhibitorii PI3K $\alpha$ . Pacientul P1 are mutații în gena PIK3CA dar nu în PTEN, deci poate fi utilizat tratamentul cu inhibitori PI3K combinat cu inhibitori EGFR. Spre exemplu, pacientul P11 este un candidat bun pentru tratamentul cu inhibitori AKT, dar pentru un efect mai bun este necesar de combinat cu alte preparate cum ar fi trastuzumabul, care la fel are efect inhibitor a căii PI3K hiperactive sau olaparibul din cauza prezenței mutației patogene în gena ATM. P3 poate beneficia de palbociclib (inhibitor CDKN2A) și GSK2636771 (în prezența dereglării căi PI3K din cauza mutației PTEN).



### 3. CELE MAI RELEVANTE REALIZĂRI OBTINUTE ÎN CADRUL PROIECTULUI

A fost instituită o colaborare multidisciplinară între colective de cercetători din diverse domenii (*Centrul Genetică Funcțională - Universitatea de Stat „Dimitrie Cantemir”, Laboratorul Imunogenetic și Morfologia Tumorilor și Laboratorul Științific Mamologie - IMSP Institutul Oncologic, Centrul Sănătatea Reproducerii și Genetica Medicală - IMSP Institutul Mamei și Copilului*), care permite efectuarea studiilor fundamentale (genetică, biologie moleculară, bioinformatică, etc) și finalizează cu aplicații practice (biologie medicală, medicină, strategii personalizate de tratament, etc).

S-a realizat secvențierea și analiza datelor la 32 pacienți cu cancer pulmonar și cancer de sân, fiind identificate o serie de mutații cunoscute în literatura de specialitate și mutații noi, care urmează a fi descrise.

S-a realizat analiza acestor mutații valorificând un sistem de algoritmi bioinformatici aplicați pentru asemenea studii (R, Cytoscape) și multiple baze de date (COSMIC, TCGA).

Cunoașterea tipului de mutație la pacienții respectivi și utilizarea sistemelor de analiză bioinformatică cu clasificarea ulterioară a acestora conform criteriilor aprobate la nivel internațional va permite sporirea efectelor de tratare a cancerului și utilizarea tratamentelor pentru un anumit tip de mutație și respectiv o cale de semnalizare specifică.

A fost elaborat un *Ghid metodic* care oferă studenților precum și tuturor cercetătorilor începători în domeniu, o descriere simplă și clară a tehnologiilor de secvențiere și a metodelor de analiză a datelor obținute. Lucrarea de față reflectă o revizuire tehnică a tipurilor de secvențiere, modalitatea de pregătire a probelor și analiza bioinformatică a secvențelor de acizi nucleici generate în cadrul NGS, precum și gama largă de aplicații pentru tehnologiile disponibile în prezent.

## 2. PARTICIPAREA ÎN PROGRAME ȘI PROIECTE INTERNAȚIONALE (ORIZONT 2020, COST...), INCLUSIV PROPUNERILE ÎNAINȚATE/PROIECTE CÂȘTIGATE ÎN CADRUL CONCURSURILOR NAȚIONALE/INTERNAȚIONALE CU TANGENȚA LA TEMATICA PROIECTULUI

Membrul echipei proiectului acad. Maria Duca a depus două proiecte individuale la programe și proiecte internaționale (ORIZONT 2020, COST).

În conformitate cu rezultatele evaluării acad. M. Duca a fost acceptată ca reprezentant al Republicii Moldova în WG: PlantEd – COST Action CA18111 – Genome editing in plants “Plant Genome Editing – State of the Art”.

## 4. COLABORĂRI ȘTIINȚIFICE INTERNAȚIONALE/NAȚIONALE

### ➤ IMSP Institutului Oncologic

*Laboratorul Imunologie și Genetică moleculară*

*Șef Laborator: dr. Valentia STARTAN*

### ➤ IMSP Institutul Mamei și Copilului

*Centrul Sănătatea Reproduserii și Genetica Medicală*

*Director Mihai STRĂȚILĂ*

## 5. VIZITE ALE CERCETĂTORILOR ȘTIINȚIFICI DIN STRĂINĂTATE

În perioada de referință, Centrul Genetică Funcțională a fost vizitat de un grup de cercetători de la Universitatea ”Alexandru Ioan Cuza” din Iași, Romania: **BOIANGIU Razvan Stefan**, **CAPATINA Luminita** și **BRINZA Ion**, care în perioada 21-22 octombrie au participat la lucrările conferinței *Life sciences in the dialogue of generations: „Connections between universities, academia and business community”*, prezentând rapoarte în cadrul Secției *Realizări și perspective în domeniul Biomedicinii*.

Discuțiile avute cu cercetătorii științifici care activează în cadrul Universității ”Alexandru Ioan Cuza” din Iași au prezentat o platformă privind posibilitățile de colaborare în domeniul dat, inclusiv mobilități ale cercetătorilor din cadrul USDC, cercetări și publicații în comun, precum și perfectarea unor proiecte în cadrul apelurilor anunțate în programele cadru la nivel internațional.

## 7. TEZE DE DOCTORAT/POSTDOCTORAT SUSȚINUTE PE PARCURSUL REALIZĂRII PROIECTULUI

În data de 18 octombrie 2019 în cadrul Consiliului Științific Specializat dh 162.02-03 din cadrul Universității de Stat „Dimitrie Cantemir”, Centrul Genetică Funcțională a avut loc susținerea publică a tezei de doctor habilitat în biologie de către Dna **SACARA Victoria** (<http://www.cnaa.md/thesis/55307/>)

**Tema tezei:** Particularitățile molecular-genetice ale patologiilor neuromusculare frecvent întâlnite în Republica Moldova

**Specialitatea:** 162.02 Genetica omului și animalelor

**Conducători științifici:**

**DUCA Maria**, doctor habilitat în științe biologice, profesor universitar, academician

**GROPPA St.**, doctor habilitat în științe medicale, academician

## 8. MANIFESTĂRI ȘTIINȚIFICE ORGANIZATE LA NIVEL NAȚIONAL/INTERNAȚIONAL

**21-22 octombrie 2019 - Conference Life sciences in the dialogue of generations: „Connections between universities, academia and business community” - Chișinău, Republica Moldova**

Organizarea acestui eveniment a avut ca scop asigurarea unui dialog eficient și transferul de cunoștințe și experiență dintre savanții notorii din țară și străinătate cu cercetătorii juniori (studenți masteranzi, doctoranzi) din diverse instituții. Un alt aspect al evenimentului se axează pe realizarea conexiunii dintre mediul de afaceri cu cercetarea științifică.

Una dintre secțiunile de lucru a Conferinței a fost **Realizări și perspective în domeniul Biomedicinii** care a oferit o platformă pentru a elucida și dezbate mai multe tematici privind implementarea unor noi metode moleculare de analiză în medicină.

La conferință au fost înregistrate 63 de instituții de cercetare cu peste 250 de reprezentanți ai comunității academice naționale și internaționale, profesori universitari, cercetători științifici din 18 universități, 36 instituții de cercetare și 9 companii private.

Au fost prezenți cercetători din 9 țări – China, Germania, Federația Rusă, Italia, Marea Britanie, România, Turcia, Ucraina, Republica Moldova.

## 9. APRECIEREA ACTIVITĂȚII ȘTIINȚIFICE PROMOVATE LA EXECUTAREA PROIECTULUI

**Maria DUCA** Premiul AȘM pentru rezultate științifice de valoare obținute în anul 2018, domeniul Biologie și ecologie „Alexandru Ciubotaru” – pentru ciclul de lucrări „Studii genetico-moleculare aplicate în biotehnologiile moderne de cultivare a florii-soarelui” (Hotărârea nr. 158 din 6 noiembrie 2019 a Prezidiului AȘM).

## 10. REZUMATUL RAPORTULUI CU EVIDENȚIEREA REZULTATULUI, IMPACTULUI, IMPLEMENTĂRILOR, RECOMANDĂRILOR

Proiectul **18.80.07.16A/PS** ” *Crearea suportului decizional în rapoartele de secvențiere de următoarea genarție pentru variantele somatice a cancerului*”, conducător: acad. Duca Maria.

**Programul de Stat *Medicina de precizie în prevenirea, diagnosticul și tratamentul patologiilor pentru anii 2018-2019.***

Direcția Strategică ”Sănătate și biomedicină” 80.07

Progresele majore din domeniul biologiei moleculare și geneticii au contribuit la dezvoltarea unor abordări cu totul noi în științele vieții, inclusiv în medicină. Descoperirea structurii dublu helix (Watson JD și colab. în 1953) a acidului dezoxiribonucleic (ADN), compusă din patru baze azotate de {A, T, C, G} a inaugurat era modernă a geneticii. Gena a devenit un segment de ADN care deține informația codificată pentru biosinteza unei proteine (mai exact, pentru așezarea secvențială specifică a aminoacizilor în proteină).

Aceste descoperiri au făcut că la ora actuală să fie dezvoltate o serie de metode care permit identificarea genelor de interes. Una dintre aceste tehnologii, cu impact major asupra cercetărilor biomedicale, este în prezent - secvențierea.

Deși prezintă numeroase dificultăți, analizele genomice au o valoare teoretică și practică deosebită. În planul cercetărilor fundamentale acestea permit studiul structurii genelor, stabilirea hărții complete a genomului, elucidarea mecanismelor de expresie genică și în special a modalităților de reglaj, cheia unor procese biologice fundamentale - diferențierea celulară, morfogeneza, senescența, funcția sistemului nervos sau a sistemului imun, etc.

În medicina practică, analiza genomică este folosită pentru studiul patologiei moleculare produsă de variantele ADN, indiferent dacă este vorba de o patologie ereditară (boli genetice) sau somatică (cancere), diagnosticul genotipic prenatal sau presimptomatic al unor boli.

Secvențierea este utilizată pentru determinarea compoziției ADN-ului în scopul determinării structurii genei, tipul de mutație, precum și a secvenței de aminoacizi pentru proteina codificată de gena respectivă în contextul analizei diversității, complexității, organizării și tratamentului bolilor. Stabilirea secvenței nucleotidice este utilă și pentru sinteza oligosondelor specifice pentru o alelă și a amorselor folosite în PCR, ambele fiind larg utilizate în diagnosticul molecular.

Astfel, având ca punct de plecare secvența totală sau parțială a genomului unui organism, poate fi determinată secvența de aminoacizi, identificate elementele de secvență înalt conservate pe linie evolutivă, de asemenea formulate o serie de ipoteze privind funcțiile acestora. Această abordare, constituie punctul de plecare pentru o serie de cercetări ce au în vedere nu numai elucidarea unor probleme strict teoretice, dar și evidențierea unor enzime cu potențiale aplicații practice

Scopul acestui proiect a constat în inițierea unor **cercetări inovatoare atât pentru Republica Moldova cât și pentru domeniul medicinei**, în special – în diagnosticul prematur și eficientizarea tratamentului de cancer. Obiectivele proiectului au constat în implementarea unor tehnici noi de biologie moleculară în analiza genetică a cancerului și aplicarea algoritmilor de bioinformatică în analiza datelor obținute.

Pe parcursul a doi ani de realizare a proiectului s-au pus bazele aplicării unor cercetări sistemice care încep cu *Pacient și analiza cazurilor concrete - Validare clinică - Identificare de gene mutaționale - Identificare de cale metabolică afectată – recomandare de tratament personalizat prin joncțiunea*

eforturilor materiale, competențelor și potențialului uman al specialiștilor din diverse domenii (*Centrul Genetică Funcțională - Universitatea de Stat „Dimitrie Cantemir”, Laboratorul Imunogenetic și Morfologia Tumorilor și Laboratorul Științific Mamologie - IMSP Institutul Oncologic, Centrul Sănătatea Reproducerii și Genetica Medicală - IMSP Institutul Mamei și Copilului*). S-a instituit o colaborare multidisciplinară, care începe de la studii fundamentale (genetică, biologie moleculară, bioinformatică, etc) și finalizează cu aplicații practice (biologie medicală, medicină, strategii personalizate de tratament, etc).

S-a realizat secvențierea și analiza datelor la 32 pacienți cu cancer pulmonar și cancer de sân, fiind identificate o serie de mutații cunoscute în literatura de specialitate și mutații noi, care urmează a fi descrise.

S-a realizat analiza acestor mutații valorificând un sistem de algoritmi bioinformatici aplicați pentru asemenea studii (R, Cytoscape) și multiple baze de date (COSMIC, TCGA).

Cunoașterea tipului de mutație la pacienții respectivi și utilizarea sistemelor de analiză bioinformatică cu clasificarea ulterioară a acestora conform criteriilor aprobate la nivel internațional va permite sporirea efectelor de tratare a cancerului și utilizarea tratamentelor pentru un anumit tip de mutație și respectiv o cale de semnalizare specifică.

A fost elaborat un *Ghid metodic* care oferă studenților precum și tuturor cercetătorilor începători în domeniu, o descriere simplă și clară a tehnologiilor de secvențiere și a metodelor de analiză a datelor obținute. Lucrarea de față reflectă o revizuire tehnică a tipurilor de secvențiere, modalitatea de pregătire a probelor și analiza bioinformatică a secvențelor de acizi nucleici generate în cadrul NGS, precum și gama largă de aplicații pentru tehnologiile disponibile în prezent.

## 11. CONCLUZII

1. Aplicarea în practică și utilizarea panelurilor multigenice în Republica Moldova în baza Secvențierii de Următoare Generație (*NGS – Next Generation Sequencing*), elaborată în 2005 care a schimbat perspectivele de analiză și înțelegere a ființelor vii, va contribui la un progres considerabil de evoluție și dezvoltare a diagnosticului molecular și analiză a datelor, va contribui la crearea profilelor genomice individuale și astfel va ajuta la selecția unui plan de tratament mai precis și personalizat.

2. Identificarea căilor de semnalizare de bază afectate și nivelul mutațiilor identificate în cadrul studiilor demonstrează particularitățile specifice la care are loc alterarea pentru fiecare pacient și oferă informații utile despre necesitatea și modul aplicării terapiilor combinate.

3. Deși rezistența dobândită la medicament este o preocupare perpetuă în ceea ce privește terapiile anticancer, găsirea combinației potrivite de medicamente în baza mutațiilor în gene și oncogene pentru pacienți individuali ar putea duce la maximizarea activității anticanceroase, reducând în același timp toxicitatea. În acest sens, țintirea tumorilor cu combinații de medicamente care inhibă membrii aceleiași căi poate fi mai puțin toxică, păstrând în același timp activitatea antitumorală.

Conducătorul proiectului

acad. Maria DUCA

## 12. BUGETUL PROIECTULUI

Volumul total al finanțării (mii lei) (pe ani)

Anul	Planificat	Executat	Cofinanțare
2018	153,5	153,5	50,0
2019	153,5	153,5	50,0

## 13. LISTA EXECUTORILOR, LISTA TINERILOR CERCETĂTORI, DOCTORANZILOR

**Lista executorilor** (funcția în cadrul proiectului, titlul științific, semnătura)

Nr d/o	Numele/Prenumele	Anul nașterii	Titlul științific	Funcția în cadrul proiectului	Semnătura
1.	DUCA Maria	1956	Doctor habilitat	Director de proiect	
2.	SAVCA Elena	1954	Doctor	Cercetător științific	
3.	MUNTEANU Viorel	1983		Cercetător științific	
4.	MARTEA Rodica	1987	Doctor	Cercetător științific	
5.	ȚUȚUIANU Valeri	1989		Cercetător științific	

**Lista tinerilor cercetători**

Nr d/o	Numele/Prenumele	Anul nașterii	Titlul științific	Funcția în cadrul proiectului
1.	MARTEA Rodica	1987	Doctor	Cercetător științific
2.	ȚUȚUIANU Valeri	1989	-	Cercetător științific

**Lista doctoranzilor**

Nr d/o	Numele/Prenumele	Anul nașterii	Titlul științific	Funcția în cadrul proiectului
1.	ȚUȚUIANU Valeri	1989	-	Cercetător științific

Conducătorul proiectului

acad. Maria DUCA

\_\_\_\_\_

### 13. LISTA PUBLICAȚILOR ȘTIINȚIFICE CE ȚIN DE REZULTATELE OBTINUTE ÎN CADRUL PROIECTULUI

#### în anul 2018

#### – Rapoarte publicate/Teze ale comunicărilor la congrese, conferințe, simpozioane, în culegeri (naționale / internaționale)

1. STRATAN, V.; MUNTEANU, V.; LOZOVANU, A.; TUTUIANU, V.; SITNIC, V.; TCACIUC, D. Using the PGM Ion Platform for molecular-genetic characterisation of breast cancer. *Romanian Journal of Laboratory Medicine*, 2018, supliment la vol. 26, nr. 2, S90. ISSN 1841-6624.

#### în anul 2019

#### Ghid metodic

1. MUNTEANU, V.; MARTEA, R.; DUCA, M. *Metode de secvențiere: Ghid metodic*, Univ. de Stat „Dimitrie Cantemir”. – [Chișinău] : Univ. de Stat „Dimitrie Cantemir”, 2019 (Tipogr. " Biotechdesign"). – 29 p.

#### Articole reviste categoria B

1. STRATAN, V., SÎTNIC, V., ȚUȚUIANU, V., BÎLBA, V., BRENIȘTER, S., ȘUTKIN, V., LOZOVANU, A., MUNTEANU, V., POPA, C. Detectarea și analiza mutațiilor somatice în unele probe de adenocarcinom pulmonar prin tehnologia de secvențiere ion torrent. *Buletinul Academiei de Științe, Științe Medicale*, 2019, 3 (acceptat spre publicare).

#### Rapoarte publicate/Teze ale comunicărilor la congrese, conferințe, simpozioane, în culegeri Internaționale

1. DUCA, M.; MARTEA, R.; MUNTEANU, V.; TUTUIANU, V. Challenges in Next-Generation Sequencing (NGS) Analysis for Cancer Research. *International Biological, Agricultural And Life Science Congress (BIALIC)*, November 7-8, 2019, Lviv, Ukraine, p. 155.
2. STRATAN, V., SÎTNIC, V., ȚUȚUIANU, V., BÎLBA, V., LOZOVANU, A., MUNTEANU, V., BRENIȘTER, S., ȘUTKIN, V. Rețeaua de interacțiune a unor gene mutante în cancerul pulmonar non-microcelular, Confer2019, 21-23 noiembrie 2019, Iași, România, volumul de rezumate, nr. 8, p. 415-416.
3. STRATAN, V., ȚUȚUIANU, V., SÎTNIC, V., LOZOVANU, A., MUNTEANU, V., SOCHIRCĂ, D. Heterogenitatea genomică intertumorală și intratumorală în cancerul de sân, Confer2019, 21-23 noiembrie 2019, Iași, România, volumul de rezumate, nr. 8, p. 412-413.
4. STRATAN, V., ȚUȚUIANU, V., SÎTNIC, V., LOZOVANU, A., MUNTEANU, V., SOCHIRCĂ, D. Utilizarea limbajului R pentru adnotarea clinică a unor gene mutante în cancerul de sân, Confer2019, 21-23 noiembrie 2019, Iași, România, volumul de rezumate, nr. 8, p. 258-259.

#### Naționale

1. MARTEA, R.; DUCA M. NGS: tendințe actuale de diagnostic molecular de laborator în cancerul somatic. *Conference Life sciences in the dialogue of generations: „Connections between universities, academia and business community”*, 21-22 octombrie 2019, Chișinău, p. 96.

Conducătorul proiectului

acad. Maria DUCA

\_\_\_\_\_

## 14. PARTICIPĂRI LA MANIFESTĂRI ȘTIINȚIFICE NAȚIONALE/INTERNAȚIONALE

### Manifestări științifice Internaționale

1. MARTEA, R. // *5th @RoBioinfo SEMINAR: BIOINFORMATICS TOOLS FOR EXPLORING PROTEIN BIOLOGY*, 4-5 aprilie 2019, Iași, România.
2. MARTEA, R. // *INTERNATIONAL BIOLOGICAL, AGRICULTURAL AND LIFE SCIENCE CONGRESS*, 7-8 noiembrie 2019, Lvov, Ucraina

POSTER : Challenges in Next-Generation Sequencing (NGS) Analysis for Cancer Research

Conducătorul proiectului

acad. Maria DUCA

\_\_\_\_\_