

Universitatea Babeș-Bolyai Cluj-Napoca  
Școala Doctorală de Biologie Integrativă

**Analiza bioinformatică a secvențelor de acizi nucleici generate prin  
studii genomice structurale și funcționale în evaluarea dezvoltării  
tumorilor maligne**

*Rezumatul tezei*

Coordonatori științifici:

**Prof. Dr. Nicolae Dragoș**

**Prof. Dr. Dan Dumitrașcu**

Doctorand:

**Roxana M. Cojocneanu (Petric)**

Cluj-Napoca, 2015

## Cuprins

1. INTRODUCERE.....	4
2. STADIUL ACTUAL AL CUNOAȘTERII.....	5
Statisticile din domeniul cancerului.....	5
Terapii convenționale în cancer .....	6
Terapii țintite.....	6
Cunoștințe curente despre cancer.....	6
Tehnologii de tip ”high throughput” .....	7
Analizele bioinformatică a datelor de tip ”high throughput” .....	7
Importanța ”Big Data” .....	8
3. CONTRIBUȚIE PERSONALĂ: Managementul datelor pentru numeroase aplicații ”omics” .....	9
3.1. STUDIUL NGS: Analizele de bioinformatică pentru secvențierea de nouă generație în cancerul de sân triplu negativ .....	9
Cancerul de sân și TNBC – caracterizare, nevoie de management mai bun .....	9
Materiale, metode și protocoale .....	10
Aspecte Etice .....	10
Pacienții.....	10
Tehnologia Ion Torrent .....	10
Extracția ADN.....	11
Cuantificarea ADN .....	11
Prepararea bibliotecii de ampliconi .....	11
Amplificarea ADN genomic cu amorse specifice .....	12
Purificarea ampliconilor .....	12
Digestia parțială a amorsoarelor.....	12
Marcarea ampliconilor .....	12
Cuantificarea bibliotecii de ampliconi.....	12
Amplificarea bibliotecii ISP (Ion Sphere Particles).....	13
Îmbogățirea ISP .....	13
Încărcarea chip-urilor ”316” .....	13
Rezultate și discuții .....	15
Validarea mutațiilor identificate.....	17
Limitări și considerații ulterioare .....	18
Concluzii .....	18
3.2. STUDIU MICROARRAY: Analiza bioinformatică a datelor microarray și extrapolarea porc-om în relație cu expunerea la Zearalenone și <i>Escherichia coli</i> .....	20
Modele animale – când și de ce le folosim? .....	20
Introducerea și motivația acestui studiu .....	20

Tehnologia Microarray – Privire de ansamblu .....	21
Materiale, metode și protocoale .....	21
Extracția de ARN total cu TRI Reagent (Sigma-Aldrich) .....	21
Purificarea ARN folosind RNeasy Mini Kit (Qiagen).....	21
Cuantificarea ARN .....	22
Sinteza sondelor.....	22
Purificarea sondelor .....	22
Cuantificarea sondelor și diluția.....	22
Hibridizarea sondelor .....	22
Procesarea datelor.....	22
Analiza datelor .....	23
Raționamentul pentru metoda de extrapolare .....	24
Sucesiunea de metode pentru extrapolare .....	24
Rezultate și discuții .....	24
Concluzii parțiale pentru metoda de extrapolare .....	24
Efectul extrapolării co-contaminării la oameni .....	25
Modelul de evaluare a expresiei genice în experimentul pe duoden .....	25
Analiza rețelelor de semnalizare.....	25
Concluzii .....	26
4. CONCLUZII GENERALE .....	27
Lista publicațiilor.....	29
Bibliografie selectivă .....	31

**Cuvinte cheie:** bioinformatică, analiză date, terapie personalizată, secvențiere de nouă generație, microarray, modele animale, extrapolare

## 1. INTRODUCERE

Atunci când știința întâlnește tehnologia de înaltă performanță, pot avea loc progrese majore, ca dovadă fiind cantitatea mare de date generate de progresele tehnologice recente din domeniul biologiei moleculare, mai precis în genomica funcțională și structurală. Printre aceste tehnologii cu impact major asupra cercetării medicale se numără secvențierea de nouă generație și tehnica microarray, ce au potențialul de a îmbunătăți metodele clinice de diagnostic și terapie țintită (1-4). Aceste progrese au facilitat studierea profilului molecular pentru numeroase patologii, cu potențialul de a conduce la descoperirea unor terapii țintite eficiente, dar și a biomarkerilor pentru diagnostic timpuriu, stratificare și prognostic.

Cantitatea mare de informații generate de aceste instrumente poate varia de la numeroase giga-perechi de baze (gpb) de date brute de secvențiere, până la tabele extinse structurate în fișiere, conținând sute, uneori chiar mii de rânduri, așa cum se întâmplă în cazul datelor expresiei genice sau a miARN obținute în urma tehnologiei microarray (5-7). Ca și în cazul oricărui test, valoarea rezultatelor constă în abilitatea cercetătorului de a interpreta și integra informațiile obținute într-o manieră cu semnificație biologică. Așadar, în încercarea de a transforma datele tip *high throughput* în date accesibile este necesară organizarea și descifrarea lor cu ajutorul metodelor bioinformaticice. Pe scurt, domeniul bioinformaticii reprezintă un sistem de management a informațiilor pentru biologia moleculară și are numeroase aplicații practice (8).

În ultimii ani, atenția medicilor și a cercetătorilor clinici a început să se mute de la tratarea pacienților deja afectați de diverse boli spre predicție, prevenire și medicina personalizată (PPMP) sau medicină de precizie, acest trend fiind foarte bine ilustrat în momentul de față în domeniul oncologiei (9). În încercarea de a asocia fiecare pacient cu cel mai bun plan terapeutic, medicii au început să se bazeze pe mai multe informații decât cele oferite de stratificarea histopatologică a bolii. Progresele recente din cercetarea din domeniul biomedicine și a geneticii oferă clinicienilor informații mai precise, precum expresia genică și moleculară în tumori, starea mutațională a genelor implicate în anumite boli, gradul de polimorfism și variație a numărului de copii. O bună congruență între profilul molecular a pacientului și răspunsul terapeutic nu reprezintă doar o îmbunătățire a rezultatului final, dar și un pas înainte în ceea ce privește medicina personalizată și precisă.

*Motivația din spatele acestei teze a fost de a sublinia modul în care analizele "big data" pot contribui la personalizarea diagnosticului și a terapiei, prin folosirea tehnologiilor pentru profilarea bio-moleculară tip "high throughput", cu scopul de a analiza semnătura moleculară a condițiilor fizice și patologice. Alături de costuri minimalizate și dezvoltare de paneluri biologice specializate – dar și cu posibilitate de a adapta aceste paneluri pentru diverse teste microarray sau de secvențiere de nouă generație – aceste tehnologii pot îndeplini o varietate largă de cerințe clinice pentru numeroase patologii. Prin folosirea bioinformaticii și a metodelor de analiză, interpretarea și integrarea a cantităților mari de date generate de aceste tehnologii, va face posibilă dezvoltarea de metode rapide și eficiente aplicabile în special în domeniul oncologiei. În primul rând, un diagnostic rapid și eficient va asigura tratamentul cel mai potrivit pentru situația patologică a fiecărui pacient. Biomarkerii preciși, nou descoperiți, vor contribui la dezvoltarea de metode de screening mai eficiente și minim invazive, ce vor stratifica potențialii pacienți și oferi posibilitatea unui tratament precis și rapid. Cu toate că nu promite să devină Sfântul Graal al medicinei, bioinformatica va schimba cel mai probabil fața diagnosticului și tratamentului oncologic, conducând fără îndoială la o viață și rată de supraviețuire îmbunătățită pentru pacienți.*

## **2. STADIUL ACTUAL AL CUNOAȘTERII**

### **Statisticile din domeniul cancerului**

Cancerul se încadrează printre principalele cauze ale decesului la nivel global, cu o medie a mortalității de 8 milioane cazuri anual, conform statisticilor furnizate de World Health Organization (10). Conform site-ului GLOBOCAN 2012, s-au estimat aproximativ 14.1 milioane de cazuri noi de cancer la ultima evaluare la nivel mondial, dintre care 7.4 milioane de cazuri la bărbați și 6.7 milioane la femei. Se preconizează că acest număr va crește la 24 de milioane până în 2035 (11, 12). Aceeași sursă prezintă și numărul de cazuri apărute în 2013 pentru fiecare tip de cancer în parte, dar și rata de mortalitate cauzată de aceste stări maligne în timpul aceluiași an.

Luând în considerare aceste cifre mari, devine clară nevoia urgentă de a găsi medicamente mai precise și eficiente, primul pas fiind înțelegerea mai amănunțită a *genomului cancerului* și a profilelor moleculare specifice pentru fiecare tumoră.

## **Terapii convenționale în cancer**

În momentul de față, atunci când se face referire la oncologie, terapiile convenționale pot fi descrise ca și un tratament ce este acceptat și folosit de majoritatea medicilor specialiști, implicând proceduri chirurgicale, administrarea de medicamente chemoterapeutice sau folosirea radițiilor.

Pe lângă posibila dobândire a rezistenței la terapie, un alt aspect important ce îi implică atât pe medici, cât și pe pacienți, este reprezentat de managementul efectelor secundare negative ale acestor medicamente, ce sunt determinate în mare parte de activitatea lor nespecifică și administrarea sistemică.

## **Terapii țintite**

Datorită dezavantajelor chemoterapiei tradiționale, noi metode țintite sunt în curs de dezvoltare, constând în medicamente ce sunt capabile să atace mecanisme celulare și moleculare specifice ce diferențiază celulele transformate de cele sănătoase (13). În funcție de particularitățile lor, aceste medicamente noi se clasifică în câteva grupuri, precum enzime sau inhibitori ai angiogenezei, anticorpi monoclonali și terapii genice.

Cu toate că numărul de studii efectuate pe subiecți umani este limitat, aceste noi medicamente țintite dețin așteptări majore pentru găsirea de terapii oncologice mai eficiente, încercările curente fiind susținute de cantitatea în continuă creștere de cunoștințe despre cancer.

## **Cunoștințe curente despre cancer**

În general, tumorile maligne sunt descrise ca și un grup de boli ce sunt caracterizate prin diviziune monoclonală intensă a unui tip particular de celule, ce invadează țesutul înconjurător și dobândește capacitatea de a coloniza țesuturi și organe (14, 15). În anii recentți, cancerul a fost descris ca și o boală poligenică, cu multiple fațete, caracterizată prin dezvoltarea de tumori generate de modificări genetice și/sau epigenetice, ce contracarează mecanismele de reparare a ADN și alte sisteme moleculare de protecție, concizând cu expresia fenotipului malign.

Alterările pot avea loc în orice pas descris de către dogma clasică a biologiei moleculare. Astfel, genele pot achiziționa mutații cu potențialul de a afecta procesul de traducere, rezultând proteine modificate, fapt ce poate avea efecte devastatoare în cazul

genelor supresoare de tumori (16, 17). Analiza datelor generate de secvențierea de nouă generație poate aduce lumină asupra tipurilor de mutații ce au loc, asupra frecvenței lor sau asupra posibilului rezultat (18-20). În alte situații, alterările pot avea loc la nivelul ARN mesager și apar ca și modificări la nivelul de expresie, sub forma unor sub- sau supraexprimări. Cu ajutorul tehnologiei microarray, produșii de transcriere pot fi monitorizați în ceea ce privește gradul de exprimare, fie atunci când se studiază particularitățile țesutului tumoral comparat cu omologul său normal sau atunci când, spre exemplu, se evaluează efectul unui tratament. Astfel, este posibilă capturarea expresiei unui număr mare de transcripți și, prin aplicarea unor algoritmi bioinformatici, generarea de liste de gene exprimate diferențiat, semnificative din punct de vedere statistic, ce pot fi interpretate apoi sub formă de termeni cu semnificație biologică (21, 22).

### **Tehnologii de tip "high throughput"**

Dezvoltarea de noi tehnologii de tip "high throughput" reprezintă atât o cauză, cât și un efect al acestei acțiuni. Pe de-o parte, noile instrumente de studiu molecular, precum secvențierea de nouă generație și platformele microarray, au făcut posibilă descoperirea și caracterizarea profilelor moleculare atât ale tipurilor de tumori, cât și ale pacienților. Pe de altă parte, aceste descoperiri au obligat rafinarea ulterioară a acestor tehnici, pentru a se potrivi nevoilor în creștere generate de medicina personalizată.

NGS, sau "secvențierea de a doua generație" atunci când este comparată cu metoda Sanger de terminare a lanțului, oferă posibilitatea de a realiza secvențieri masive paralele de genomuri complete, transcriptom, miRnom, secvențiere țintită de ampliconi sau interacțiuni ADN-proteine, cu costuri ce continuă să scadă, în timp ce performanța crește.

În timp ce NGS este în mare parte direcționat spre identificarea și caracterizarea variațiilor structurale, tehnologia microarray este folosită în principal pentru a efectua studii funcționale, precum expresia genică sau cuatificarea microARN. Astfel, este posibilă identificarea de anomalii moleculare de la nivelul întregului genom ce sunt conectate cu dezvoltarea și progresia tumorilor.

### **Analizele bioinformatică a datelor de tip "high throughput"**

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” (23, 24).

### **Importanța “Big Data”**

Datele de tipul “high throughput”, produse fie prin experimente proprii sau descărcate din baze de date publice, pot deveni o unealtă valoroasă în mâinile cercetătorilor. Aceste date, de la mutații în genele supresoare de tumori, până la reglarea post-transcripțională, expresia genică și inhibiția ARNm via microARN, au potențialul de a produce îmbunătățiri în ceea ce privește managementul terapeutic al pacienților oncologici. Așa cum s-a menționat anterior și va fi evidențiat în continuare, rezultatele acestor analize moleculare pot ajunge în domeniul clinic prin numeroase căi (25, 26).

Pe de-o parte, tehnologiile moderne duc la dezvoltarea de noi medicamente țintite, specifice, concepute pentru fiecare tip particular de cancer. În același timp, abilitatea de a determina profilul molecular pentru fiecare pacient va oferi clinicienilor posibilitatea de a alege tratamentul ce țintește cel mai eficient particularitățile fiecărei tumori în parte, cu eficiență maximă și efecte secundare reduse (27, 28).

Altă arie de interes pentru potențialul translational de interpretare al datelor de tipul “high throughput” este prevenția, mai exact prin dezvoltarea de metode de screening mai sensibile. Studii recente au demonstrat potențialul biomarkerilor, inclusiv microARN, ce pot deveni baza dezvoltării testelor de screening capabile de a reduce în final ratele de incidență, și implicit mortalitate, provocate de tumorile maligne (29, 30).



### **3. CONTRIBUȚIE PERSONALĂ: Managementul datelor pentru numeroase aplicații “omics”**

#### **3.1. STUDIUL NGS: Analizele de bioinformatică pentru secvențierea de nouă generație în cancerul de sân triplu negativ**

(Părți din acest capitol au fost publicate în Clujul Medical)(31)

#### **Cancerul de sân și TNBC – caracterizare, nevoie de management mai bun**

Conform ultimelor rapoarte despre incidența cancerului centralizate de GLOBOCAN, în anul 2012 malignitățile de sân se situau pe locul doi în topul celor mai comune tipuri de cancer la nivel global, cu aproximativ 1.67 milioane de cazuri noi diagnosticate în cazul femeilor, plasând cancerul de sân pe primul loc în lista celor mai comune cazuri oncologice la femei, atât în țările dezvoltate, cât și în cele în curs de dezvoltare. În România, cancerul de sân reprezintă cea mai comună malignitate din rândul femeilor, având în 2012 o incidență de 25.22%. De asemenea, reprezintă principala cauză de deces în femeile diagnosticate cu cancer, cu o rată de mortalitate de 16.74%. Cu privire la prevalența acestui tip de cancer, EUCAN prezintă următoarele valori: 12.58% pentru un an, 34.54% pentru 3 ani și o prevalență de 5 ani de 52.88% (11).

Cancerul de sân triplu negativ cuprinde tumori ce nu exprimă receptorul pentru estrogen, receptorul pentru progesteron și receptorul pentru factorul uman de creștere epidermal 2 (HER2) și reprezintă aproximativ 15-20% din totalul de cazuri de cancer de sân diagnosticate (32). Prin lipsa exprimării receptorilor pentru hormoni, tumorilor cancerului de sân triplu negativ le lipsește principala țintă terapeutică folosită în terapiile hormonale, ce se adaugă faptului că aceste subtipuri de tumori mamare sunt mai agresive, prezintă șanse mai scăzute de supraviețuire a pacientului și apar la vârste mai tinere (33-35).

Fiecare tip de cancer afișează mutații somatice specifice ce pot influența oncogeneza, iar diferitele subtipuri de cancer de sân nu fac excepție (36). Anumite studii prezintă importanța folosirii secvențierii de nouă generație pentru evaluarea mutațiilor din diferite gene și tipuri de cancer, datorită sensibilității mai crescute a acestei metode și capacității de a identifica mai multe mutații decât metoda secvențierii Sanger, fiind de asemenea capabilă să genereze date de tip “high throughput”(37-40).

Scopul acestui studiu constă în aplicarea bioinformaticii în analiza mutațiilor somatice din 46 de gene implicate în 31 de tumori de cancer de sân triplu negativ, evaluate cu ajutorul secvențierii de nouă generație, și anume platforma de secvențiere Ion Torrent PGM (Life Technologies).

## **Materiale, metode și protocoale**

### Aspecte Etice

Toate protocoalele experimentale au fost supervizate și aprobate de către Comitetul de Etică al Institutului Oncologic “Prof. Dr. I Chiricuta” și al Universității de Medicină și Farmacie “Iuliu Hațieganu” Cluj-Napoca. Toți pacienții incluși în studiu au citit și semnat formularul de consimțământ, ce se supune cerințelor legale naționale și europene.

### Pacienți

Acest studiu include 31 de pacienți ce au fost diagnosticați cu cancer de sân triplu negativ, diagnostic ce a fost stabilit folosind criterii universale. Toți pacienții prezintă o urmărire de 5-6 ani.

### Tehnologia Ion Torrent

Ion Torrent, numit și “Personal Genome Machine” (PGM™) distribuit 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, nu necesită dNTP marcate fluorescent sau chemiluminiscente, sau o detecție bazată pe imagine a nucleotidelor incorporate. În loc de depistarea pirofosfatului eliberat în timpul incorporării nucleotidelor, chip-ul construit pe baza tehnologiei semiconductoare detectează protonul ce este de asemenea eliberat în timpul acestui proces. Cu alte cuvinte, chip-ul Ion Torrent funcționează ca și un pH-metru, sesizând schimbările subtile de pH ce pot să apară la formarea unei legături fosfodiesterice din timpul elongării catenei de secvențiere (41).

Fluxul de lucru pornește odată cu efectuarea de librării de ADN. ADNc sau genomic este în primă fază fragmentat în catene de 200-400 pb, capetele sunt reparate, iar librăria este amplificată și purificată; apoi se atașează adaptorii specifici la capetele fragmentelor, iar golurile sunt completate. Atunci când se efectuează secvențiere multiplex, precum secvențierea ADN pentru mai mult de un pacient, fiecărui fragment i se adaugă un cod specific. Apoi biblioteca este selectată pe bază de mărime și verificată calitativ, fie prin folosirea unui Bioanalizator Agilent sau prin metoda tradițională de PCR cantitativ (PCRc). Următorul pas constă în amplificarea via PCR în emulsie (1). După amplificare, biblioteca

de ampliconi este supusă unui proces de îmbogățire. Amplificarea PCR și îmbogățirea bibliotecii sunt conduse folosind sistemul “Ion OneTouch”, ce automatizează procesul și reduce erorile cauzate de manipularea manuală (42). După acest pas, librăria ADN este încărcată pe chip și supusă centrifugării pentru a asigura ocuparea eficientă a chip-ului de către matriță, iar procesul de secvențiere este apoi inițiat (43).

### Extracția ADN

ADN a fost extras din țesutul incorporat în parafină (FFPE) folosind kitul PureLink Genomic DNA Mini Kit produs de Invitrogen, conform protocolului aferent kit-ului. Acesta permite izolarea și purificarea ADN genomic într-o manieră rapidă și eficientă, începând de la o varietate de probe biologice, precum celule, țesut din parafină sau înghețat, sânge integral, etc.. Principiul pe care se bazează acest kit constă în abilitatea ADN de a se lega selectiv de membrana de silicagel din coloane, în prezența de săruri chaotropice. Un alt avantaj prezentat de acest kit îl constituie faptul că folosește soluție de tampon puternică, capabilă de liza celulelor și țesuturilor prin simpla incubare la 55°C în prezența enzimei proteinază K, fără a fi nevoie de folosirea unei metode mecanice pentru ruperea membranelor celulare.

### Cuantificarea ADN

Cantitatea de ADN obținută în timpul extracției și purificării a fost cuantificată cu ajutorul spectrofotometrului NanoDrop-1000 (Thermo Scientific). Avantajul folosirii acestui instrument pentru cuantificarea acizilor nucleici constă în timpul scurt de efectuare a determinării și cantitatea mică de ADN nediluat (1  $\mu$ l), așadar acest tip de analiză poate avea loc imediat după extracție și purificare.

### Prepararea bibliotecii de ampliconi

Pentru fiecare probă, biblioteca de ampliconi a fost preparată urmând strict pașii din protocolul producătorului, folosind o serie de kit-uri Ion AmpliSeq Kits de la Applied Bioscience. Aceste kit-uri și amorsele aferente au fost concepute cu scopul folosirii lor în reacția PCR multiplex utilizată pentru un număr mare de ținte genomice, asigurând în același timp specificitate ridicată și acoperire uniformă.

### Amplificarea ADN genomic cu amorse specifice

Amplificarea ADN genomic a fost efectuată folosind kit-ul “Ion Ampliseq Cancer Panel” de la Applied Bioscience, ce conține amorse pentru 46 de gene ce sunt cunoscute pentru faptul că prezintă mutații legate de cancer (AKT1, BRAF, FGFR1, GNAS, IDH1, FGFR2, KRAS, NRAS, PIK3CA, MET, RET, EGFR, JAK2, MPL, PDGFRA, PTEN, TP53, FGFR3, FLT3, KIT, ERBB2, ABL1, HNF1A, HRAS, ATM, RB1, CDH1, SMAD4, STK11, ALK, SRC, SMARCB1, VHL1, CTNNB1, KDR, FBXW7, APC, CSF1R, NMP1, SMO, ERBB4, CDKN2A, NOTCH1, JAK3, PTPN11). Prin folosirea a doar 10 ng de ADN genomic, acest kit oferă acoperire de 97% a genelor țintă, acoperind 739 de mutații cunoscute înregistrate în baza de date COSMIC (Catalogue of Somatic Mutations in Cancer).

### Purificarea ampliconilor

Pasul de purificare este necesar pentru a înlătura orice resturi rămase de ADN genomic și amorse și pentru a asigura calitatea înaltă a ampliconilor. Procesul are loc folosind un suport magnetic DynaMag-2 și reactivul “Agencourt AMPure XP” de la Beckman Coulter ce conține particule magnetice ce leagă fie ADN genomic sau ampliconii, în funcție de rata volumetrică dintre particulele magnetice și probă.

### Digestia parțială a amorseelor

Pentru a putea continua cu pasul de secvențiere a ampliconilor, amorsele folosite pentru amplificarea țintită prin PCR trebuie îndepărtate.

### Marcarea ampliconilor

După pasul de purificare, ampliconii trebuie marcați cu un cod specific, ce permite desfășurarea reacțiilor multiplex posibile pe secvențiatorul Ion Torrent.

Amplificarea și Nick-translation a bibliotecii marcate are loc conform protocolului producătorului, folosind reactivul Platinum PCR SuperMix High Fidelity și Library Amplification Primer Mix.

### Cuantificarea bibliotecii de ampliconi

Cuantificarea bibliotecii are loc folosind kit-ul “Ion Library Quantification Kit” (Life Technologies) și platforma ViiA7 RT-PCR (Life Technologies), urmărind în același timp instrucțiunile producătorului. Acest pas include prepararea unor serii de diluții pentru

biblioteca de control *E coli* DH10B, fiind cunoscută concentrația inițială, diluții ce sunt folosite pentru efectuarea unei curbe standard pentru determinarea concentrației probelor de ampliconi.

După ce toate bibliotecile de ampliconi au fost diluate la o concentrație uniformă, probele au fost repartizate în grupe de 4 într-o manieră echimolară, în vederea procesării ulterioare și a secvențierii. Acest pas a fost posibil datorită capacităților multiplex oferite de folosirea codurilor unice.

### Amplificarea bibliotecii ISP (Ion Sphere Particles)

“Ion Sphere Particles” (ISPs) reprezintă particule mici de polistiren de care sunt atașați ampliconii via adaptorilor AP ce au fost legați anterior de biblioteci. Cu ajutorul tehnologiei, librăriile sunt supuse unei amplificări clonale în sistemul Ion OneTouch, sistem ce face parte din platforma Ion Torrent PGM.

### Îmbogățirea ISP

Acest pas din procesul de preparare a bibliotecilor asigură o încărcare eficientă a chip-ului de secvențiere și o calitate superioară a secvențierii în sine. Scopul îmbogățirii ISP constă în eliminarea particulelor ISP ce nu au librării de ampliconi atașate, iar acest pas este obținut cu ajutorul complexului biotină-streptavidină.

### Încărcarea chip-urilor “316”

Înainte de încărcarea ISP pe chip-urile de secvențiere, anumiți reactivi trebuie adăugați bibliotecilor, precum particule de control, amorse și enzima polimerază ce va adăuga nucleotidele prin formarea unei legături fosfodiesterice. Apoi, fiecare grup de 4 librării codate a fost încărcat pe chip-urile “316”, ce sunt folosite pentru obținerea a 100 Mb de date. Chip-ul este plasat pe platforma Ion Torrent PGM, ce a trecut anterior printr-o serie de etape de spălare și inițiere. Torrent Server, ce este conectat la aparatul PGM, permite utilizatorului să monitorizeze online procesul de secvențiere, inclusiv parametrii de încărcare pe chip, procentul de ISP cu matrițe, și particulele de clonare, cantitatea de complexe primer-dimer, etc

Programul Torrent Suite V4.4 preinstalat pe server-ul conectat la platforma de secvențiere a fost folosit pentru a realiza primii pași ai analizei datelor, în principal

procesarea semnalului și identificarea bazelor. Toate secvențele obținute au fost aliniate la Genomul Uman 19 (hg19), iar variantele au fost identificate folosind Variant Caller 4.4.0.6 cu scopul de a detecta mutațiile somatice, setând parametrii regiunilor țintă pentru panelul AmpliSeq CP.20131001.

Tabelul 3.1.6 arată anumite date statistice pentru fiecare chip folosit pentru secvențierea celor 31 de probe tisulare de cancer de sân triplu negative, informații obținute din raportul generat de programul Torrent Suite la sfârșitul fiecărei ture de secvențiere.

Tabel 3.1.6 Date statistice pentru cele 9 chip-uri folosite pentru secvențierea probelor

Tura (numărul chip-ului)	Numărul total de baze (Mbp)	Numărul total de baze Q20 (Mbp)	Numărul total de citiri	Lungime medie (bp)	Cea mai lungă citire (bp)	Cea mai lungă aliniere AQ20 (bp)
1	142.59	123.68	1,661,227	85	197	167
2	180.07	153.88	2,125,563	84	195	170
3	78.93	65.52	983,722	80	202	166
4	14.65	13	173,907	84	195	169
5	27	24.2	327,619	82	190	165
6	208.91	191.85	2,731,127	76	368	270
7	217.34	195.63	2,878,130	75	389	241
8	50.8	47.26	666,223	76	340	252
9	201.16	175.6	2,728,373	73	377	265

Odată ce variantele prezentate în proba de secvențiere au fost identificate prin aplicarea unor „plug-in”-uri prezente în programul Torrent Suite, aceste modificări de nucleotide au trebuit caracterizate. În acest sens, fișierele *.vcf* generate pentru fiecare din cele 31 de probe au fost descărcate și transferate pe un calculator pe care rulează Ubuntu 10.04, iar pentru a analiza datelor și adnotarea variantelor a fost folosit programul ANNOVAR în linie de comandă (44). Aceste unelte folosesc informații din diferite baze de date precum COSMIC, 1000genomes, refGene, dbSNP, sau ClinVar, pentru a adnota mutațiile descoperite în experimentul de secvențiere, dar și pentru a atribui un scor PolyPhen-2 (Polymorphism Phenotyping version2) mutațiilor nonsinonime, estimând astfel posibilele daune și statusul benign.

## Rezultate și discuții

Experimentul de secvențiere condus pe cele 31 de probe de țesut la parafină de la pacienți diagnosticați cu cancer de sân triplu negativ a evaluat mutațiile din 46 de oncogene și gene supresoare de tumori, din care s-au evidențiat mutații în 37 de gene. După filtrarea variantelor cu parametri de calitate reduși, numărul total de mutații găsite în cele 37 de gene a ajuns la 165. Variantele sinonime au fost de asemenea excluse, datorită faptului că aceste tipuri de substituții nucleotidice nu determină nicio modificare a aminoacizilor.

Evaluarea clinică a acestor mutații a fost făcută conform scorului PolyPhen-2, ce prezice modificările funcționale de la nivelul proteinelor ca rezultat al mutațiilor punctiforme non-sinonime (nsSNP). Din cele 165 de mutații identificate în probele TNBC, 70 au fost clasificate ca și dăunătoare, 22 posibil dăunătoare și 27 au fost considerate benigne, conform scorului PolyPhen-2 pentru mutații somatice. Celelalte 46 de mutații nu au avut semnificație clinică conform acestui algoritm (Figura 3.1.8).

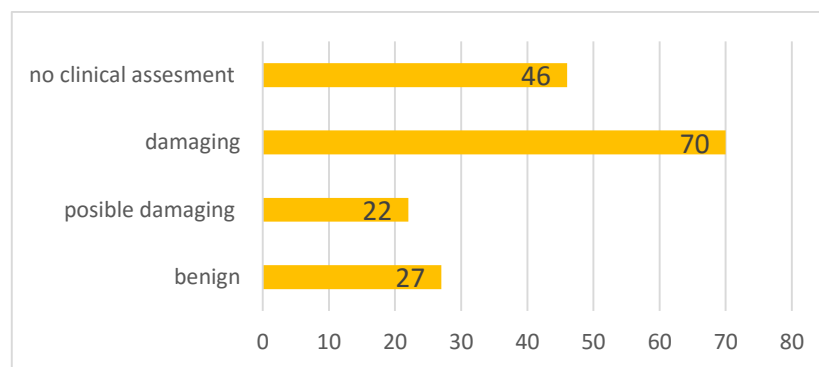


Figure 3.1.8 Semnificația clinică a mutațiilor identificate conform scorului PolyPhen-2

Din numărul total de variante, 25 erau cunoscute – din care 16 au fost deja descrise în baza de date COSMIC - în timp ce celelalte 140 erau mutații noi. Cele 138 de mutații exonice au fost grupate în patru categorii: mutații missense – ce de obicei sunt mutații punctiforme ce cauzează schimbarea unui aminoacid; mutații nonsense – unde schimbul de nucleotidă conduce la apariția unui codon stop prematur, ducând la apariția unui ARNm nefuncțional; mutații de decalare a cadrului de citire (frameshift) – unde se modifică codonul și se schimbă cadrul de citire, generând un model translational diferit, și deleții fără schimbare a cadrului de citire (no frameshift deletions) – ce reprezintă deleții de multipli de trei nucleotide, ce nu modifică citirea codonilor, dar conduc la deleții în numărul corespunzător de aminoacizi, generând astfel o proteină modificată.

Majoritatea mutațiilor ce nu au avut niciun impact clinic ca rezultat al analizei datelor au fost observate în proba 14. Dintre aceste mutații, cea cu cea mai mare frecvență în grupul studiat s-a regăsit în gena *AKT1*, o mutație intronică  $G > A$ , deja prezentă în baza de date dbSNP, ce a fost identificată în 9 probe.

Mutațiile posibil dăunătoare, cărora li s-a atribuit un scor PolyPhen-2 ce variază între 0.505 și 0.956, au fost observate în 13 gene, dintre care cele mai frecvent mutante au fost *KDR*, cu mutații observate în 12 probe. Celelalte mutații au fost observate în mare parte doar într-o singură probă, cu excepția unora – una în *ATM* și una în *SMAD4*, ce au fost observate în două probe din 31. Mutațiile cu cel mai mare scor PolyPhen-2, încadrat între 0.965 și 1.0, au fost observate în 29 dintre gene, majoritatea fiind găsite doar într-o probă.

Din totalul de 37 de gene ce au fost găsite mutante în acest experiment, genele ce au prezentat cele mai frecvente mutații au fost *TP53* (mutant în 19 probe), *KDR* (în 13 probe), *PIK3CA* (în 12 probe), *AKT* și *ATM* (fiecare mutante în 11 probe), *JAK3* (în 5 probe), *MET*, *SMAD4*, *FGFR2* și *FGFR3* (fiecare mutante în 4 probe), *PTEN*, *ABL1* și *ERBB4* (fiecare găsite mutante în 3 probe). Aceste gene prezintă un interes nu doar prin faptul că sunt cel mai frecvent mutante în probele noastre și fiecare având cel mai mare număr de mutații individuale, dar și prin faptul că au fost caracterizate în numeroase studii ca gene cu mutații legate de cancer. Fiind implicate în multe căi de semnalizare și procese celulare precum ciclul celular, repararea ADN, creștere și proliferare, orice modificare ce apare în aceste gene are potențialul de a genera sau de a fi implicate în modificări conectate cu tumorigeneza, progresia sau răspunsul la terapie (45, 46).

Aplicarea metodelor de bioinformatică cu scopul de a descoperi mutații prezente în oncogene sau gene supresoare de tumori de la pacienți cu cancer de sân și cu accent asupra sub-tipului mai agresiv triplu negativ, prezintă chiar mai multă importanță atunci când sunt corelate rezultatele cu informațiile clinice ale pacienților, precum existența unor posibile metastaze. Genele ce au prezentat cele mai multe mutații în probele de la pacienți cu metastaze au fost *TP53*, *KDR*, *AKT1* și *PIK3CA*, ce de asemenea coincid cu cele mai mutante gene din din toate cele 31 de probe.

Rezultatele generate de studiul prezent au fost în concordanță cu activitatea de cercetare condusă de alte echipe de cercetare. Analizând probe tisulare de cancer de sân sau plasma de la pacienți cu malignități mamare și folosind aceeași platformă de secvențiere, Ion Torrent Personal Genome Machine, și aceleași kituri de preparare a librăriilor – reactivi AmpliSeq, alte echipe au descoperit de asemenea că numărul cel mai mare de mutații



somatrice se regăesc în genele *PIK3CA*, *PTEN*, *AKT1*, *TP53*, *SMAD4*, oferind un prim grad de validare acestui studiu (47, 48).

### Validarea mutațiilor identificate

Pentru a lua în considerare inițierea posibilității de a traduce analiza bioinformatică a datelor SNG în date cu semnificație clinică pentru diagnostic, prognostic sau cu scop de biomarkeri, este obligatoriu să se valideze mutațiile obținute cu ajutorul unei tehnici bazate pe alte principii moleculare decât secvențierea. În această situație, metoda aleasă a fost validarea TaqMan SNP Genotyping de la Life Technologies. Activitatea acestui test este bazată pe două sonde ce sunt complementare cu alelele sălbatice și respective cu cele mutante. Prin folosirea unor coloranți fluorescenți diferiți pentru fiecare probă, testul dezvăluie prezența celei mai frecvente alele, validând sau infirmând mutațiile identificate în timpul analizei bioinformatică a datelor.

Am ales un număr total de 9 mutații diferite găsite în mai mult de o probă, din top 13 cele mai frecvent mutate gene. Validarea prin genotipare SNP a fost efectuată pentru toate cele 31 de probe supuse secvențierii de nouă generație. O comparație între frecvența mutațiilor, exprimată în procente, obținută după efectuarea celor două metode este prezentată în figura 3.1.11.

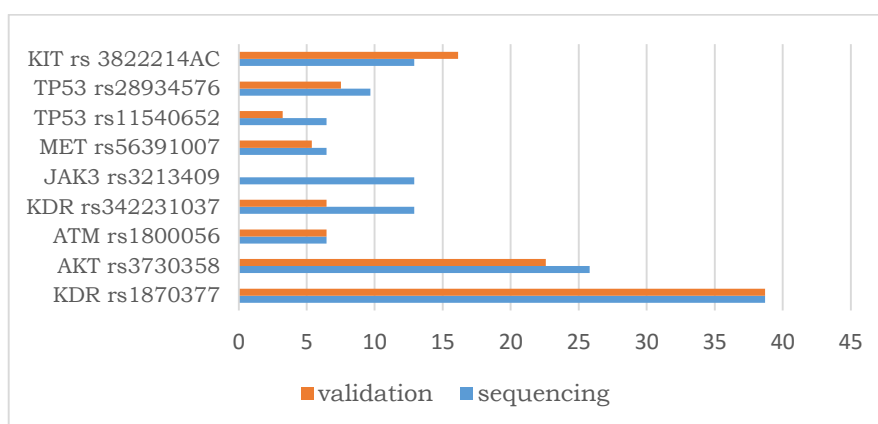


Figura 3.1.11. Frecvența mutațiilor identificate în analizele de secvențiere și validare

Mutațiile observate după analiza datelor NGS au fost validate folosind testul de genotipare SNP, cu excepția genei *JAK3*. Pentru aceasta, mutația ce a fost identificată cu tehnologia NGS în câteva din probele analizate, nu a fost găsită în timpul procesului de validare, cel mai probabil din cauza lipsei amplificării în etapa de PCR a studiului. Acest

fapt se poate datora cantității și/sau calității insuficiente a probei de ADN folosită, ținând cont că acizii nucleici au fost extrași direct din secțiuni obținute din blocuri de parafină, fără a mai trece printr-o microdisecție preliminară.

Cu toate acestea, faptul că o bună majoritate a mutațiilor punctiforme ce au fost identificate în timpul secvențierii au fost validate cu testul de genotipare, demonstrează că datele obținute în urma Secvențierii de Nouă Generație, analizate cu instrumente bioinformatică potrivite și interpretate corect, au potențialul de a deveni un instrument clinic folositor, ducând în final la un management mai bun al bunăstării pacienților.

### **Limitări și considerări ulterioare**

Ca și o limitare pentru acest studiu – ce reprezintă în același timp o motivare în a continua – este amintit faptul că a fost condus pe un număr relativ mic de pacienți, mărimea cohorței fiind constrânsă de nevoia de a avea, pe lângă material biologic potrivit, rezultate semnificative ce pot fi urmărite și informații recente despre starea vitală a subiecților. O altă limitare a fost reprezentată de informațiile relativ incomplete despre pacienți, ce a îngreunat realizarea de corelații semnificative dintre mutațiile observate în probe, datele clinice și rezultatul final. Aceste limitări trebuie depășite de studii ulterioare pe grupuri mai mari de pacienți, cu rezultate ulterioare și informații clinice suficiente, pentru a fi posibilă corelarea semnificativă între mutații și fenotipurile pe care le-ar putea determina.

### **Concluzii**

Prin secvențierea a 31 de probe tisulare de cancer de sân triplu negativ provenite din blocuri de parafină, folosind platforma Ion Torrent PGM pentru secvențierea de nouă generație, am obținut atât mutații cunoscute cât și necunoscute în 34 de gene implicate în numeroase căi de semnalizare ce au loc în timpul dezvoltării și progresiei tumorale. Prin folosirea metodelor de bioinformatică pentru analiza datelor cu ajutorul unor alinieri succesive cu diferite baze de date, am reușit să identificăm o serie de mutații, o parte regăsindu-se în literatură ca fiind implicate în cancerul de sân, fapt ce demonstrează, până la un anumit punct, eficiența acestei metode. Prin folosirea metodei de genotipare SNP bazată pe tehnologia PCR, am validat o parte din mutațiile prezente în mai mult de o probă, în opt dintre cele 13 cele mai frecvent mutate gene, demonstrând eficiența analizelor de bioinformatică aplicate datelor obținute în urma Secvențierii de Nouă Generație. Aceste rezultate preliminare generate de studiul ce se situează printre primele desfășurate asupra

populației din România, demonstrează că bioinformatica reprezintă o unealtă puternică ce poate completa golul dintre cantitatea mare de date obținută în urma efectuării secvențierii de nouă generație și interpretarea și integrarea informației semnificative ce poate fi folosită în final în beneficiul pacienților.

### **3.2. STUDIU MICROARRAY: Analiza bioinformatică a datelor microarray și extrapolarea porc-om în relație cu expunerea la Zearalenone și *Escherichia coli*** (Părți din acest capitol au fost submise spre publicare în PLOS ONE)

#### **Modele animale – când și de ce le folosim?**

Modelele animale au ajutat la înțelegerea unui număr mare de boli, precum cancerul, diabetul, boli cardiovasculare, osteoporoza, HIV, dar și a unor diferite tulburări ale sistemului nervos. Multe studii comparative au condus nu numai la progresul medicinei umane, dar și la acumularea rapidă de cunoștințe referitoare la medicina veterinară, ajutându-ne astfel să diagnosticăm mai precis și să îngrijim animalele noastre de companie și cele ce împart habitatul cu noi. În același timp, multe specii de animale sunt crescute de către oameni ca hrană sau pentru a fi folosite în procese biotehnologice farmaceutice. Aceste efective de animale trebuie ținute în condiții optime de sănătate, iar rezultatele studiilor comparative pot fi de asemenea folositoare și pentru aceste tipuri de întreprinderi (49-51).

Pe lângă animale tradiționale de laborator pentru precum șoareci sau șobolani, și alte specii au devenit folositoare pentru diferite studii preclinice, cum ar fi primatele, porcii, vacile și alte animale domestice. În primul rând, multe dintre ele împart aceleași boli și același tip de răspuns la tratamente corespunzătoare cu oamenii. Aceste similarități avansează și mai profund, la nivel celular și subcelular, împărțind căi moleculare și procese biochimice echivalente (52-54). Dar aceste similarități există și la scară mai largă, ținând cont de faptul că oamenii și animalele locuiesc în același habitat, fiind expuși la aceiași factori ecologici și la aceleași condiții de trai.

În consecință, folosirea modelelor de animale pentru cercetări preclinice, în special în cazul studiilor de tip ”proof of concept”, este necesară și deține promisiuni pentru viitoare dezvoltări interdisciplinare. Dar, așa cum vom demonstra în continuare, rezultatele generate de aceste studii pot fi analizate și interpretate folosind unelte de bioinformatică cu scopul de a le extrapola la oameni.

#### **Introducerea și motivația acestui studiu**

Zealarenona (ZEA) este un metabolit secundar produs de anumite specii ce aparțin genului *Fusarium*, un grup comun de fungi (55, 56). Acestea sunt toxine comune prezente în aproape toate tipurile de recolte, inclusiv cele folosite pentru hrana animalelor de fermă, dar și cereale consumate de către oameni (57-59). Altă sursă de contaminare pentru ambele

specii este bacteria *Escherichia coli*, un bacil Gram negativ ce include numeroase tulpini ce populează majoritar intestinele multor specii de animale. Intestinul reprezintă o barieră importantă între organism și mediul înconjurător, așadar interacționează extins atât cu microflora existentă, dar și cu eventualii agenți patogeni ce pot să apară (60-62).

Modelul animal folosit în acest studiu este porcul, datorită asemănărilor histopatologice amintite mai devreme dintre *Sus scrofa* și *Homo sapiens*. De asemenea, porcii consumă cantități mari de porumb, cereală ce este predispusă la infecția micotoxică cu *Fusarium* (63) și care este și un aliment de bază în alimentația oamenilor.

### **Tehnologia Microarray – Privire de ansamblu**

Tehnologia microarray poate fi folosită pentru caracterizarea profilului expresiei genice a tumorilor, luând în considerare faptul că expresia genelor – în special a celor implicate în oncogeneză, precum oncogenele și genele supresoare de tumori – este direct responsabilă de comportamentul celulelor transformate (22, 64). Prin analiza simultană a unui număr mare de transcripti și producerea consecutivă a unei cantități mari de date, această tehnologie deține abilitatea de a captura modificările relaționate cu tumorigeneza la nivel de genom complet.

### **Materiale, metode și protocoale**

Probele biologice ce constau în 15 probe tisulare de splină și 15 de duoden au fost colectate de la porci nou-născuți. Animalele au fost expuse anterior unei contaminări experimentale de 100 ppb cu zearalenone (ZEA) și *Escherichia coli*, fie ca și agenți de contaminare singulari, fie în combinație.

#### Extracția de ARN total cu TRI Reagent (Sigma-Aldrich)

Probele tisulare au fost omogenizate în prezența TRI Reagent-ului folosind mixer-ul Polytron. Apoi, cantitatea totală de ARN a fost extrasă folosind metoda fenol-clorofom.

#### Purificarea ARN folosind RNeasy Mini Kit (Qiagen)

Purificarea cantității totale de ARN a fost condusă în concordanță cu protocolul recomandat de producătorul kitului de extracție.

## Cuantificarea ARN

Evaluarea cantitativă a ARN extras și purificat este realizată prin folosirea spectrofotometrului NanoDrop-1000 (Thermo Scientific), în timp ce calitatea probelor ARN a evaluată folosind Bioanalyzer 2100 (Agilent Technologies) ce practică o tehnică bazată pe electroforeză, miniaturizată la nivel de chip – tehnologia în sine fiind patentată sub numele de “Lab-on-a-Chip”. Prepararea sondelor pentru microarray, hibridizarea, scanarea și pre-procesarea datelor au fost conduse urmărind protocolul recomandat furnizat de către producător (Agilent).

## Sinteza sondelor

Sondele folosite pentru evaluarea expresiei genice prin microarray, ARNc-Cy3, au fost generate folosind metoda unei singure culori, metodă realizată cu ajutorul Agilent Low Input Quick Amp Labeling Kit (5190-2305).

## Purificarea sondelor

Kit-ul folosit pentru purificarea sondelor de microarray a fost RNeasy Mini Kit de la Qiagen.

## Cuantificarea sondelor și diluția

Cu scopul de a asigura parametrii potriviți pentru probele ARNc, a fost condus un control de calitate folosind aparatul NanoDrop-1000, prin selectarea secțiunii *Măsurare Microarray*, iar apoi randamentul și diluțiile au fost calculate folosind drept orientare datele furnizate de către producător în protocolul original (Agilent manual: G4140-90040).

## Hibridizarea sondelor

Sondele au fost hibridizate pe lame Agilent 8 x 60k, conținând un panel personalizat (AMADID 056850) compus din 60 de oligomeri pentru mai mult de 59,000 de sonde reprezentative pentru un număr mare de transcripti de *Sus scrofa*, folosind Gene Expression Hybridization Kit de la Agilent, conform protocolului.

## **Procesarea datelor**

Scanarea lamelor microarray generează un fișier de tip *.tiff* pentru fiecare probă, fișier ce poate fi încărcat în programul Feature Extraction, versiunea 11.0.1.1. Acest program

citește, interpretează și integrează fișierele microarray tip imagine prin introducerea în mod automat a grilei potrivite, marcare și excluderea pixelilor ce se găsesc ca fiind aberanți și prin calcularea intensității și ratei fiecărei trăsături. După generarea parametrilor pentru raportul de control al calității, programul generează apoi un fișier de ieșire tip *.txt* ce conține date numerice pentru analiza datelor.

### Analiza datelor

Analiza bioinformatică a datelor de microarray din probele porcine de duoden și splină a fost condusă cu ajutorul programului GeneSpring GX, versiunea 13.0, dezvoltat de Agilent Technologies. Fișierul *.txt* ce corespunde fiecărei probe a fost încărcat în program, iar fiecare probă a fost identificată și adnotată conform tipului de tratament/expunere la contaminare. Abordarea din studiul prezent a constatat în tratarea probelor din fiecare tip de țesut diferit, fapt ce creează în consecință două experimente diferite: unul pentru țesutul duodenal și unul pentru cel splenic. Analiza diferențială a fost efectuată după ce s-au aplicat filter și teste statistice.

Un factor de expresie (fold change) de 2.0 a fost aplicat întregilor liste pentru cele trei perechi de condiții din fiecare experiment: ZEA vs control, *E coli* vs control, și ZEA+*E coli* vs control, atât pentru duoden cât și pentru splină, iar transcriptiile au fost considerați semnificativi atunci când aveau o valoare *p* mai mica de 0.05.

Table 3.2.5. Numărul de transcripti exprimați diferențial și semnificativi din punct de vedere statistic în toate perechile de condiții

Tip de țesut	Perechi de condiții	FC	Valoare p	Nr de gene semnificative statistic	Număr de gene semnificative statistic cu FDR
Duoden	E coli_vs_ctrl	2.0	0.05	3,058	2,875
	Zea_vs_ctrl	2.0	0.05	4,023	0
	Zea_E coli_vs_ctrl	2.0	0.05	804	316
Splină	E coli_vs_ctrl	2.0	0.05	3,677	3,446
	Zea_vs_ctrl	2.0	0.05	467	23
	Zea_E coli_vs_ctrl	2.0	0.05	570	141

Dintre acești transcripți a căror expresie a fost considerată modificată ca rezultat la expunerea la contaminanți, unii sunt suprexprimați, în timp ce pentru unii transcrierea este semnificativ redusă.

### **Raționamentul pentru metoda de extrapolare**

În experimentul nostru de genomică funcțională și structurală am folosit paneluri de gene personalizate, fapt ce prezintă numeroase avantaje, dar și dezavantaje, datorită lipsei caracteristicilor comerciale, precum adnotări extinse și posibilități de extrapolare complete. Astfel, doar o parte din secvențele sondelor conținea un nume de genă corespunzător nomenclaturii cunoscute, iar unele dintre sonde conțineau nume de identificare ce variau din punct de vedere taxonomic, multe din ele nefiind incluse în bazele de date comune.

### **Sucesiunea de metode pentru extrapolare**

Pentru a putea extrapola rezultatele obținute, cu scopul de a identifica echivalentul la om al transcripțiilor exprimați diferit față de grupul de control din experiment, a fost necesară efectuarea unor etape succesive de alinieri și conversii, folosind programe ca și Galaxy Suite și Linux.

### **Rezultate și discuții**

Ca rezultat al acestei metode, am obținut un fișier ce conține lista de ID-uri pentru probele originale într-o coloană, iar în cealaltă, numele genei corespunzătoare la om. Cu ajutorul programului Excel, am adnotat în continuare fiecare din transcripții originali porcini exprimați diferențiat ce au fost generați de către analiza GeneSpring, cu ortologul său uman.

### **Concluzii parțiale pentru metoda de extrapolare**

Am reușit să identificăm echivalentul genelor umane ce sunt exprimate diferențiat într-o manieră semnificativă din punct de vedere statistic, înțelegând astfel mai bine modalitatea prin care co-contaminarea poate afecta atât porcii, cât și oamenii, două specii ce împart multe similarități anatomice, fiziologice, de dietă și de mediu. În această eră “omics”, unde situații similare pot fi întâlnite de către cercetători ce folosesc diverse modele animale,



noua noastră metodă de extrapolare de la animal la om a numelor de gene oferă o soluție inovativă pentru analiza eficientă și interpretarea datelor.

### **Efectul extrapolării co-contaminării la oameni**

După ce procesul de extrapolare s-a finalizat, a fost posibilă deducerea potențialelor efecte ale micotoxinei zearalenone, singulare sau atunci când întâlnește tulpini patogene de *E coli*, asupra unor părți importante ale tractului digestiv uman. Genele au fost ordonate conform status-ului lor, sub- sau supraexprimate, și ținând de asemenea cont de valoarea *p*.

### **Modelul de evaluare a expresiei genice în experimentul pe duoden**

Din cauza numărului mare de gene ale căror alterări influențează un număr la fel de mare de căi de semnalizare, precum și din cauza constrângerilor de spațiu, interpretarea rezultatelor a fost concentrată pe efectul zearalenonei asupra probelor de duoden. Raționamentul din spatele alegerii acestui organ se regăsește în faptul că duodenul este primul segment al intestinului subțire, reprezentând una din interfețele dintre organism și potențialii patogeni, prima barieră intestinală ce protejează corpul.

Uitându-ne la transcripții ce prezintă alterări semnificative, modelul ce reiese constă în implicarea acestor transcripți în căi de semnalizare ce includ procese moleculare conectate cu evenimente celulare ca și apoptoza, ciclul celular, replicare și diferențiere. Corelarea cu anumite semne distinctive ale cancerului confirmă ipoteza noastră conform căreia efectele zearalenonei contribuie la transformarea malignă, atât singură, cât și în co-contaminare cu *E coli*.

### **Analiza rețelelor de semnalizare**

Folosind programul Ingenuity Pathway Analysis (IPA) (Qiagen), am reușit să investigăm în continuare implicațiile moleculare și celulare a acestor gene a căror expresie a fost semnificativ alterată ca și răspuns la expunerea la ZEA la nivelul duodenului. Așadar, am identificat cele mai importante căi canonice de semnalizare, la fel ca și topul funcțiilor biologice și patologice ale celor mai afectate gene, prin introducerea în programul IPA a listei de gene generate după aplicarea unui fold change de 2.0.

Tabel 3.2.7. Top 5 căi canonice modulate de ZEA la nivelul duodenului

	<b>Căi canonice</b>	<b>Valoare-p</b>	<b>Suprapunere</b>
<b>1</b>	Rolul macrofagelor, fibroblaștilor și celulelor endoteliale în artrita reumatoidă	0.000000181	18.1 % 54/298
<b>2</b>	Semnalizarea TREM1	0.00000145	26.7 % 20/75
<b>3</b>	Semnalizarea kinazei proteice A	0.00000163	15.3% 59/385
<b>4</b>	Reglarea cancerului de sân de către Stathmin1	0.00000429	18.3 % 35/191
<b>5</b>	Feedback-ul Dopamine-DARPP32 în semnalizarea cAMP	0.00000522	19.3 % 31/161

Precum se poate observa, majoritatea moleculelor cu un pattern de expresie alterat aparțin căilor de semnalizare implicate în inflamație, creștere celulară și diferențiere, apoptoză, formarea citoscheletului, etc., procese ce sunt conectate cu transformarea malignă (65-67).

## Concluzii

Mulți dintre transcripții ce au prezentat niveluri de expresie alterate fac parte din căi de semnalizare implicate în procese celulare și moleculare cheie, dar și în promovarea inflamației, ce, după cum a fost demonstrat de numeroase studii, este conectată direct cu dezvoltarea și progresia tumorilor maligne. Așadar, se poate concluziona că expunerea prelungită la zearalenonă, o micotoxină comună găsită ca și contaminant în cereale, e posibil să fie corelată cu activarea unor procese moleculare inflamatorii de promovare a dezvoltării tumorilor (68).

Cu toate că informația în sine este importantă, valoarea studiului se poate regăsi în metoda nouă de bioinformatică ce a fost dezvoltată pentru acest experiment, fiind astfel posibilă extrapolarea rezultatelor la oameni, având în vedere faptul că aceste două specii împart numeroase caracteristici, de la anatomie, fiziologie, patologie, până la factori de mediu și contaminanți. Așadar, folosirea bioinformaticii pentru descifrarea informațiilor de tip “high throughput” generate de tehnologiile moleculare moderne și studiile pe modele de animale, pot contribui în final la îmbunătățirea vieții umane.

#### 4. CONCLUZII GENERALE

Scopul prezentei teze a fost de a evidenția modalitatea în care analizele bioinformatică pot să contribuie la personalizarea diagnosticului și terapiilor, prin folosirea de tehnologii bio-moleculare de tip "high throughput" cu scopul de a analiza profilul molecular al condițiilor fiziologice și patologice. Prin folosirea uneltelor și metodelor de analiză bioinformatică pentru interpretarea și integrarea cantității mari de date generate de aceste tehnologii, va fi în curând posibilă dezvoltarea de metode rapide și eficiente aplicabile în special în domeniul oncologiei.

Analiza, integrarea și interpretarea rezultatelor obținute în urma acestor experimente de cercetare duc la formularea următoarelor concluzii generale:

Pentru experimentul ce conține secvențierea de nouă generație, prin folosirea metodelor de bioinformatică pentru analiza datelor cu ajutorul alinierilor succesive cu diferite baze de date, am reușit să identificăm o serie de mutații, o parte din ele fiind deja descrise în literatură ca fiind implicate în cancerul de sân, fapt ce dovedește până la un anumit punct, eficiența metodei folosite. Unele mutații găsite nu au putut fi corelate în mod direct cu cancerul de sân, bazându-ne exclusiv pe literatură, însă au fost descrise ca fiind implicate în afectarea unor căi de semnalizare tumorigenice. Metoda de genotipare SNP bazată pe tehnologia PCR a contribuit la validarea unor mutații identificate, *demonstrând eficiența analizelor bioinformatică a datelor obținute în urma Secvențierii de Nouă Generație*. Mutațiile noi ce au fost descoperite sunt supuse nevoii de caracterizare ulterioară, dar aceste rezultate preliminare, printre primele de acest gen din România, demonstrează că *bioinformatica reprezintă o unealtă puternică, capabilă să completeze golul dintre cantitatea mare de date generate de secvențierea de nouă generație și interpretarea și integrarea lor sub forma unor informații semnificative ce pot fi folosite apoi în beneficiul pacienților*.

În ceea ce privește studiul microarray, mulți dintre transcripții ce au prezentat niveluri de expresie alterate sunt implicați în căi de semnalizare cruciale pentru anumite procese celulare și moleculare, corelate cu dezvoltarea și progresia tumorilor maligne. Așadar, se poate susține că expunerea prelungită la zearalenonă, o micotoxină comună găsită ca și contaminant în cereale, poate fi conectată cu activarea proceselor moleculare inflamatorii de promovare a tumorilor. Cu toate că informația în sine este importantă, valoarea studiului se poate regăsi și în metoda nouă de bioinformatică ce a fost dezvoltată

pentru acest experiment, fiind astfel posibilă extrapolarea rezultatelor la oameni, având în vedere că aceste două specii împart numeroase caracteristici, de la anatomie, fiziologie, patologie, până la factori de mediu și contaminanți. Așadar, *folosirea bioinformaticii pentru descifrarea informațiilor de tipul “high throughput” generate de tehnologiile moleculare moderne și studiile pe modele animale, pot contribui în final la îmbunătățirea vieții umane.* În era “omics”, atunci când situații similare pot fi des întâlnite de către cercetători ce folosesc diferite modele animale, metoda noastră nouă de extrapolare a genelor de la animal la om oferă o soluție inovativă pentru analiza și interpretarea eficientă a datelor.

## Lista publicațiilor

### Jurnale ISI

1. **Cojocneanu Petric R**, Braicu C, Raduly L, Zanoaga O, Dragos N, Monroig P, Dumitrascu D, Berindan-Neagoe I. Phytochemicals modulate carcinogenic signaling pathways in breast and hormone-related cancers. *Onco Targets Ther.* 2015 Aug 6;8:2053-66. doi: 10.2147/OTT.S83597
2. **Roxana Cojocneanu Petric**, Cornelia Braicu, Cristian Bassi, Laura Pop, Ionelia Taranu, Nicolae Dragos, Dan Dumitrascu, Massimo Negrini, Ioana Berindan-Neagoe. Interspecies Gene Name Extrapolation – A New Approach. Acceptat spre publicare în PLOS ONE
3. Zaharie F\*, **Cojocneanu-Petric R\***, Muresan M, Frinc I, Dima D, Petrushev B, Tanase A, Berce C, Chitic M, Berindan-Neagoe I, Pileczki V, Irimie A, Tomuleasa C. Small molecules against B-RAF (BRAF) Val600Glu (V600E) single mutation. *Int J Nanomedicine.* 2015 Jul 31;10:4897-9. doi: 10.2147/IJN.S87405
4. Tudoran O, Soritau O, Balacescu L, Visan S, Barbos O, **Cojocneanu-Petric R**, Balacescu O, Berindan-Neagoe I. Regulation of stem cells-related signaling pathways in response to doxorubicin treatment in Hs578T triple-negative breast cancer cells. *Mol Cell Biochem.* 2015 Jul 18
5. Cornelia Braicu, Valentina Pileczki, Laura Pop, **Roxana Cojocneanu Petric**, Sergiu Chira, Eve Pointiere, Patriciu Achimas-Cadariu, Ioana Berindan-Neagoe, Dual targeted therapy with p53 siRNA and epigallocatechingallate in a triple negative breast cancer cell model, *PLoS ONE* 04/2015; 10(4). DOI:10.1371/journal.pone.0120936
6. Cornelia Braicu, **Roxana Cojocneanu-Petric**, Sergiu Chira, Anamaria Truta, Alexandru Floares, Patriciu Achimas-Cadariu, Ioana Berindan-Neagoe. Clinical and pathological implications of miRNA in bladder cancer. *International Journal of Nanomedicine* 2015;10 1–10
7. Muresan M, Zaharie F, Bojan A, Frinc I, Dima D, Selicean S, Gafencu GA, Petrushev B, **Cojocneanu-Petric R**, Tefas C, Cioca A, Irimie A, Berce C, Berindan-Neagoe I, Tomuleasa C, Achimas-Cadariu P. MicroRNAs in liver malignancies. *Basic science applied in surgery. J BUON.* 2015 Mar-Apr;20(2):361-75
8. Claudia Gherman, Matea Cristian Tudor, Bele Constantin, Tabaran Flaviu, Razvan Stefan, Bindea Maria, Sergiu Chira, Cornelia Braicu, Laura Pop, **Roxana**

- Cojocneanu Petric**, and Ioana Berindan-Neagoe Pharmacokinetics Evaluation of Carbon Nanotubes Using FTIR Analysis and Histological Analysis. *J. Nanosci. Nanotechnol.* 15, 2865-2869 (2015)
9. Calin Ionescu, Cornelia Braicu, Roxana Chiorean, **Roxana Cojocneanu Petric**, Emilian Neagoe, Laura Pop, Sergiu Chira, Ioana Berindan-Neagoe. TIMP-1 expression in human colorectal cancer is associated with SMAD3 gene expression levels: a pilot study. *J Gastrointestin Liver Dis* 2014 Dec;23(4):413-8
  10. Laura-Ancuța Pop, Emil Puscas, Valentina Pileczki, **Roxana Cojocneanu-Petric**, Cornelia Braicu, Patriciu Achimas-Cadariu, Ioana Berindan-Neagoe. Quality control of ion torrent sequencing library. *Cancer Biomark* 2014 ;14(2-3):93-101
  11. Cornelia Braicu, Ioana Berindan-Neagoe, Valentina Pileczki, **Roxana Cojocneanu-Petric**, Laura-Ancuța Pop, Emil Puscas, Alexandru Irimie, Rares Buiga. Breast tumor bank: an important resource for developing translational cancer research in Romania. *Cancer Biomark* 2014 ;14(2-3):119-27
  12. A Irimie, C Braicu, **R Cojocneanu Petric**, I Berindan Neagoe, RS Campian , Novel technologies for oral squamous carcinoma biomarkers in diagnostics and prognostics. *Acta Odontologica Scandinavica.* 2014;73(3):161-8.
  13. Berindan-Neagoe I, Braicu C, Pileczki V, **Cojocneanu Petric R**, Miron N, Balacescu O, Iancu D, Ciuleanu T. 5-Fluorouracil potentiates the anti-cancer effect of oxaliplatin on Colo320 colorectal adenocarcinoma cells. *J Gastrointestin Liver Dis.* 2013 Mar;22(1):37-43

### Jurnale BDI

1. **Roxana Cojocneanu Petric**, Laura-Ancuta Pop, Ancuta Jurj, Lajos Raduly, Dan Dumitrascu, Nicolae Dragos, Ioana Berindan Neagoe. Next Generation Sequencing Application for Breast Cancer Research. *Clujul Medical*, [S.l.], v. 88, n. 3, p. 278-287, jul. 2015. ISSN 2066-8872
2. C. Gherman, V. Pileczki, **R. Cojocneanu Petric**, S. Rapuntean, S. Gherman, I. Berindan Neagoe, Molecular mechanisms of action and prediction of response to oxaliplatin in colorectal cancer cells. *Annals of the Romanian Society for Cell Biology* 2012; 17(1):194-200.
3. Gherman C, Pileczki V, **Cojocneanu Petric R**, Braicu C, Răpuntean S, Berindan Neagoe I. In vitro studies for evaluation the antitumoral and immunomodulator effect of EGCG on Ehrlich Ascites. *Archiva Zootechnica* 15:2, 79-87, 2012

## Bibliografie selectivă

1. Casey G, Conti D, Haile R, Duggan D. Next generation sequencing and a new era of medicine. *Gut*. 2013;62(6):920-32.
2. Metzker ML. Sequencing technologies - the next generation. *Nature reviews Genetics*. 2010;11(1):31-46.
3. Shoemaker DD, Schadt EE, Armour CD, He YD, Garrett-Engel P, McDonagh PD, et al. Experimental annotation of the human genome using microarray technology. *Nature*. 2001;409(6822):922-7.
4. Trevino V, Falciani F, Barrera-Saldana HA. DNA microarrays: a powerful genomic tool for biomedical and clinical research. *Molecular medicine*. 2007;13(9-10):527-41.
5. Hsi-Yang Fritz M, Leinonen R, Cochrane G, Birney E. Efficient storage of high throughput DNA sequencing data using reference-based compression. *Genome research*. 2011;21(5):734-40.
6. Batley J, Edwards D. Genome sequence data: management, storage, and visualization. *BioTechniques*. 2009;46(5):333-4, 6.
7. Zhang J, Chiodini R, Badr A, Zhang G. The impact of next-generation sequencing on genomics. *Journal of genetics and genomics = Yi chuan xue bao*. 2011;38(3):95-109.
8. Luscombe NM, Greenbaum D, Gerstein M. What is bioinformatics? A proposed definition and overview of the field. *Methods of information in medicine*. 2001;40(4):346-58.
9. Bodrova TA, Kostyushev DS, Antonova EN, Slavin S, Gnatenko DA, Bocharova MO, et al. Introduction into PPPM as a new paradigm of public health service: an integrative view. *The EPMA journal*. 2012;3(1):16.
10. de Martel C, Ferlay J, Franceschi S, Vignat J, Bray F, Forman D, et al. Global burden of cancers attributable to infections in 2008: a review and synthetic analysis. *The Lancet Oncology*. 2012;13(6):607-15.
11. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International journal of cancer Journal international du cancer*. 2015;136(5):E359-86.
12. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *International journal of cancer Journal international du cancer*. 2010;127(12):2893-917.
13. Slavicek L, Pavlik T, Tomasek J, Bortlicek Z, Buchler T, Melichar B, et al. Efficacy and safety of bevacizumab in elderly patients with metastatic colorectal cancer: results from the Czech population-based registry. *BMC gastroenterology*. 2014;14:53.
14. Valastyan S, Weinberg RA. Tumor metastasis: molecular insights and evolving paradigms. *Cell*. 2011;147(2):275-92.
15. Guise T. Examining the metastatic niche: targeting the microenvironment. *Seminars in oncology*. 2010;37 Suppl 2:S2-14.

16. Studer RA, Dessailly BH, Orengo CA. Residue mutations and their impact on protein structure and function: detecting beneficial and pathogenic changes. *The Biochemical journal*. 2013;449(3):581-94.
17. Hochman J, Insel PA, Bourne HR, Coffino P, Tomkins GM. A structural gene mutation affecting the regulatory subunit of cyclic AMP-dependent protein kinase in mouse lymphoma cells. *Proceedings of the National Academy of Sciences of the United States of America*. 1975;72(12):5051-5.
18. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nature methods*. 2010;7(4):248-9.
19. Kumar S, Suleski MP, Markov GJ, Lawrence S, Marco A, Filipinski AJ. Positional conservation and amino acids shape the correct diagnosis and population frequencies of benign and damaging personal amino acid mutations. *Genome research*. 2009;19(9):1562-9.
20. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature protocols*. 2009;4(7):1073-81.
21. Tarca AL, Romero R, Draghici S. Analysis of microarray experiments of gene expression profiling. *American journal of obstetrics and gynecology*. 2006;195(2):373-88.
22. Slonim DK, Yanai I. Getting started in gene expression microarray analysis. *PLoS computational biology*. 2009;5(10):e1000543.
23. Altobelli G. Bioinformatics applied to gene transcription regulation. *Journal of molecular endocrinology*. 2012;49(2):R51-9.
24. Viceconti M, Hunter P, Hose R. Big Data, Big Knowledge: Big Data for Personalized Healthcare. *IEEE journal of biomedical and health informatics*. 2015;19(4):1209-15.
25. Idris SF, Ahmad SS, Scott MA, Vassiliou GS, Hadfield J. The role of high-throughput technologies in clinical cancer genomics. *Expert review of molecular diagnostics*. 2013;13(2):167-81.
26. Tran B, Dancey JE, Kamel-Reid S, McPherson JD, Bedard PL, Brown AM, et al. Cancer genomics: technology, discovery, and translation. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2012;30(6):647-60.
27. Hansen AR, Bedard PL. Clinical application of high-throughput genomic technologies for treatment selection in breast cancer. *Breast cancer research : BCR*. 2013;15(5):R97.
28. Hoelder S, Clarke PA, Workman P. Discovery of small molecule cancer drugs: successes, challenges and opportunities. *Molecular oncology*. 2012;6(2):155-76.
29. Ren A, Dong Y, Tsoi H, Yu J. Detection of miRNA as non-invasive biomarkers of colorectal cancer. *International journal of molecular sciences*. 2015;16(2):2810-23.
30. Sozzi G, Boeri M, Rossi M, Verri C, Suatoni P, Bravi F, et al. Clinical utility of a plasma-based miRNA signature classifier within computed tomography lung cancer screening: a correlative MILD trial study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2014;32(8):768-73.
31. Cojocneanu Petric R, Pop LA, Jurj A, Raduly L, Dumitrascu D, Dragos N, et al. Next Generation Sequencing Applications for Breast Cancer Research. *Clujul Medical*. 2015;88(3):278-87.



32. Cleere DW. Triple negative breast cancer: a clinical update. *Community Oncology*. 2010;7(5):8.
33. Lawrence RT, Perez EM, Hernandez D, Miller CP, Haas KM, Irie HY, et al. The proteomic landscape of triple-negative breast cancer. *Cell reports*. 2015;11(4):630-44.
34. Lehmann BD, PiTENpol JA. Identification and use of biomarkers in treatment strategies for triple-negative breast cancer subtypes. *The Journal of pathology*. 2014;232(2):142-50.
35. Boyle P. Triple-negative breast cancer: epidemiological considerations and recommendations. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*. 2012;23 Suppl 6:vi7-12.
36. Stratton MR. Exploring the genomes of cancer cells: progress and promise. *Science (New York, NY)*. 2011;331(6024):1553-8.
37. Buttitta F, Felicioni L, Del Grammastro M, Filice G, Di Lorito A, Malatesta S, et al. Effective assessment of egfr mutation status in bronchoalveolar lavage and pleural fluids by next-generation sequencing. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2013;19(3):691-8.
38. Tripathy D, Harnden K, Blackwell K, Robson M. Next generation sequencing and tumor mutation profiling: are we ready for routine use in the oncology clinic? *BMC medicine*. 2014;12:140.
39. Tattini L, D'Aurizio R, Magi A. Detection of Genomic Structural Variants from Next-Generation Sequencing Data. *Frontiers in bioengineering and biotechnology*. 2015;3:92.
40. Shen T, Pajaro-Van de Stadt SH, Yeat NC, Lin JC. Clinical applications of next generation sequencing in cancer: from panels, to exomes, to genomes. *Frontiers in genetics*. 2015;6:215.
41. Niedringhaus TP, Milanova D, Kerby MB, Snyder MP, Barron AE. Landscape of next-generation sequencing technologies. *Analytical chemistry*. 2011;83(12):4327-41.
42. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, et al. Performance comparison of benchtop high-throughput sequencing platforms. *Nature biotechnology*. 2012;30(5):434-9.
43. Stranneheim H, Lundeberg J. Stepping stones in DNA sequencing. *Biotechnology journal*. 2012;7(9):1063-73.
44. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research*. 2010;38(16):e164.
45. Balko JM, Giltane JM, Wang K, Schwarz LJ, Young CD, Cook RS, et al. Molecular profiling of the residual disease of triple-negative breast cancers after neoadjuvant chemotherapy identifies actionable therapeutic targets. *Cancer discovery*. 2014;4(2):232-45.
46. Zhong X, Yang H, Zhao S, Shyr Y, Li B. Network-based stratification analysis of 13 major cancer types using mutations in panels of cancer genes. *BMC genomics*. 2015;16 Suppl 7:S7.
47. Bai X, Zhang E, Ye H, Nandakumar V, Wang Z, Chen L, et al. PIK3CA and TP53 gene mutations in human breast cancer tumors frequently detected by ion torrent DNA sequencing. *PloS one*. 2014;9(6):e99306.

48. Liu S, Wang H, Zhang L, Tang C, Jones L, Ye H, et al. Rapid detection of genetic mutations in individual breast cancer patients by next-generation DNA sequencing. *Human genomics*. 2015;9:2.
49. Gauthier C, Griffin G. Using animals in research, testing and teaching. *Revue scientifique et technique*. 2005;24(2):735-45.
50. Festing S, Wilkinson R. The ethics of animal research. Talking Point on the use of animals in scientific research. *EMBO reports*. 2007;8(6):526-30.
51. Eliasof S, Lazarus D, Peters CG, Case RI, Cole RO, Hwang J, et al. Correlating preclinical animal studies and human clinical trials of a multifunctional, polymeric nanoparticle. *Proceedings of the National Academy of Sciences of the United States of America*. 2013;110(37):15127-32.
52. Frazer KA, Elnitski L, Church DM, Dubchak I, Hardison RC. Cross-species sequence comparisons: a review of methods and available resources. *Genome research*. 2003;13(1):1-12.
53. Yang CS, Wang X, Lu G, Picinich SC. Cancer prevention by tea: animal studies, molecular mechanisms and human relevance. *Nature reviews Cancer*. 2009;9(6):429-39.
54. Cui X, Vinar T, Brejova B, Shasha D, Li M. Homology search for genes. *Bioinformatics*. 2007;23(13):i97-103.
55. Caldwell RW, Tuite J, Stob M, Baldwin R. Zearalenone production by *Fusarium* species. *Applied microbiology*. 1970;20(1):31-4.
56. Hidy PH, Baldwin RS, Greasham RL, Keith CL, McMullen JR. Zearalenone and some derivatives: production and biological activities. *Advances in applied microbiology*. 1977;22:59-82.
57. Yazar S, Omurtag GZ. Fumonisin, trichothecenes and zearalenone in cereals. *International journal of molecular sciences*. 2008;9(11):2062-90.
58. Fazekas B, Tar A. Determination of zearalenone content in cereals and feedstuffs by immunoaffinity column coupled with liquid chromatography. *Journal of AOAC International*. 2001;84(5):1453-9.
59. Iqbal SZ, Rabbani T, Asi MR, Jinap S. Assessment of aflatoxins, ochratoxin A and zearalenone in breakfast cereals. *Food chemistry*. 2014;157:257-62.
60. Tenaillon O, Skurnik D, Picard B, Denamur E. The population genetics of commensal *Escherichia coli*. *Nature reviews Microbiology*. 2010;8(3):207-17.
61. Sekirov I, Russell SL, Antunes LC, Finlay BB. Gut microbiota in health and disease. *Physiological reviews*. 2010;90(3):859-904.
62. Huffnagle G, Noverr MC. GI microbiota and regulation of the immune system. Preface. *Advances in experimental medicine and biology*. 2008;635:v-vi.
63. Oswald IP, Desautels C, Laffitte J, Fournout S, Peres SY, Odin M, et al. Mycotoxin fumonisin B1 increases intestinal colonization by pathogenic *Escherichia coli* in pigs. *Applied and environmental microbiology*. 2003;69(10):5870-4.
64. Berretta R, Moscato P. Cancer biomarker discovery: the entropic hallmark. *PloS one*. 2010;5(8):e12262.
65. Rakoff-Nahoum S. Why cancer and inflammation? *The Yale journal of biology and medicine*. 2006;79(3-4):123-30.

66. Lu H, Ouyang W, Huang C. Inflammation, a key event in cancer development. *Molecular cancer research : MCR*. 2006;4(4):221-33.
67. Kiraly O, Gong G, Olipitz W, Muthupalani S, Engelward BP. Inflammation-induced cell proliferation potentiates DNA damage-induced mutations in vivo. *PLoS genetics*. 2015;11(2):e1004901.
68. West AC, Jenkins BJ. Inflammatory and Non-Inflammatory Roles for Toll-Like Receptors in Gastrointestinal Cancer. *Current pharmaceutical design*. 2015;21(21):2968-77.