

Tumor heterogenitás vizsgálata szolid tumorokban újgenerációs DNS szekvenálás segítségével

Doktori tézisek

Pongor Lőrinc Sándor

Semmelweis Egyetem
Patológia Tudományok Doktori Iskola



Témavezető: Dr. Győrffy Balázs, az MTA doktora, tudományos tanácsadó

Hivatalos bírálók: Dr. Kocsis István, Ph.D., egyetemi docens
Dr. Mészáros Bálint, Ph.D., tudományos munkatárs

Szigorlati bizottság elnöke: Dr. Veres Gábor, az MTA doktora, egyetemi tanár
Szigorlati bizottság tagjai: Dr. Alexa Anita, Ph.D., tudományos munkatárs
Dr. Újházy András, Ph.D., egyetemi adjunktus

Budapest
2018

1. Bevezetés

A daganatok növekedése során dinamikusan változik a genetikai összetétel, mely során genetikai változások halmozódnak fel. Ennek következménye, hogy a tumorok nem csak betegek közt mutatnak inter-tumor heterogenitást, hanem egyazon tumoron belül megjelenik az intra-tumor heterogenitás is. A genetikai változékonyság hatására megjelenhetnek olyan mutációk, melyek befolyásolhatják a betegek túlélését, valamint a kezeléssel szembeni rezisztenciát. Ismertebb példák a KRAS gén mutációinak kapcsolata a receptor tirozin kináz terápia rezisztenciával, vagy a KI67 gén (mint proliferációs marker) kifejeződés szintjének összefüggése a betegek túlélésével.

A klinikai diagnosztikában egyre gyakrabban alkalmaznak újgenerációs szekvenálást, mely nem csak a genetikai változások meglétéről, de a tumorban lévő arányukról is képes információt adni. Ezen vizsgálatok előnye, hogy képesek több száz gént, esetleg teljes genomot egy mérésben vizsgálni, így lehetőség van korábban nem ismert változások azonosítására is.

Az onkológiai vizsgálatok célja olyan genetikai változások azonosítása, amely prognosztikus lehet a beteg túlélésére, vagy prediktív a kezelésének hatékonyságára. Az utóbbi években számos példa utalt arra, hogy egy gén mutációja alapján nem lehetséges tökéletesen előre jelezni a választ bizonyos terápiákra vagy betegek túlélésére. Ilyen esetekben gyakran a vizsgált mutáció által érintett jelátviteli útvonal felelős a változásért, amely háttérében más genetikai mutációk is állhatnak. Például, receptor tirozin kináz terápiával szembeni rezisztenciát okozhatnak a BRAF, az NRAS, vagy akár a PIK3CA gének szomatikus mutációi, melyek a KRAS génhez hasonlóan a sejt növekedési jelpályán helyezkednek el. A PARP inhibitorok hatékonyak bizonyultak azon daganatok esetén, ahol nem megfelelő a DNS kettős-szál törés javítás homológ rekombináció által. A homológ rekombináció hiánya az ismert BRCA1/2 mutációkon kívül létrejöhet a PALB2, az ATM, vagy a CHEK1/2 génmutációk hatására is.

2. Célkitűzések

Célkitűzéseim a következők:

- Megvizsgálni a tumor heterogenitás és a tumorban kimutatható szomatikus mutációk kapcsolatát. Munkám során az alábbi részcélokat tűztem ki:
 1. Sejt mozgás hatásának vizsgálata az újgenerációs szekvenálással mérhető összetételre *in vitro* kísérletekkel.
 2. Újgenerációs szekvenálás pontosságának vizsgálata ismert összetételű sejtvonal keverékek felhasználásával.
 3. A szekvenált tumor méret és a detektált szomatikus mutációk száma közötti összefüggés vizsgálata petefészek tumor minták segítségével.

- A génmutáció és génkifejeződés közti kapcsolat vizsgálata, amelyhez a következő részcélokat tűztem ki:
 4. A génmutációk eredményeképpen létrejövő génkifejeződés változás hatásának vizsgálata az emlőtumoros betegek túlélésére.
 5. A TP53 génmutációra visszavezethető megváltozott génkifejeződési mintázat elemzése és értelmezése.

3. Módszerek

3.1. Genotype-2-Outcome elemzőrendszer adatainak előkészítése

A TCGA (The Cancer Genome Atlas) adatbázisból összesen 762 emlőtumoros és 555 nem-kissejtes tüdőrákos beteg adatait dolgoztam fel munkám során. Az exom szekvenálási adatokból a mutációk azonosítását a *Mutect* szoftverrel végeztem. A leszűrt mutációk annotálását az *SNPeff* programcsomag felhasználásával végeztem. A TCGA adatbázisban elérhető tumor minta RNS-szekvenálásból származó adatok normalizálását RSEM algoritmussal végeztem.

A microarray alapú génkifejeződési adatokat az EGA (European Genome-Phenome Archive) és GEO (Gene Expression Omnibus) adatbázisokból értük el. Az adatbázisban 5,934 beteg nyers Affymetrix CEL formátumú génkifejeződési fájlja volt elérhető, melyeket MAS5 normalizálással dolgoztunk fel a *Bioconductor Affy* könyvtárával, R (v3.02) környezetben.

3.2. Genotype-2-Outcome elemzőrendszer algoritmus

Az algoritmus három adatbázist használ fel: TCGA mutációs adatok, TCGA RNA-seq adatok, valamint független validációs adatbázisként egy microarray alapú génkifejeződési adatbázist. A módszer működése két fő lépésre bontható, bemenetként egy gén nevet vagy azonosítót kell adni, kimenetként a program Kaplan-Meier túlélés-elemzést végez.

Az elemzés első szakaszában a szoftver kikeresi, hogy mely gének kifejeződése változik meg a kiválasztott génmutáció állapotának függvényében a TCGA adatok felhasználásával. A mutációval rendelkező és vad típusú betegek közti génkifejeződés változást ROC (Receiver Operating Characteristic) elemzéssel végzi a szoftver.

A második szakaszban történik a túlélés-elemzés a független microarray adatbázis felhasználásával. Ez alatt kiszámolja betegenként a megváltozott kifejeződést mutató gének átlagos kifejeződési szintjét. A kapott metagén kifejeződési szintekből kikeresi a medián értéket, amely alapján a betegeket magas vagy alacsony kifejeződési csoportokra bontja. Végül a kettébontott csoportok felhasználásával készít túlélés-elemzést és Kaplan-Meier ábrákat.

3.3. Sejtvonalas inváziós és kalibrációs kísérletek

A sejt motilitás mérésekhez négy melanóma sejtvonalat választottunk ki (A375, MEL-JUSO, SK-MEL-28 és MEWO). Ezen sejtvonalak esetén 3 génben igazoltunk specifikus mutációkat Sanger szekvenálással. Újgenerációs szekvenáláshoz az A375 / MEWO, valamint az SK-MEL-28 / MEWO sejtvonal párosokat használtuk fel. Az A375 / MEL-JUSO sejtvonal párok esetén az inváziót fluoreszcens videó-mikroszkópiával követtük.

A gyűrűs inváziós kísérleteket FlexiPERM[®] conB sejt-kizáró szilikon gyűrű felhasználásával végeztük el. A kísérletek során az első sejtvonalat a tenyésztő csésze belső régiójára raktuk le, majd miután a sejtek letapadtak, a szilikon gyűrűt eltávolítottuk és hozzáadtuk a második sejtvonalat a teljes felületre. 72 órás inkubálást követően DNS izolálást végeztünk a belső gyűrű kerülete és a sejttenyésztő csésze külső kerülete mentén.

Megfelelő mutáció frekvencia összehasonlítás érdekében készítettünk kalibrációs keverékeket az A375 és a MEWO sejtvonalakkal. A kalibrációs sorban a két sejtvonal arányát 2%, 5%, 10%, 25% valamint 50%-ra állítottuk be, amelyben az A375 volt a minor sejtvonal.

Az inváziós és kalibrációs kísérletek során izolált DNS mintákat Ion 314 chip-en szekvenáltuk Ion PGM 200 eszközön 600x lefedettséggel. A szekvenáláshoz célzottan terveztünk amplikonokat 25 sejtvonal specifikus mutációra.

3.4. Petefészek tumor minták gyűjtése és előkészítése

A petefészek tumorok multi-régió szekvenálását öt betegből végeztük. Minden beteg esetén egy normál vérmintát, valamint a tumorból három részletet gyűjtöttünk. A minták az Országos Onkológiai Intézetben műtött betegekből származnak. A vér és tumor mintákból a „Qiagen DNeasy Blood and Tissue Kit” segítségével izoláltunk DNS-t a gyártó protokollja alapján (Qiagen GmbH, Hilden, Németország).

Minden betegből 3 mintát készítettünk elő: (1) biopszia minta, (2) lokális minta, ahol a DNS a biopszia környékéről származott, valamint egy (3) globális minta, ahol a beteg három tumor részletéből izoláltunk DNS-t. Ezeket a mintákat küldtük ki teljes exom szekvenálásra ~100x lefedettséggel.

3.5. Újgenerációs szekvenálási adatok bioinformatikai feldolgoása

A szekvenálási leolvasások minőségi ellenőrzését a *FastQC* szoftverrel ellenőriztem, a leolvasások nyírását a *trimmomatic* szoftverrel végeztem el. A leolvasásokat a *BWA MEM* programmal illeszttem a humán referencia genomra. Az illesztett leolvasásokat a *samtools* szoftvercsomaggal rendeztem és formáztam a tömörített BAM formátumra.

Az illesztett és rendezett BAM formátumú leolvasásokban a duplikátumokat a *picard-tools* szoftver segítségével megjelöltem. Az illesztett leolvasásokban inszerciók és deléciók környékén újra-illesztést végeztem a *GATK realignertargetcreator* és *indelrealigner* segítségével.

Sejtvonalas kísérletek adatainak esetén a mutációkat a *samtools mpileup* szoftverrel kerestem. Az *mpileup* kimenetéből kiválasztottam azon részeket, melyek a mutációkat lefedik. Ezen régiókból ki lehetett számolni, hogy hány leolvasás volt referencia bázis szerinti, vagy hány leolvasás tartalmazta a sejtvonal specifikus mutációkat.

A petefészek minták feldolgoása esetén a szomatikus mutációk azonosítását a *GATK mutect2* szoftverével végeztem. Mivel egy beteghez több minta tartozott, egy olyan mutáció kereső szoftvercsomagot fejlesztettem, amely együttesen dolgozza fel a mintákat a *mutect2* által azonosított mutációk alapján. Az öröklött mutációkat a *GATK HaplotypeCaller* segítségével azonosítottam. A mutációk génekhez való kapcsolását az *SNPeff* programmal végeztem. Az öröklött mutációkat továbbá annotáltam a dbSNP és ClinVar adatbázisokkal. A kópiaszám változás elemzést, valamint a mintákban lévő tumorsejt arány számítását a *sequenza* programmal végeztem.

4. Eredmények

4.1. Sejtmozgás hatásának vizsgálata újgenerációs szekvenálásra

Újgenerációs szekvenálás segítségével sikerült detektálni a mozgásból származó sejtösszetétel változást az *in vitro* inváziós kísérletekben. Az invázió mértéke 18,6% volt az A375 és a MEWO sejtvonala párok esetén, míg az SK-MEL-28 és a MEWO párok esetén csak 8,6% volt. Bár az SK-MEL-28 és az A375 sejtvonala mozgása lényegesen nem tért el egymástól monokultúras körülmények között, az általuk elért invázió mértéke szignifikánsan eltért ($p=0,011$), amikor a homozigóta mutációk frekvenciáját hasonlítottuk össze. Ezzel szemben a heterozigóta mutációkkal végzett elemzés esetén nem volt szignifikáns a különbség ($p=0,39$).

A kalibrációs szekvenálási kísérletekben sejtvonala párok ismert összetételű keverékeit szekvenáltuk. A mutációk frekvenciájának szórása egyre magasabb volt, ahogy a két sejtvonala aránya kiegyenlítődött. A legmagasabb szórás értéket két sejtvonala 50-50%-os aránya esetén kaptam. Heterozigóta mutációk frekvenciájának szórásai lényegesen nagyobbak voltak, amely alacsony szekvenálási lefedettség esetén tovább növekedett.

A mutációk frekvenciájának szórása akár a 17%-ot is elérhette egy kísérleten belül mind a kalibrációs, mind az inváziós szekvenálások esetén is. Ezzel szemben, amikor a mutációkat párosítottam a technikai ismétléseknél, ott a szórás csak 5,6% volt. A szimulációk alapján számított szórás alacsony (<50x) lefedettség esetén elérhette a 12,9%-ot. Nagyon magas, 1000x lefedettség esetén ez lényegesen kisebb volt, ebben az esetben 2% volt a legmagasabb érték.

4.2. A tumor mérete és a detektált szomatikus mutációk száma közötti összefüggés vizsgálata

A vizsgált betegek közül három esetben sikerült azonosítani olyan heterozigóta öröklött mutációkat, melyek kulcsszerepűek a DNS kettős-szál törés javításban. Ezen betegeknél a tumorban törődött a vad típusú allél, emiatt a mutáció homozigóta státuszú lett. A szomatikus mutációikban azonosítható volt a BRCA hiányhoz asszociált mintázat („Signature 3”). Esetükben a kópiaszám változás elemzés során sok kromoszóma-

részletet érintő amplifikációt és deléciót azonosítottam. Mindhárom tumorban jelen volt TP53 gént érintő szomatikus mutáció.

A negyedik betegnél csak az első kromoszómában azonosítottam egy darab amplifikációt. A szomatikus mutációkban az APOBEC citidin-dezamináz hibával járó mutációs mintázatokat határoztam meg, amelynél emelkedett a C>T és C>G bázis változások aránya. Négy klonális szomatikus mutáció a PIK3 növekedési jelpályát érintő génekben helyezkedett el (PTEN tumorszupresszorban inaktíváló mutáció, az FGFR2 receptor tirozin kinázban egy szubsztitúció és a PIK31 génben két heterozigóta inaktíváló mutáció).

Az ötödik beteg esetén a tumorban több teljes kromoszómát érintő kópiaszám változás volt megfigyelhető. Ezen tumor hipermutáns jellegű volt, a többi tumorhoz képest 3x több szomatikus mutáció volt található a biopszia mintában. A tumorban kimutattam a KRAS gén leggyakoribb p.G12D aminosav cserével járó aktiváló mutációját, valamint a PIK3CA gén leggyakoribb p.E454K aminosav cserével járó mutációját.

A négy nem hipermutáns tumor egyes mintavételi régióiban a szomatikus mutációk 69,3%-93,8%-ban egyeztek meg. A tumorokban azonosított mutációk száma átlagosan 149 volt ($118 < x < 200$ intervallumban).

A hipermutáns tumor biopszia mintájában 688 mutációt találtam, melyek 15,5%-a volt megtalálható mind a lokális, mind a globális mintákban. Az arány magasabb volt a lokális minta mutációi esetén, ahol a mutációk 25%-a azonosítható volt a többi régióban, viszont a kimutatható mutációk száma csupán 392 volt. A globális minta szekvenálása esetén már csak 140 mutáció volt detektálható, amely 71,4%-ban megegyezett a biopszia és lokális mintákkal.

4.3. Génmutáció és génkifejeződés kapcsolata

A Genotype-2-Outcome (G-2-O) rendszer segítségével mutáció és génkifejeződés adatok együttes felhasználásával végeztem túlélés-elemzést. Az elemzőrendszerrel végzett túlélés-elemzés során nem közvetlenül a mutációk hatását vizsgáltam, hanem a mutációk által előidézett génkifejeződési változások hatását. A GATA3 gén példáján látható volt, hogy a G-2-O által végzett túlélés-elemzés (HR=1,66; $p < E-16$) lényegesen szignifikánsabb, mint egyszerűen mutációval ($p=1,3E-08$) vagy kifejeződési szinttel (HR=0,71; $p=1,3E-08$) végzett túlélés-elemzés emlőtumoros betegeken. A MAP3K1 gén

magas kifejeződése rossz prognózissal járt ($HR=1,8$; $p<E-16$) és a gén mutációjához kötött génkifejeződés mintázat szintén rosszabb prognózissal társult ($HR=1,6$; $p<E-16$).

Tüdődaganatok esetén nagyon gyakori a KRAS onkogén mutációja, amely egy fontos negatív prediktív biomarker EGFR-t célzó kezelésekre. Túlélés-elemzésekből látható volt, hogy sem a génmutáció állapota, sem kifejeződési szintje nem hozható összefüggésbe a betegek prognózisával. A génmutációjához kötött génkifejeződési mintázat viszont rossz prognózissal társult ($HR=2,4$; $p=1,24E-12$).

A TP53 génmutációja esetén 23 olyan gén kifejeződési szintje emelkedett meg, mely a sejtciklushoz kötődik emlőtumorok ($p=1,3E-16$) és nem-kissejtes tüdődaganatok ($p=1,1E-23$) esetén is. Ezen gének magas kifejeződési szintje szinte mindegyik esetben rosszabb prognózissal járt a független géncsip adatbázisban. Az első 10 legszignifikánsabb p-értékkel rendelkező gén kombinációjából készített mintázat magas kifejeződése lényegesen rosszabb prognózissal járt ($HR=2,43$, $p<E-16$), mint az egyes gének külön ($HR<2.1$, $p<E-16$). Érdekes megfigyelés volt, hogy a kombinált mintázat hasonlóan rossz prognózissal társult csak TP53 gén vad típusú tumorok esetén ($HR=2,36$, $p=1,8E-3$), viszont jobb prognózist eredményezett a TP53 génmutáns tumoroknál ($HR=0,52$, $p=3,7E-2$).

5. Következtetések

Az utóbbi években több vizsgálat bemutatta, hogy magas a tumorok inter- és intra-tumor heterogenitása. Ez a genetikai diverzitás befolyásolja a klinikai szekvenálások eredményét, valamint nehezíti a prognosztikus és prediktív biomarkerek azonosítását.

Az eredményeim alapján levonható következtetések:

1. Újgenerációs szekvenálás felhasználásával követhető a mozgásból származó sejtösszetétel-változás *in vitro* inváziós kísérletekben. Ismert összetételű sejtvonalak szekvenálása esetén pontosabban lehetett megállapítani az arányokat azon esetekben, amikor egy sejtvonal van túlsúlyban.
2. Mind az inváziós, mind a kalibrációs kísérletek során a mutáció frekvenciák magas szórással rendelkeztek a biológiai ismétlések és a technikai ismétlések között. A magas szórások jelentősen befolyásolhatják az elemzésekből levont következtetéseket és a további elemzéseket. *In silico* modellezéssel bemutattam, hogy a szórás lényegesen csökken magas lefedettség esetén, így célszerű lehet magas lefedettségű szekvenálást végezni a klinikai diagnosztikai mérések esetén.
3. Nagyobb tumorrészt lefedő szekvenálás lényegesen nem befolyásolja a detektálható mutációk számát. A jelenlegi klinikai diagnosztikai szekvenálások mintavételi gyakorlata eredményeim alapján is helyesnek tűnik, de törekedni kell a minél nagyobb szekvenálási lefedettségre.
4. Szomatikus mutációk hatására gyakran nem csak egy fehérje, hanem az általa szabályozott jelátviteli útvonal és a sejtben lévő RNS szintek is megváltoznak. A Genotype-2-Outcome elemzőrendszer felhasználásával bemutattam, hogy onkogén vagy tumorszupresszor gén mutációjához kötött génkifejeződés-változások prognosztikusak a betegek túlélésére. A mutációhoz kötött génkifejeződési változásokkal végzett túlélés-elemzés sok esetben jobb eredményt produkált az adott gén mutációja, vagy kifejeződési szintje alapján elvégzett elemzéshez képest. Ezért ez a módszer új biomarker azonosításánál jól kiegészítheti a csak génmutáción vagy génkifejeződésen alapuló módszereket.
5. A TP53 génmutációval rendelkező emlőtumoros betegekben több olyan gén kifejeződési szintje emelkedett meg, melyek a sejtosztódást szabályozzák. Ezen génekről több közleményben leírták, hogy a magas kifejeződésük rossz prognózissal társul. Munkámban kimutattam, hogy a talált gének magas kifejeződése valóban rossz prognózissal jár a TP53 gén vad típusú betegekben, viszont jobb a prognózis TP53 génmutáns betegek esetén. A megfigyelésekből

arra következtethetünk, hogy ezen gének kifejeződése és a betegek prognózisának kapcsolata inkább indirekt módon tükrözi a TP53 gén mutációs állapotát.

6. Saját publikációk jegyzéke

6.1. Disszertációhoz kapcsolódó publikációk jegyzéke

1. **Pongor LS**, Kormos M, Hatzis C, Pusztai L, Szabo A, Gyorffy B. (2015) A genome-wide approach to link genotype to clinical outcome by utilizing next generation sequencing and gene chip data of 6,697 breast cancer patients. *Genome medicine*, 7: 104. **IF=5,846**
2. Jiang T, Shi W, Wali VB, **Pongor LS**, Li C, Lau R, Gyorffy B, Lifton RP, Symmans WF, Pusztai L, Hatzis C. (2016) Predictors of Chemosensitivity in Triple Negative Breast Cancer: An Integrated Genomic Analysis. *PLoS medicine*, 13: e1002193. **IF= 11,862**
3. Harami-Papp H*, **Pongor LS***, Munkacsy G, Horvath G, Nagy AM, Ambrus A, Hauser P, Szabo A, Tretter L, Gyorffy B. (2016) TP53 mutation hits energy metabolism and increases glycolysis in breast cancer. *Oncotarget*, 7: 67183-67195. **IF= 5,168**
4. Nagy A, **Pongor LS**, Szabo A, Santarpia M, Gyorffy B. (2017) KRAS driven expression signature has prognostic power superior to mutation status in non-small cell lung cancer. *International journal of cancer*, 140: 930-937. **IF=6,513**
5. **Pongor LS**, Harami-Papp H, Mehes E, Czirok A, Gyorffy B. (2017) Cell Dispersal Influences Tumor Heterogeneity and Introduces a Bias in NGS Data Interpretation. *Scientific reports*, 7: 7358. **IF= 4,259**

*Megosztott első szerző

6.2. Disszertációtól független publikációk jegyzéke

1. **Pongor LS**, Pinter F, Petak I. (2013) HeurAA: accurate and fast detection of genetic variations with a novel heuristic amplicon aligner program for next generation sequencing. *PloS one*, 8: e54294. **IF= 3.534**

2. **Pongor LS**, Vera R, Ligeti B. (2014) Fast and sensitive alignment of microbial whole genome sequencing reads to large sequence datasets on a desktop PC: application to metagenomic datasets and pathogen identification. PloS one, 9: e103441. **IF= 3.234**

3. Cai W, Xiong Chen Z, Rane G, Satendra Singh S, Choo Z, Wang C, Yuan Y, Zea Tan T, Arfuso F, Yap CT, **Pongor LS**, Yang H, Lee MB, Cher Goh B, Sethi G, Benoukraf T, Tergaonkar V, Prem Kumar A. (2017) Wanted DEAD/H or Alive: Helicases Winding Up in Cancers. Journal of the National Cancer Institute, 109. **IF= 12.589**