



Európa Kollégium Egyetemista Központ, Újvidék

Pletikoszity Gábor

A DSPP-GÉN PROTEINKÓDOLÓ  
SZEKVENCIÁJÁBAN FELLÉPŐ  
PONTMUTÁCIÓK SZÁMÍTÓGÉPES  
ELEMZÉSE

ORVOSTUDOMÁNYI ÉS EGÉSZSÉGVÉDELMI MŰHELY

EGYETEMISTA:

Pletikoszity Gábor

Újvidéki Egyetem

Orvostudományi Kar, Fogorvosi szak

IV. évfolyam

MENTOR:

Dr. Stojišić-Milosavljević Anastazija

## ELŐSZÓ

Legyen szó bármilyen biológiai jelenségről, ismertetőjegyről, tulajdonságról, nincs olyan, amelyet egy kicsit ne határozná meg a genetika. Manapság már közismert a gének szerepe, erre az ismeretre pedig teljesen új tudományágak épültek a molekuláris genetikától kezdve, az epigenetikán át egészen a modern orvosi genetikáig. Segítségével közvetlen bizonyítékot találhatunk számos betegség kialakulására, öröklődésére, és szinte látnoki képességgel meg tudjuk állapítani annak lefolyását, előre láthatjuk továbböröklődését. Ma, a modern genetika korában természetes ennek a tudásnak a jelenléte, azonban ha csak néhány évtizedre visszatekintünk, láthatjuk, mekkorát változott az ismeretkör ezen a területen.

A legelső látványos lépés a genetika történetében a mendeli genetika létrejötte. Habár kezdetben nem vették komolyan, számos, máig érvényes alapelvet felállított, például az öröklődés folyamatának magyarázatát. 1953-ban Watson és Crick DNS-modellje számos addigi problémát megmagyarázott. Ekkor vált ismertté a DNS kettős helix szerkezete, a replikáció, valamint a DNS és RNS közötti kapcsolat. Az 1990 és 2006 között zajló Human Genom Project segítségével szinte az egész génkészletet sikerült felfedezni, számos rendelkezést sikerült megmagyarázni, és innentől kezdve a genetika már egy új szakaszba lépett, a mai értelemben vett modern genetika korszakába.

A kutatás keretein belül a Human Genom Project eredményeként létrejött DSPP-gén<sup>1</sup> standardizált génváltozata lett elemezve. Ez számos génadatbázisban elérhető, például az általam is választott Ensembl adatbázisban. Az ismert bázissorrend és elsősorban kombinatorikai műveletek segítségével könnyen meg lehet határozni bizonyos mutációk gyakoriságát és helyét is. Ez a munka nem szól másról, mint a gén pontmutációinak ilyen szemszögű vizsgálatáról.

## BEVEZETÉS

A DSPP-gén az emberi kromoszómakészlet 4. kromoszómájának q karján található gén. Megközelítőleg 8.35kb<sup>2</sup> hosszúságú és 5 exonból áll, amelyek közül az utolsó négy közvetlenül a protein kódolásában vesz részt. A proteinkódoló szekvenciáját összesen 3906 bázispár alkotja, amelyről az 1301 aminosavból álló dentin sialophosphoprotein íródik át. Ezt a dentint képző odontoblaszt sejtek választják ki, majd poszttranszlációs modifikáció során két kisebb egységre

<sup>1</sup> Dentin sialophosphoprotein gén.

<sup>2</sup> Kilobázis.

esik szét, a dentin sialoproteinre és a dentin phosphoproteinre. Mindkét fehérje a dentin mineralizációjában játszik fontos szerepet.

A gén mutációja során a fogakat érintő egyik leggyakoribb és legsúlyosabb kór, az úgynevezett dentinogenesis imperfecta lép fel. Ez egy autoszomális dominánsan öröklődő betegség, amelyek következtében a fog normális nagyságú, ám a dentin erősen hipomineralizált, szerkezete szabálytalan; a dentincsatornák tágak, számuk csökkent és gyakran egymásba csavarodnak. A dentin a szokástól lazább szerkezetű és könnyebben kopik. A fog szürkés vagy barnás elszíneződése is megfigyelhető. Mutációjával emellett még számos dentin dysplasiás rendellenesség is felléphet.

## A KUTATÁS PROBLÉMAFELVETÉSE ÉS CÉLKITŰZÉSE

A DSPP-gén mutációjának lehetséges következményeit elemezve nyilvánvalóvá válik a gén fontossága és jelentősége nemcsak a laboratóriumi, hanem klinikai kutatásban is. A kutatás számítógépes elemzésen alapszik. Természetesen ez nem helyettesítheti a klinikán, illetve a laboratóriumban végzett kutatásokat, viszont néhány, elsősorban matematikán alapuló jelenségre fényt deríthet, és egy jó útírányt adhat egy esetleges, későbbi klinikai vagy laboratóriumi megfigyeléshez. Az elérhető forrásokat figyelembe véve a következő célok fogalmazódtak meg:

- kimutatni, hogy az adott gén hányféleképpen mutálódhat;
- az adott mutációk közül hány nem okozza a képződő protein módosulását;
- a gént alkotó nitrogénalapú bázisok közül melyiket milyen gyakran érintheti mutáció;
- a különféle aminosavakat milyen gyakorisággal érinti mutáció;
- egymáshoz képest milyen gyakoriak az aminosavak szinonim mutációi;
- mekkora eséllyel jön létre nonsense mutáció abban az esetben, ha csak egy darab bázis mutálódik, és ezek a mutációk hol találhatóak meg a szekvencián.


## A KUTATÁS HIPOTÉZISE

A génszekvencia összetételéből már következtetni lehet, hogy az esetlegesen fellépő pontmutációk elsősorban azokat a nitrogénalapú bázisokat érinti, amelyeknek az előfordulási gyakorisága a legnagyobb; azokat az aminosavakat

érinti leggyakrabban mutáció, amelyeknek megfelelő kodonjai<sup>3</sup> a leggyakoribbak, szinonim változatainak száma pedig a gyakorisága mellett a legtöbb szinonim kodonnal rendelkező kodon jelentkezik. Az összes szinonim mutáció a hipotézis szerint elenyésző az összes mutációhoz képest, kevesebb, mint 0.1%-ot képvisel, ugyanis nincs olyan aminosav, amely több mint 6 szinonim kodonnal rendelkezik a lehetséges 40 féle variációból, valamint léteznek olyan aminosavak is (metionin és triptofán), amelyeket csak egyféle kodonnal kódol. Az egy darab pontmutációval fellépő nonsense mutációk száma pedig nem haladja meg a 200-at, tekintettel arra, hogy a szekvencia csak egy start és stopkodont tartalmazhat, és mindössze csak 3 kodon kódol stopkodont, így viszonylag kevés esélyt adva más, missense mutációhoz képest. A szekvencia hosszát tekintve a lehetséges összes mutáció számának nagyságrendje nagyobb a  $10^{1000}$  értékénél.

## A KUTATÁS MÓDSZERE

A kutatás matematikai jellegét tekintve teljes egészében kivitelezhető informatikai környezetben. Az erre a célra alkalmas, kész szoftver hiányában saját fejlesztésűt alkalmaztam, ami PyCharm fejlesztőkörnyezetben készült.<sup>4</sup> A program kódja összesen 591 sorból áll, amely a beviteli mezőbe bevitt génszakasz bázissorrendje alapján logikai, kombinatorikai műveleteket és százalékszámítást végez.



*1. ábra: A program beviteli mezője*

A program elsősorban ellenőrzi, hogy a bevitt elem megfelel-e a gén szokásos összetételének, tehát hogy a szakasz adenin-, guanin-, timin- és citozinbázisokból áll-e. Abban az esetben, ha a program a felsorolt bázisokon kívül mást is talál, rögtön kiírja a probléma forrását<sup>5</sup> és működése leáll.

<sup>3</sup> A DNS- vagy RNS-szekvencián található bázishármas, a program tripletként hivatkozik rájuk.

<sup>4</sup> A program a következő linkről tölthető le: <https://docs.google.com/uc?export=download&id=13NUAshbiGNAcYoOuLaycdXgGJeVoFPPc>

<sup>5</sup> Ebben az esetben „A bevitt elem nem génszakasz (Nem áll nukleotidokból)” üzenetet írja ki.

```

gen = input("Génszekvencia: ")
gen = gen.upper()
for item in gen:
    if item != "A" and item != "G" and item != "T" and item != "C":
        print("A bevitt elem nem génszakasz (Nem áll nukleotidokból)")
        kilep = input("A kilépéshez nyomj ENTER-t!")
        sys.exit()

```

2. ábra: A beviteli szakasz és az összetételt ellenőrző szakaszok

Ezek után leméri a szakasz hosszát, megszámolja a benne lévő bázisok számát, majd csoportosítja őket és százalékos értékben kimutatja előfordulásukat.

```

hossz = len(gen)
adenin = 0
guanin = 0
timin = 0
citozin = 0
for item in gen:
    if item == "A":
        adenin = adenin + 1
    elif item == "T":
        timin += 1
    elif item == "G":
        guanin += 1
    elif item == "C":
        citozin += 1

szazalek_adenin = round(adenin / hossz * 100, 2)
szazalek_timin = round(timin / hossz * 100, 2)
szazalek_guanin = round(guanin / hossz * 100, 2)
szazalek_citozin = round(citozin / hossz * 100, 2)

```

3. ábra: A szakasz hosszát és összetételét elemző szakaszok

A továbbiakban kiszámolja a lehetséges összes mutáció számát, és tízes alapú hatványá alakítja, majd négy tizedes pontossággal mutatja ki őket.

```

lehetseges_muatciok = decimal.Decimal(pow(5, hossz) - 1)
lehetseges_muatciok_tizedes = lehetseges_muatciok
kitevo = 0
while lehetseges_muatciok_tizedes >= 10:
    lehetseges_muatciok_tizedes = lehetseges_muatciok_tizedes / 10
    kitevo = kitevo + 1

lehetseges_muatciok_tizedes = round(lehetseges_muatciok_tizedes, 4)

```

4. ábra: A lehetséges mutációk számát kiszámító szakasz és az ezeket tízes hatványalakba alakító szakaszok

Ezt követően a kodonokat ellenőrzi, és csoportosítja a neki megfelelő aminosavakkal, emellett az aminosavnak sorszámot ad. Ha a kodon csak egy bázisban

különbözik valamelyik stopkodontól,<sup>6</sup> a sorszámát elmenti, valamint annak a bázisnak a sorrendjét is, ahol a mutáció nonsense mutációt hozhat létre.

```
triplet_első = 0
triplet_masodik = 1
triplet_harmadik = 2
...
while i < hossz:
    i = i + 3
    x = gen[triplet_első]
    triplet_első = triplet_első + 3
    y = gen[triplet_masodik]
    triplet_masodik = triplet_masodik + 3
    z = gen[triplet_harmadik]
    triplet_harmadik = triplet_harmadik + 3
```

5. ábra: Aminosavakat sorszámozó szakaszok

Az aminosavak ellenőrzése úgy történik, hogy az x, y és z változó helyére a kodonban található első, második és harmadik bázis jelét (A, T, G és C) veszi fel. Első lépésben az x értékét hasonlítja bázisokhoz, ha megtalálta a megfelelőt, áttér az y-ra. Amennyiben itt is megtalálta, a neki megfelelőt, a z értéket kezdi hasonlítani. Abban az esetben, ha minden x, y és z változóra megtalálta a neki megfelelő A, T, G és C értéket, a program felismeri a neki megfelelő kodont, és így már kimutatható a szinonim kodonok száma, az egy mutációval járó nonsense mutációk száma, a sorszáma. Ezután a ciklus újraindul és a következő kodont figyeli.

```
if x == "G":
    if y == "G":
        if z == "G":
            gly = gly + 1
            polimorf = polimorf * 4
        elif z == "A":
            gly = gly + 1
            polimorf = polimorf * 4
            stop_szerez = stop_szerez + 1
            aminosav_index.append(round(i / 3))
            gly_nonsense = gly_nonsense + 1
            g_nonsense = g_nonsense + 1
            bázis_index.append(i - 2)
```

6. ábra: Példa glicint kimutató sorokra

Abban az esetben, ha a szakasz nem startkodonnal<sup>7</sup> kezdődik, a program kiírja a hiba okát és leáll.

<sup>6</sup> A stopkodonok a TAG, TAA és TGA.

<sup>7</sup> Csak ATG lehet startkodon.

```

if i == 3:
    if x != "A" or y != "T" or z != "G":
        print("A bevitt elem nem start kodonnal kezdődik")
        kilep = input("A kilépéshez nyomj ENTER-t!")
        sys.exit()

```

7. ábra: A startkodont ellenőrző szakaszok

Ugyanezt elvégzi a stopkodonra is.

```

if i == hossz:
    if x != "T" or y != "A" or z != "A":
        if x != "T" or y != "A" or z != "G":
            if x != "T" or y != "G" or z != "A":
                print("A bevitt szakasz nem teljes (Stop kodon hiányzik)")
                kilep = input("A kilépéshez nyomj ENTER-t!")
                sys.exit()

```

8. ábra: A stopkodont ellenőrző szakaszok

Ha a bevitt szakasz nem kodonokból áll, a legutolsó lépésnél a program nem tudja felállítani a kodont,<sup>8</sup> kiírja a hiba okát és leáll.

```

try:
    ...
except IndexError:
    print("A bevitt elem nem kódol fehérjét (Nem áll tripletekből)")
    kilep = input("A kilépéshez nyomj ENTER-t!")

```

9. ábra: A hibakódot megjelenítő szakasz

Ezek után a szinonim mutációkat számolja ki, a számát tízes alapú hatvány-nyá alakítja, és négy tizedes pontossággal elmenti. Ugyanígy százalékosan is kimutatja kéttizedes pontossággal az összes mutációhoz képest. Abban az esetben, ha a százalékos érték kisebb, mint 0,1, helyére „<0.1%” kerül.

```

polimorf = decimal.Decimal(polimorf - 1)
polimorf_tizedes = polimorf
polimorf_kitevo = 0
while polimorf_tizedes >= 10:
    polimorf_tizedes = polimorf_tizedes / 10
    polimorf_kitevo = polimorf_kitevo + 1

polimorf_tizedes = round(polimorf_tizedes, 4)
polimorf_szazalek = polimorf / lethetseges_muatciok * 100
if polimorf_szazalek < 0.01 and polimorf_szazalek != 0:
    polimorf_szazalek = "<0.01"
else:
    polimorf_szazalek = round(polimorf_szazalek, 2)

```

10. ábra: A szinonim mutációkat kiszámító és kiírató szakaszok

<sup>8</sup> Abban az esetben, ha csak egy vagy két bázis marad a szekvencia végéig.

Ezt követően kiszámítja az összes aminosav számát, és az előzőhöz hasonlóan kiszámítja a nonsense mutációk előfordulási gyakoriságát. Az eredményeket kimutatja tizedes és százalékos formában is.

```

osszes_aminosav = hossz / 3

stop_szerez_szazalek = stop_szerez / osszes_aminosav * 100
if stop_szerez_szazalek < 0.01 and stop_szerez_szazalek != 0:
    stop_szerez_szazalek = "<0.01"
else:
    stop_szerez_szazalek = round(stop_szerez_szazalek, 2)

nonsense_szazalek = (stop_szerez + stop-veszit + start_kodon) /
osszes_aminosav * 100
if nonsense_szazalek < 0.01 and nonsense_szazalek != 0:
    nonsense_szazalek = "<0.01"
else:
    nonsense_szazalek = round(nonsense_szazalek, 2)

nonsense_bazis = t_nonsense + g_nonsense + c_nonsense + a_nonsense
nonsense_bazis_szazalek = nonsense_bazis / hossz
if nonsense_bazis_szazalek < 0.01 and nonsense_bazis_szazalek != 0:
    nonsense_bazis_szazalek = "<0.01"
else:
    nonsense_bazis_szazalek = round(nonsense_bazis_szazalek, 2)

```

11. ábra: Az aminosavak számát és nonsense mutációkat kimutató szakaszok

A program működése során szerzett eredmények – többek között a bázisok és aminosavak száma és százalékos összetétele, az összes mutáció száma, a szinonim mutációk száma tizedes és százalékos formában, a nonsense mutációk összetétele, részaránya százalékokban és ezek helyei a szekvencián – a kimeneti ablakban jelennek meg a felhasználó számára érthető formában.

```

legvalószínűbb nonsense mutációk száma (egy darab mutáció a tripletben elég a kialakításához):
(nonsense_bazis)      Az összes triplet (nonsense_bazis_szazalek)%-a
tripletek száma, ahol legvalószínűbben jelentkezik nonsense mutáció: {stop_szerez +
stop-veszit + start_kodon}      Az összes triplet (nonsense_szazalek)%-a
Ezen tripletek helyei a szekvencián: {aminosav_index}

Ezen pontmutációk helyei a szekvencián: {bazis_index}

```

12. ábra: A legvalószínűbb nonsense mutációk számát és helyeit kiírató sorok

```

Génszekvencia:

Bázisok lehetséges mutációi:

Timin: Százalékban:
Adenin: Százalékban:
Citozín: Százalékban:
Guanin: Százalékban:

Mutációk tripletenként:

Fenilalanin: Százalékban: Szinonim mutációk száma: Százalékban:
Leucin: Százalékban: Szinonim mutációk száma: Százalékban:
Szerin: Százalékban: Szinonim mutációk száma: Százalékban:
Tirozín: Százalékban: Szinonim mutációk száma: Százalékban:
Cisztein: Százalékban: Szinonim mutációk száma: Százalékban:
Triptofán: Százalékban: Szinonim mutációk száma: Százalékban:
Prolin: Százalékban: Szinonim mutációk száma: Százalékban:
Hisztidin: Százalékban: Szinonim mutációk száma: Százalékban:
Glutamin: Százalékban: Szinonim mutációk száma: Százalékban:
Arginin: Százalékban: Szinonim mutációk száma: Százalékban:
Izoleucin: Százalékban: Szinonim mutációk száma: Százalékban:
Metionin: Százalékban: Szinonim mutációk száma: Százalékban:
Treonin: Százalékban: Szinonim mutációk száma: Százalékban:
Aszparagin: Százalékban: Szinonim mutációk száma: Százalékban:
Lizin: Százalékban: Szinonim mutációk száma: Százalékban:
Valin: Százalékban: Szinonim mutációk száma: Százalékban:
Alanin: Százalékban: Szinonim mutációk száma: Százalékban:
Glutaminsav: Százalékban: Szinonim mutációk száma: Százalékban:
Glicin: Százalékban: Szinonim mutációk száma: Százalékban:
Stop kodon: Százalékban: Szinonim mutációk száma: Százalékban:
Összes triplet: Százalékban:

Lehetséges összes mutáció:
Polimorf mutációk: Százalékban:

Tripletek száma, ahol a mutáció a fehérjeszintézist idő előtt megállítja: Az összes triplet %-a
Tripletek száma, ahol a mutáció megakadályozza a fehérjeszintézis normális leállítását:
Tripletek száma, ahol a mutáció megakadályozza a fehérjeszintézis normális elkezdődését:

Legvalószínűbb nonsense mutációk száma (egy darab mutáció a tripletben elég a kialakításához): Az összes triplet %-a
Tripletek száma, ahol legvalószínűbben jelentkezik nonsense mutáció: Az összes triplet %-a
Ezen tripletek helyei a szekvencián:

Ezen pontmutációk helyei a szekvencián:

A nonsense mutációk a következő aminosavakat kódoló tripletteket érinti:

Fenilalanin: Százalékban: Az összes triplet %-a
Leucin: Százalékban: Az összes triplet %-a
Szerin: Százalékban: Az összes triplet %-a
Tirozín: Százalékban: Az összes triplet %-a
Cisztein: Százalékban: Az összes triplet %-a
Triptofán: Százalékban: Az összes triplet %-a
Prolin: Százalékban: Az összes triplet %-a
Hisztidin: Százalékban: Az összes triplet %-a
Glutamin: Százalékban: Az összes triplet %-a
Arginin: Százalékban: Az összes triplet %-a
Izoleucin: Százalékban: Az összes triplet %-a
Metionin: Százalékban: Az összes triplet %-a
Treonin: Százalékban: Az összes triplet %-a
Aszparagin: Százalékban: Az összes triplet %-a
Lizin: Százalékban: Az összes triplet %-a
Valin: Százalékban: Az összes triplet %-a
Alanin: Százalékban: Az összes triplet %-a
Glutaminsav: Százalékban: Az összes triplet %-a
Glicin: Százalékban: Az összes triplet %-a
Stop kodon: Százalékban: Az összes triplet %-a

A nonsense mutációk a következő nitrogén alapú bázisokat érinti:

Timin: Százalékban: Az összes bázis %-a
Adenin: Százalékban: Az összes bázis %-a
Citozín: Százalékban: Az összes bázis %-a
Guanin: Százalékban: Az összes bázis %-a
Összes: Az összes bázis %-a

```

13. ábra: A lefutott program kimeneti ablaka eredmények nélkül

## A KUTATÁS EREDMÉNYEI

A génszakasz beillesztése és a program lefuttatása során a következő eredmények születtek. A génszakasznak összesen  $1.5026 \times 10^{2730}$  szubsztitúciós és pontdeléciós pontmutációja lehetséges, ebből mindössze  $7.2409 \times 10^{714}$  a szinonim változatok száma, tehát kevesebb mint 0.1%-a az összes mutációnak.

A nitrogénalapú bázisok közül legnagyobb számban az adenin található meg, ezt követi a guanin, majd citozin és timin. Ebből kiindulva a mutáció legnagyobb valószínűséggel adenint érint (1457 helyen, tehát 37.30% eséllyel), majd guanint (1140 helyen, 29.19%), citozint (712 helyen, 18.23%) és legkisebb valószínűséggel timint (597 helyen, 15.28%).

Ami a kodonokat érinti, a szerin aminosavat kódoló kodonból található meg a legtöbb, ezért az őt érintő mutációk száma lesz a leggyakoribb (542 helyen, tehát 41.63% részarányban), azt követi az aszparagin (259 helyen, 19.89%), glicin (80 helyen, 6.14%), glutaminsav (64 helyen, 4.92%), lizin (43 helyen, 3.3%), treonin (29 helyen, 2.23%), izoleucin (24 helyen, 1.84%), valin és alanin (mindkettő 23 helyen, 1.77% valószínűséggel), glutamin (21 helyen, 1.61%), hisztidin (15 helyen, 1.15%), prolin és arginin (mindkettő 14 helyen, 1.08% valószínűséggel), leucin (10 helyen, 0.77%), tirozin (8 helyen, 0.61%), triptofán és metionin (mindkettő 4 helyen, 0.31% valószínűséggel), fenilalanin (3 helyen, 0.23%), cisztein (2 helyen, 0.15%), és legkevésbé valószínű stopkodon (1 helyen, 0.08%). Szinonim mutációk legnagyobb valószínűséggel a szerinnél jelentkeznek (2710 mutáció, az összes szinonim mutáció 71.28%-a), azt követi az aszparagin (259 mutáció, 6.81%), glicin (240 mutáció, 6.31%), treonin (87 mutáció, 2.29%), arginin (70 mutáció, 1.84%), valin és alanin (mindkettő 69 ilyen mutációval rendelkezik, az ilyen mutációk 1.81%-a), glutaminsav (64 mutáció, 1.68%), leucin (50 mutáció, 1.32%), izoleucin (48 mutáció, 1.26%), lizin (43 mutáció, 1.13%), prolin (42 mutáció, 1.01%), glutamin (21 mutáció, 0.55%) hisztidin (15 mutáció, 0.39%), tirozin (8 mutáció, 0.21%), fenilalanin (3 mutáció, 0.08%) cisztein és stopkodon (mindkettő 2 mutáció, 0.05%-a az összes mutációnak). Triptofán és metionin nem rendelkezik szinonim mutációval.

A szekvencia 3906 nukleotidjából összesen 210 olyan nitrogénalapú bázis található, mely mutációja nonsense mutációt idéz elő a szekvenciában, ezzel az összes bázis 5.38%-át képviselve. Megoszlásuk szerint 101 (48.1%) guanin, 56 (26.67%) adenin, 38 (18.1%) citozin és 15 (7.14%) timin, ezzel az összes bázis 2.59%, 1.43%, 0.97% és 0.38%-át képviselve. A szekvencián az 1., 2., 3., 4., 18., 24., 29., 30., 41., 42., 58., 64., 73., 76., 85., 88., 101., 115., 119., 128., 133., 139., 143., 154., 163., 166., 187., 196., 199., 205., 208., 217., 229., 232., 235., 250., 254., 255., 259., 265., 271., 288., 296., 304., 307., 319., 326., 327., 340., 343., 349., 357., 367., 379., 382., 385., 409., 412., 415., 439., 451., 466., 478., 487., 520., 538., 541.,

547., 553., 562., 586., 592., 601., 612., 613., 619., 634., 646., 658., 661., 673., 676., 682., 700., 703., 748., 757., 763., 769., 787., 793., 796., 808., 811., 832., 838., 841., 847., 859., 862., 889., 896., 904., 907., 912., 915., 922., 928., 931., 946., 955., 961., 970., 976., 979., 1000., 1015., 1018., 1024., 1033., 1036., 1048., 1051., 1057., 1066., 1072., 1081., 1084., 1088., 1090., 1108., 1114., 1120., 1123., 1129., 1135., 1156., 1168., 1171., 1180., 1189., 1195., 1198., 1204., 1207., 1210., 1216., 1226., 1231., 1243., 1249., 1252., 1255., 1270., 1273., 1282., 1285., 1289., 1291., 1297., 1303., 1309., 1345., 1350., 1359., 1372., 1381., 1384., 1408., 1436., 1438., 1462., 1465., 1479., 1489., 1493., 1495., 1520., 1522., 1525., 1531., 1556., 1615., 1619., 1630., 1636., 1640., 1867., 1871., 1882., 1886., 1888., 1915., 1919., 2098., 2572., 2659., 2773., 3247., 3445., 3697., 3802., 3814., 3820., 3844., 3868., 3890., 3904., 3905. és a 3906. nukleotid helyén találhatóak.

A génszekvenciában összesen 202 kodonban elég egy darab pontmutáció, amely következtében nonsense mutáció lép fel. Ebből egy a startkodon mutációja, melyek következtében a fehérjeszintézis nem indul el, egy a stopkodon mutációja, mely következtében a fehérjeszintézis nem áll le a megfelelő helyen. A maradék 200 kodon a 2., 6., 8., 10., 14., 20., 22., 25., 26., 29., 30., 34., 39., 40., 43., 45., 47., 48., 52., 55., 56., 63., 66., 67., 69., 70., 73., 77., 78., 79., 84., 85., 87., 89., 91., 96., 99., 102., 103., 107., 109., 114., 115., 117., 119., 123., 127., 128., 129., 137., 138., 139., 147., 151., 156., 160., 163., 174., 180., 181., 183., 185., 188., 196., 198., 201., 204., 205., 207., 212., 216., 220., 221., 225., 226., 228., 234., 235., 250., 253., 255., 257., 263., 265., 266., 270., 271., 278., 280., 281., 283., 287., 288., 297., 299., 302., 303., 304., 305., 308., 310., 311., 316., 319., 321., 324., 326., 327., 334., 339., 340., 342., 345., 346., 350., 351., 353., 356., 358., 361., 362., 363., 364., 370., 372., 374., 375., 377., 379., 386., 390., 391., 394., 397., 399., 400., 402., 403., 404., 406., 409., 411., 415., 417., 418., 419., 424., 425., 428., 429., 430., 431., 433., 435., 437., 449., 450., 453., 458., 461., 462., 470., 479., 480., 488., 489., 493., 497., 498., 499., 507., 508., 509., 511., 519., 539., 540., 544., 546., 547., 623., 624., 628., 629., 630., 639., 640., 700., 858., 887., 925., 1083., 1149., 1233., 1268., 1272., 1274., 1282., 1290., 1297. és az 1302. kodon helyén található, a kodonok 15.51%-át érintve. Az ilyen fajta nonsense mutáció a kodonok 15.36%-át érinti. Összetételük szerint 64 (31.68%) glutaminsavat, 43 (21.29%) lizint, 27 (13.37%) glicint, 21 (10.4%) glutamint, 15 (7.43%) szerint, 12 (5.94%) arginint, 8 (3.96%) tirozint, 4-4 (1.98%) triptofánt és leucint, 2 (0.99%) ciszteint és 1-1 (0.5%) stop- és startkodont, ezzel az összes kodon 4.92%, 3.3%, 2.07%, 1.61%, 1.15%, 0.92%, 0.61%, 0.31%, 0.31%, 0.15% és 0.15%-át képviselve. A többi kodonnál mindössze egy darab bázis mutációjával a szekvenciában nem jelentkezhet nonsense mutáció.

## A KUTATÁS DISZKUSSZIÓJA

A kapott eredmények rávilágítanak arra, hogy a gén rengeteg lehetséges mutációval rendelkezik, amelyeknek csak kizárólag egy töredéke szinonim mutáció, így nagy valószínűséget adva a keletkező protein szerkezetének megváltozására. Különös szerepet játszik egy-egy elváltozás keletkezésében az a 210 bázismutáció, amely nonsense mutációt idézhet elő, ugyanis ezek mutációja a fehérje hosszát vagy drasztikusan lerövidíti, és a fehérjeszintézis idő előtt leáll, vagy ellenkezőleg, meghosszabbítja, ugyanis stopkodon mutációja következtében a fehérjeszintézis nem áll le, vagy a fehérjeszintézis elkezdését akadályozza meg a startkodon mutációjával. A bázisok és a kodonok összetételét ismerve következtetni lehet arra, hogy a mutáció melyik aminosavnál, vagy a szekvenciát figyelve melyik kodonnál vagy bázisnál jelentkezhet, ezzel kiszámíthatóbbá téve az elváltozások mechanizmusait.

A kutatás során felhasznált információk többszörösen is ellenőrizve lettek. A génszekvencia standardizált változata több génbázisban is elérhető (Ensembl, National Center for Biotechnology Information, UniProt), amely szerverek által feltüntetett génszekvenciák között nem található eltérés.

## ÖSSZEFOGLALÓ

A megközelítőleg 8.35kb hosszúságú DSPP-gén elsősorban a fogászatban jelentős gén. Az általa termelt dentin sialophosphoprotein modifikáción esik át, és két kisebb egységgé válik szét, melyek a dentin mineralizációját segítik. Mutációjakor dentinogenesis imperfecta léphet fel, mely a dentint érintő kóros elváltozás. Következtében a dentin szervesanyag-tartalma kritikusan lecsökken, ami maga a dentin minőségi romlásához vezet. A kutatás egy saját fejlesztésű szoftverrel történt, amellyel elsősorban logikai és kombinatorikai módszereket alkalmazva a gén proteinkódoló szekvenciája összes lehetséges mutációjának előfordulását lehet elemezni. A kutatás eredményeként arra lehet következtetni, hogy a génnek  $1.5026 \times 10^{2730}$  pontmutációja van, amelyből  $7.2409 \times 10^{714}$  szinonim. A mutáció leggyakrabban adenint érint (1457 helyen), legkisebb valószínűséggel pedig timint (597 helyen). A kodonok közül legnagyobb valószínűséggel a szerint kódoló kodon mutálódik (542 helyen, 41.63%-ot adva) valamint nála jelentkezhet a legtöbb szinonim mutáció (2710 mutáció, 71.28%-ot adva), legkevésbé valószínű pedig, hogy a stopkodon mutálódik (csak 1 helyen, 0.08%-ot képviselve), legkevésbé szinonim mutációval pedig a triptofánt és a metionint kódoló kodonok rendelkeznek (mindkét esetben 0). Emellett összesen

210 nitrogénalapú bázis és 202 olyan kodon van, amelynél egy mutáció elég nonsense mutáció kialakításához, ebből egy a stopkodon, egy a startkodon, a maradék 200 ilyen kodon pedig elszórtan található meg a szekvencián.

## IRODALOM

- BUDAKOV, Pavle 2011. Patologija. Medicinski fakultet Novi Sad, Novi Sad.
- DEÁK Veronika 2014. Általános genetika. Typotex Elektronikus Kiadó Kft, Budapest.
- European Bioinformatics Institute (EMBL-EBI), SIB Swiss Institute of Bioinformatics, Protein Information Resource (PIR): UniProt. URL: <https://www.uniprot.org/> (Letöltés: 2022. 02. 11.)
- European Molecular Biology Laboratory's European Bioinformatics Institute: Ensembl. URL: <https://www.ensembl.org/> (Letöltés: 2022. 02. 11.)
- HUGO Gene Nomenclature Committee: Gene Names. URL: <https://www.genenames.org/> (Letöltés: 2022. 02. 23.)
- McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine: OMIM® – Online Mendelian Inheritance in Man. URL: <https://omim.org/> (Letöltés: 2022. 02. 24.)
- National Library of Medicine: National Center for Biotechnology. URL: <https://www.ncbi.nlm.nih.gov/> (Letöltés: 2022. 02. 24.)
- POPIĆ-PALJIĆ, Feodora 2016. Humana genetika. Medicinski fakultet Novi Sad, Novi Sad.
- TURNPENNY, Peter – ELLARD Sian 2009. Emerijevi osnovi medicinske genetike. Ford. Novaković, Ivana; Vapa, Ljiljana. Data Status, Beograd.
- U. S. National Library of Medicine: Medline Plus. URL: <https://medlineplus.gov/> (Letöltés: 2022. 02. 24.)

E számunk támogatói: a Tartományi Művelődési,  
Tájékoztatási és Vallásügyi Titkárság és a Magyar Nemzeti Tanács

