

UNIVERZA V LJUBLJANI
FAKULTETA ZA FARMACIJO

KAJA MIVŠEK

**OPREDELITEV GENETSKIH VZROKOV
PREZGODNJE PUBERTETE
THE IDENTIFICATION OF THE GENETIC
CAUSE OF PRECOCIOUS PUBERTY**

MAGISTRSKI ŠTUDIJSKI PROGRAM LABORATORIJSKA BIOMEDICINA

Ljubljana, 2017/2018

Magistrsko nalogo sem opravljala na Pediatrični kliniki, UKC Ljubljana, Služba za specialno laboratorijsko diagnostiko pod mentorstvom izr. prof. dr. Katarine Trebušak Podkrajšek, univ. dipl. kem., spec. med. biokem., spec. lab. med. gen. in somentorstvom doc. dr. Magdalene Avbelj Stefanije, dr. med.

Izjava o avtorstvu

Izjavljam, da sem magistrsko nalogo samostojno izdelala pod mentorstvom izr. prof. dr. Katarine Trebušak Podkrajšek, univ. dipl. kem., spec. med. biokem., spec. lab. med. gen. in somentorstvom doc. dr. Magdalene Avbelj Stefanije, dr. med.

Kaja Mivšek

ZAHVALA

Izr. prof. dr. Katarina Trebušak Podkrajšek, univ. dipl. kem., spec. med. biokem., spec. lab. med. gen., zahvaljujem se Vam za pripravljenost in pomoč ter vso podporo pri magistrskem delu. Hvala, ker ste bili moja mentorica.

Doc. dr. Magdalena Avbelj Stefanija, dr. med. iz Kliničnega oddelka za endokrinologijo, diabetes in bolezni presnove na Pediatrični kliniki, UKC Ljubljana, iskrena hvala za posvečen čas, potrpežljivost in strokovno pomoč.

Dr. Jernej Kovač, uni. dipl. biokem., najlepša hvala za izjemno vodstvo skozi praktični del, za vse nasvete in ideje, ki so me usmerjali in mi omogočili dosego končnega cilja.

Mirjana Zupančič, mag. med. biokem., hvala, da sem praktični del magistrske naloge lahko opravljala v Službi za specialno laboratorijsko diagnostiko Pediatrične klinike.

Tamara Obreza, mag. lab. biomed., iz srca hvala za vse nasvete, velikodušnost in pomoč pri magistrski nalogi. Hvala ti za čas in pogovore, ki so me vedno razveselili.

Zahvaljujem se vsem **ostalim iz Službe za specialno laboratorijsko diagnostiko na Pediatrični kliniki, UKC Ljubljana**, za prijaznost, pomoč in skupno preživeti čas tekom mojega raziskovalnega dela.

Prijatelji, iskreno hvala za smeh, druženja in skupno preživete trenutke.

Posebna zahvala gre **moji družini, bratu Luku, mami in očetu**. Iz srca hvala, da ste mi stali ob strani v vseh trenutkih, me podpirali in mi omogočili študij. Hvala vam za vse pogovore, nasvete in mnoga skupna druženja, brez vas bi bilo življenje težje.

Tomaž, srčno hvala, da verjameš vame, mi stojiš ob strani in me osrečuješ.

Bog, hvala za življenje, upanje in vero.

KAZALO VSEBINE

POVZETEK	I
ABSTRACT	II
SEZNAM OKRAJŠAV IN SIMBOLOV	III
1 UVOD.....	1
1.1 NASTOP PUBERTETE.....	1
1.1.1 PREZGODNJA PUBERTETA	1
1.1.2 KLASIFIKACIJA PREZGODNJE PUBERTETE.....	2
1.1.3 DRUŽINSKA CENTRALNA PREZGODNJA PUBERTETA.....	3
1.1.4 KLINIČNE ZNAČILNOSTI PREZGODNJE PUBERTETE.....	3
1.1.5 DIAGNOSTIKA PREZGODNJE PUBERTETE.....	4
1.1.6 GENETSKO OZADJE CENTRALNE PREZGODNJE PUBERTETE	6
1.1.7 ZDRAVLJENJE	8
1.2 GENOMSKA DNA IN TEHNOLOGIJA ANALIZE DNA S SEKVENCIRANJEM NASLEDNJE GENERACIJE	9
1.2.1 OSNOVE SEKVENCIRANJA NASLEDNJE GENERACIJE (angl. <i>NEXT GENERATION SEQUENCING – NGS</i>).....	9
1.2.2 BIOINFORMATSKA OBDELAVA REZULTATOV PRIDOBLENIH Z PRISTOPOM SEKVENCIRANJA CELOTNEGA GENOMA	10
2 NAMEN DELA.....	15
3 PREISKOVANCI IN METODE.....	16
3.1 PREISKOVANCI.....	16
3.2 METODE	16
3.2.1 IZOLACIJA GENOMSKE DNA.....	16
3.2.2. SPEKTROFOTOMETRIČNO PREVERJANJE ČISTOSTI IZOLIRANE DNA	17
3.2.3 AGAROSNA ELEKTROFOREZA	17
3.2.4. FLUORIMETRIČNO MERJENJE KONCENTRACIJE DNA	18
3.2.5 PRIPRAVA DNA KNJIŽNICE IN SEKVENCIRANJE CELOTNEGA GENOMA.....	19
3.2.6 BIOINFORMATSKA ANALIZA OBDELANIH REZULTATOV PRIDOBLENIH S PRISTOPOM SEKVENCIRANJA CELOTNEGA GENOMA – ISKANJE VZROČNE GENETSKE SPREMEMBE V GENOMU.....	26

3.2.7 POTRJEVANJE S SEKVENCIRANJEM PO SANGERJU.....	29
4 REZULTATI.....	35
4.1 KLINIČNE ZNAČILNOSTI PREISKOVANCEV	35
4.2 SEKVENCIRANJE CELOTNEGA GENOMA IN BIOINFORMATSKA ANALIZA	35
4.3 POTRREDITVENA ANALIZA	40
5 RAZPRAVA.....	42
5.1 IZVOR NAPAK PRI SEKVENCIRANJU CELOTNEGA GENOMA IN BIOINFORMATSKI ANALIZI REZULTATOV	43
5.1.1 NAPAKE V OBDELAVI PODATKOV PRIDOBLENIH S SEKVENCIRANJEM CELOTNEGA GENOMA.....	43
5.1.2 SLABŠA OBČUTLJIVOST IN TOČNOST SEKVENCIRANJA CELOTNEGA GENOMA.....	44
5.1.3 KOLIČINA PODATKOV	45
5.1.4 KOMPLEKSNOST GENOMA.....	46
5.1.5 NARAVA DEDNE BOLEZNI	46
5.2 POMEN POTRJEVANJA REZULTATOV SEKVENCIRANJA NASLEDNJE GENERACIJE.....	48
5.3 NADALJNI KORAKI ZA OPREDELITEV GENETSKEGA VZROKA	49
5.3.1 ANALIZA GENETSKIH SPREMEMB VRST CNV IN SV	49
5.3.2 INTERPRETACIJA NEKODIRAJOČIH PODROČIJ GENOMA	50
5.3.3 PREISKAVA CELOTNEGA GENOMA S CILJEM OPREDELITVE GENETSKIH SPREMEMB PO POSAMEZNI DRUŽINI	50
5.3.4 PREISKAVA CELOTNEGA GENOMA S CILJEM OPREDELITVE GENETSKIH SPREMEMB SKUPNIH VSEM HČERKAM	50
5.3.5 EPIGENETSKE PREISKAVE.....	51
6 SKLEP.....	51
7 LITERATURA	53
8 PRILOGA.....	61

KAZALO PREGLEDNIC

PREGLEDNICA I. Delitev prezgodnje pubertete.	3
PREGLEDNICA II. Pojav sprememb v času pubertete.	4

PREGLEDNICA III. Lastnosti genov in proteinov, ki jih ti geni kodirajo in so povezani z družinsko in idiopatsko CPP.	7
PREGLEDNICA IV. Lastnosti kandidatnih genov in proteinov ki jih ti kandidatni geni kodirajo in so povezani z družinsko in idiopatsko CPP. ...	8
PREGLEDNICA V. Vrste genetskih sprememb in pripadajoči tipi v kodirajočih regijah genoma.	13
PREGLEDNICA VI. Razmerje med QPHRED vrednostjo in e-vrednostjo.	14
PREGLEDNICA VII. Oligonukleotidna začetnika z opisanimi lastnostmi.	30
PREGLEDNICA VIII. Podatki o značilnostih preiskovancev.	35
PREGLEDNICA IX. Z algoritmom cadd opredeljene genetske spremembe v genu <i>ZNF717</i>	38
PREGLEDNICA X. Nadaljevanje <i>Preglednice IX</i>	39

KAZALO SLIK

SLIKA 1. Shema poteka obdelave rezultatov WGS.	10
SLIKA 2. Odsek gDNA ligiran z indeksi označenimi adapterji.	20
SLIKA 3. Delovna shema za pripravo knjižnice DNA.	21
SLIKA 4. Popravilo in fosforilacija koncev odsekov gDNA.	22
SLIKA 5. Primer ligacije odseka DNA z indeksi označenimi adapterji.	23
SLIKA 6. Premostitveni PCR (<i>angl. Bridge PCR</i>).	25
SLIKA 7. Primer Vennovega diagrama presekov sprememb mame, očeta in hčerke družine 4.	26
SLIKA 8. Pokritost zaporedja DNA.	27
SLIKA 9. Shema postopka filtracije sprememb pri posamezni družini.	28
SLIKA 10. Rezultat WGS analize.	36
SLIKA 11. Rezultat bioinformatične analize rezultatov pridobljenih z metodo WGS.	40
SLIKA 12. Z elektroforezo ločeni odseki gDNA.	41
SLIKA 13. Rezultat potrditvene analize.	41
SLIKA 14. Mesto nahajanja gena <i>ZNF717</i> na kromosomu 3.	42
SLIKA 15. Rezultat bioinformatične obdelave podatkov pridobljenih z metodo WGS.	62

POVZETEK

Interakcije med genetskimi in okoljskimi dejavniki določajo čas začetka spolnega dozorevanja. Puberteta se prične, ko se hipotalamično-hipofizno-gonadna os ponovno aktivira. Takrat opazimo razvoj žleznega tkiva dojke pri deklicah oz. povečanje mod pri dečkih ter sočasno pospešeno rast in dozorevanje kosti. Kadar se omenjeni znaki pojavijo pred pričakovano starostjo, je možen vzrok zanje centralna prezgodnja puberteta.

Prezgodnji spolni razvoj pomeni premik rodne dobe, ki ga povežemo s tveganjem za razvoj nekaterih bolezni, večjo nagnjenostjo k nižji odrasli postavi in možnostjo nastanka psihosocialnih motenj. Za opredelitev morebitnega dednega vzroka bolezni so potrebne molekularne genetske preiskave.

Namen magistrske naloge je bil odkriti genetske spremembe, ki bi lahko bile odgovorne za prezgodnjo puberteto. Določili smo nukleotidno zaporedje celotnega genoma štirih družin s prezgodnjo puberteto, ki se deduje po materi. Uporabili smo pristop sekvenciranja celotnega genoma in z bioinformatično analizo skušali opredeliti možne genetske vzroke za prezgodnjo puberteto.

Z uporabo orodij in algoritmov smo našli spremembe, ki so pri vseh štirih družinah ležale v istem genu *ZNF717* in jih tako prepoznali kot možno vzročne. Kvantitativno smo ocenili patogenost omenjenih sprememb. Pri članih četrte družine smo prisotnost najverjetnejše patogene spremembe preverjali s potrditvenim sekvenciranjem po Sangerju.

Rezultat potrditvene analize je pokazal odsotnost iskane spremembe, kar kaže na napako v pristopu sekvenciranja celotnega genoma oziroma sledeče bioinformatične analize.

Razpravljali smo o vzrokih za lažno pozitiven rezultat ter predstavili slabosti pristopa s sekvenciranjem celotnega genoma in bioinformatične analize.

Rezultati prvotnih analiz niso bili zanesljivi, kar pomeni, da potencialno vzročne genetske spremembe pri analiziranih družinah s prezgodnjo puberteto ostajajo neopredeljene. Naši rezultati kažejo, da je preverjanje rezultatov sekvenciranja naslednje generacije s potrditveno metodo nujno potrebno za zanesljivo genetsko diagnostiko.

Ključne besede: Družinska centralna prezgodnja puberteta, sekvenciranje celotnega genoma, bioinformatična analiza, sekvenciranje po Sangerju, lažno pozitiven rezultat.

ABSTRACT

Interactions between genetic and environmental factors determine the timing of sexual maturation. Puberty begins when the hypothalamic-pituitary-gonadal axis is reactivated. At that time, mammary gland in girls and testes in boys are developed and growth and bone maturation are accelerated. When this pubertal characteristics occur before the expected age, possible cause for them is central precocious puberty. Premature pubertal development results in the shift of a fertile period, which is associated with the risk of developing certain diseases, a greater tendency towards the lower final adult stature and increased potential for psychosociological disorders. Molecular genetic testing is needed to identify the potential hereditary cause of the disease.

The aim of this study was to detect genetic variants, that could be responsible for precocious puberty. We determined the sequence of the entire genome in members of four families with maternally inherited precocious puberty. We applied the whole genome sequencing followed by bioinformatic approaches in order to identify the possible genetic causes of precocious puberty.

Using tools and algorithms, we identified variants in all four families located in the same gene *ZNF717* and therefore determined them as potentially pathogenic. We quantitatively assessed the pathogenicity of these variants. In members of the fourth family, we verified the presence of the most likely pathogenic variant by confirmative Sanger sequencing.

Results of Sanger analysis showed, that the desired variant was not present, indicating error in the approach of whole genome sequencing and subsequent bioinformatic analysis. The causes of the false positive result were discussed and the weaknesses of the whole genome sequencing method and the bioinformatic analysis were presented.

The results of the initial analyses were not reliable, therefore potentially causative genetic variants in analyzed families with precocious puberty remain undetermined. Our results showed that verification of the results of next generation sequencing with a secondary method is necessary for reliable genetic diagnostics.

Key words: Familial central precocious puberty, sequencing of the entire genome, bioinformatics analysis, Sanger sequencing, a false positive result.

SEZNAM OKRAJŠAV IN SIMBOLOV

A	adenin
A260	absorbanca pri 260 nanometrih
A280	absorbanca pri 280 nanometrih
AD	avtosomno dominantno
BA	starost kosti
BAM	binarna mapa poravnave (angl. <i>Binary Alignment Map</i>)
bp	bazni par
BWA	algoritem Burrows-Wheller Aligner za poravnavo nukleotidnega zaporedja
C	citozin
CA	kronološka starost
CADD	algoritem za ocenjevanje patogenosti genetskih sprememb (angl. <i>Combined Annotation Dependent Depletion - CADD</i>)
ChIP-Seq	kombinacija kromatinske imunoprecipitacije in sekvenciranja naslednje generacije (angl. <i>Chromatin immunoprecipitation with sequencing – ChIP-Seq</i>)
CNV	spremembe v številu kopij (angl. <i>Copy number variants - CNV</i>)
CPP	centralna prezgodnja puberteta
Cq	kvantifikacijski cikel
CŽS	centralni živčni sistem
ddNTP	dideoksinukleotid trifosfat
<i>de novo</i>	na novo pridobljena genetska sprememba v generaciji
DNA	deoksiribonukleinska kislina
dNTP	deoksinukleotid trifosfat
dsDNA	dvojnovijačna deoksiribonukleinska kislina
e	stopnja napake zaporedja (angl. <i>sequence error rate</i>)
EDTA	etilendiamintetraocetna kislina
F	smer naprej
FSH	folikel-stimulirajoči hormon
G	gvanin
GB	gigabajt
gDNA	genomska deoksiribonukleinska kislina

GnRH	gonadotropin sproščujoči hormon
hCG	humani horionski gonadotropin
HHG	hipotalamično-hipofizno-gonadna os
INDEL	spremembe insercija-delecija (angl. <i>Insertion-deletion - INDEL</i>)
LH	luteinizirajoči hormon
min	minuta
MRI	magnetna resonanca
mRNA	informacijska RNA
NGS	sekvenciranje naslednje generacije (angl. <i>Next generation sequencing</i>)
PCR	verižna reakcija s polimerazo
PPP	periferna prezgodnja puberteta
pre-mRNA	predhodna informacijska RNA
qPCR	verižna reakcija s polimerazo v realnem času
Qphred	enota za kvaliteto določitve baze (angl. <i>base quality</i>)
R	obratna smer
RNA	ribonukleinska kislina
s	sekunda
SAM	sekvenčna mapa poravnave (angl. <i>Sequence Alignment Map</i>)
SNP	polimorfizem posameznega nukleotida
SNV	eno-nukleotidne spremembe (angl. <i>Single nucleotide variants - SNV</i>)
SV	strukturne spremembe (angl. <i>Structural variants - SV</i>)
T	timin
TBE	Tris-Borat-Etilendiamintetraocetna kislina
UV	ultravijolični del spektra
WES	sekvenciranje celotnega eksoma (angl. <i>Whole exome sequencing</i>)
WGBS	bisulfitno sekvenciranje celotnega genoma (angl. <i>Whole genome bisulfite sequencing - WGBS</i>)
WGS	sekvenciranje celotnega zaporedja genoma (angl. <i>Whole genome sequencing</i>)

1 UVOD

1.1 NASTOP PUBERTETE

Puberteta je življenjsko obdobje, v katerem posameznik spolno dozori in postane ploden. Zanj so značilne fiziološke in psihološke spremembe (1). Na zunaj opazimo tipične spremembe, ki jim pravimo tudi označevalci pubertete: telesna postava raste in se preoblikuje; razvijejo se primarni in sekundarni spolni znaki (2).

V puberteti se aktivira hipotalamično-hipofizno-gonadna os (v nadaljevanju HHG os), kar imenujemo tudi gonadarha (3). Pri deklicah se kaže s telarho (razvoj prsi) in kasneje v razvoju z menarho (prva menstrualna krvavitev) (3). Pri dečkih gonadarho označuje razvoj in povečanje mod (3). Pri obeh spolih se neodvisno od gonadarhe pojavi adrenarha (zvišanje androgenih hormonov nadledvične žleze), ki ima med drugim za posledico pubarho (rast pazdušnih in sramnih dlak) (3).

HHG os je v resnici aktivna že v embrionalnem obdobju in v času takoj po rojstvu, vendar se v zgodnjem otroštvu utiša (3). V času pubertete se preko mreže inhibitornih in ekscitatornih dejavnikov ponovno aktivira in doseže vrhunec delovanja (3). Čas začetka njene ponovne aktivacije označuje pričetek pubertete, ki je različen med rasami, spoloma in posamezniki (2, 3). Običajno se pri deklicah pojavi med 8. in 13. letom, pri dečkih pa kasneje – med 9. in 14. letom; poleg tega pri populacijah afriškega izvora spremembe nastopijo leto prej kot pri populacijah evropskega izvora (1, 3).

1.1.1 PREZGODNJA PUBERTETA

Primerjave zdajšnjih podatkov s tistimi izpred 50. let kažejo, da se starost ob pričetku pubertete znižuje (4, 5). Puberteto smatramo kot prezgodnjo, kadar se označevalci pojavijo pred 8. letom pri deklicah in pri dečkih pred 9. letom starosti (3, 6). Znaki centralne prezgodnje pubertete (gonadarhe) so razvoj žleznega tkiva dojk pri deklicah in povečanje mod pri dečkih ter pospešena rast, kar pomeni, da so se kosti začele pospešeno daljšati in dozorevati (3). Prezgodnji začetek pubertete nosi tveganje za razvoj bolezni; obstaja večja verjetnost za razvoj raka dojke in endometrioze, nagnjenosti k debelosti, sladkorni bolezni tipa 2 in srčno-žilnim boleznim (4, 7, 8, 9, 10). Zgodnejša menstruacija pri ženskah pomeni zgodnejšo menopavzo, kar zaradi omejenega števila jajčec premakne obdobje rodne dobe.

Ker je za prezgodnjo puberteto značilno hitrejšo zorenje skeleta, se rast predhodno zaključuje, zato so bolniki nagnjeni k nizki odrasli postavi. V primeru nezdravljene centralne prezgodnje pubertete (v nadaljevanju CPP) je višina odraslih deklet nižja za 8-12 cm in 12-20 cm pri odraslih dečkih (11). Pojavljajo se lahko psihosocialne motnje v obnašanju in vedenju (2, 3, 12). Otroci s prezgodnjo puberteto imajo večje tveganje za zgodnejše spolne odnose, zlorabo prepovedanih substanc in slabše šolske rezultate (13).

Trenutna pojavnost prezgodnje pubertete v populaciji je od 1/500 do 1/10.000 in je 10-krat pogostejša pri deklicah kot pri dečkih (11, 14, 15).

1.1.2 KLASIFIKACIJA PREZGODNJE PUBERTETE

Obstajata dva tipa prezgodnjega pubertetnega razvoja. Razdelimo ju glede na vzroke, ki ju povzročijo, kar prikazuje *Preglednica I* (povzeto po 1, 3, 16):

- a) Od gonadotropinov odvisna oblika je imenovana tudi centralna prezgodnja puberteta (CPP). Značilna je prezgodnja reaktivacija HHG osi, kar vodi v prezgodnje sproščanje gonadotropin sproščujočega hormona (GnRH) iz hipotalamusa, ki vpliva na hipofizo, da sprosti gonadotropina; luteinizirajoči hormon (LH) in folikel-stimulirajoči hormon (FSH). Slednja delujeta na jajčnike in moda tako, da spodbujata izločanje spolnih steroidnih hormonov, estrogenov za ženske in androgenov za moške spolne značilnosti (1). Vzroki so okvirno v 70 - 80 % genetski in v 30 - 20 % okoljski in epigenetski (prehrana in z njo povezani endokrini motilci) (4). V manjšem deležu so vzrok razna obolenja, okužbe ali poškodbe centralnega živčnega sistema (CŽS). Kadar je vzrok nepoznan je to idiopatska CPP.
- b) Od gonadotropinov neodvisna oblika je imenovana tudi periferna prezgodnja puberteta (PPP) ali prezgodnja pseudopuberteta. Je redkejša oblika, za katero je značilno povečano nastajanje spolnih steroidnih hormonov zaradi motenj v delovanju spolnih ali nadledvičnih žlez. Spolni steroidi so lahko tudi eksogenega izvora, če jih vnesemo iz okolja.

Preglednica I. Delitev prezgodnje pubertete (PP – prezgodnja puberteta, CPP – centralna prezgodnja puberteta, PPP – periferna prezgodnja puberteta) (povzeto po 3, 17, 19).

Od gonadotropinov odvisna PP ali CPP	Od gonadotropinov neodvisna PP ali PPP
<p>Brez sprememb v CŽS:</p> <ul style="list-style-type: none"> - genetski vzroki (družinska CPP), - kronična izpostavljenost spolnim steroidnim hormonom, - izpostavljenost endokrinim motilcem, - idiopatska CPP. 	<p>Tumorji:</p> <ul style="list-style-type: none"> - nadledvičnih žlez, jajčnikov, mod, - ciste jajčnikov.
<p>Prisotne spremembe v CŽS (nevrogena CPP):</p> <ul style="list-style-type: none"> - tumorji, - kongenitalne malformacije: hipotalamični hamartom, arahnoidne ciste in drugo, - pridobljene bolezni: infekcije in vnetja CŽS. 	<p>Hujši dolgoročni nezdravljeni primarni hipotiroidizem.</p>
	<p>Genetski vzroki:</p> <ul style="list-style-type: none"> - testotoksikoza, - McCune-Albrightov sindrom, - Prader-Willijev sindrom, Williamov sindrom, Templov sindrom, - mutacije v signalnih poteh gonadotropinov.
	<p>Izpostavljenost eksogenim spolnim steroidom ali njihovim analogom.</p>

1.1.3 DRUŽINSKA CENTRALNA PREZGODNJA PUBERTETA

O družinski CPP govorimo, kadar obstaja več kot en član znotraj družine ali v rodovniku, ki je doživel prezgodnji pubertetni razvoj (18, 19, 20). Vpliv genetskih faktorjev na čas začetka pubertete je 50-80 % (16, 17, 19, 20). Družinska zgodovina s prezgodnjim pubertetnim razvojem je pozitivna pri 19-27, 5 % bolnikih s CPP (19, 20).

1.1.4 KLINIČNE ZNAČILNOSTI PREZGODNJE PUBERTETE

V *Preglednici II* so zbrane in opisane fizične spremembe, ki jih opazimo v posameznih fazah pri normalnem nastopu pubertete (povzeto po 21, 22). Znaki so klinično pomembni in so lahko znak prezgodnje pubertete, kadar se pojavijo pred 8. letom pri deklicah in pred 9. letom starosti pri dečkih.

Preglednica II. Pojav sprememb v času pubertete, razdeljenih po posameznih fazah (znana Tanner-jeva razdelitev faz pubertetnega razvoja (povzeto po 21, 22)).

Faza	Ženski spol				Moški spol			
	Starost (v letih)	Razvoj prsi	Sramne dlake	Ostale spremembe	Starost (v letih)	Povečanje mod	Povečanje penisa	Sramne dlake
I	0–15	/	Brez	/	0–15	Ni (≤ 2.5 cm)	/	/
II	8–15	telarha, areolarna hiperplazija z manjšim deležem tkiva dojke	Daljšje puhaste sramne dlake blizu sramnih ustnic	Povečana rast se pogosto pojavi kmalu po nastopu II. faze	10–15	Povečanje mod, pigmentacija skrotuma	Minimalna ali je ni	Daljšje puhaste dlake
III	10–15	Nadaljnjo povečanje dojk in areole, ni jasno vidne meje med njima	Podaljšanje in pigmentacija	Pozno v III. stopnji pri 2% deklic pojav menarhe	1.5–16.5	Nadaljnjo povečanje	občutno povečanje premera	Podaljšanje, kodranje
IV	10–17	Razločimo areolo in bradavico, nad dojko se oblikuje kupček	Podobne odraslemu, a ne še v načinu porazdeljenosti	1-3 leta po telarhi se pri večini deklic v fazi IV pojavi menarha	Spremenljivo: 12–17	Nadaljnjo povečanje	Nadaljnjo povečanje, predvsem v premeru	Podobne odraslemu, a ne še v načinu porazdeljenosti
V	12.5–18	Polna velikost dojk	Porazdelitev kot pri odraslem	Pri 10% deklet se pojavi menarha v fazi V	13–18	Odrasla velikost	Odrasla velikost	Porazdelitev kot pri odraslem (tudi po stegnih, prsih in trebuhu)

1.1.5 DIAGNOSTIKA PREZGODNJE PUBERTETE

Zgodnja prepoznavna in potrditev bolezni je pomembna zato, ker je lahko znanilec življenjsko ogrožajočih bolezni in, da se po potrebi prične ustrezen način zdravljenja, s katerim preprečimo škodljive vplive na rast in omogočimo razvoj v primerni starosti.

Nujna je izključitev drugih možnih s prezgodnjo puberteto povezanih bolezni, kot so: hCG tumor, okvarjena tiroidna funkcija, tumorji centralnega živčevja, jajčnikov in nadledvične žleze, neskladen pubertetni razvoj (McCune Albright sindrom, ciste jajčnikov, izolirana telarha, izolirana vaginalna krvavitev zaradi infekcije, spolne zlorabe, tumorja ali prisotnega tujka) (3, 11, 23).

Postavitev diagnoze temelji na podatkih iz laboratorijskih preiskav in kliničnih značilnostih. Potek diagnostike je opisan v nadaljevanju (povzeto po 11, 21-28):

- PODATKI O BOLNIKU

Potrebno je pridobiti informacije o starosti ob nastopu pubertetnega razvoja, trenutno prisotnih sekundarnih spolnih znakih, ocenjeno hitrost razvoja ter vzorec rasti v zadnjih 6-12 mesecih.

- TELESNI PREGLED

Določi se fazo razvoja glede na Tanner-jevo razdelitev pubertetnega razvoja po starostnih obdobjih (glej *Preglednico II*), izmeri se višino, težo in poda telesna razmerja (23).

- SKELETNA STAROST

Bolj kot kronološka starost, ki je določena z datumom rojstva, nam o biološki in strukturalni starosti pove starost kosti. Določimo jo z radiografsko metodo, pri kateri se z x-žarki presvetli kosti ne-dominantne roke in dlani (25). Kadar je starost kosti povečana za več kot 2SD kronološke starosti, obstaja verjetnost, da gre za prezgodnji pubertetni razvoj. Razmerje med starostjo kosti in kronološko starostjo ($\Delta BA / \Delta CA$) je nad 1-2 pri večini pacientov s progresivno CPP (26).

- HORMONSKO STANJE

Standardni GnRH test je visoko občutljivi imunokemiluminiscenčni test, ki omogoča določitev povišanih serumskih bazalnih in stimuliranih koncentracij LH. Za zadnje je značilno, da močno korelirajo z LH stimuliranim z endogenim GnRH, zato se ta test uporablja za presejanje bolnikov s CPP (27, 28). Intravensko se injicira GnRH, nato pa se v času 0, po 20, 30 in 60 minutah izmeri koncentracije LH. Za prezgodnjo puberteto so klinično pomembne tudi zvišane bazalne koncentracije plazemskega testosterona pri dečkih. Diagnostična vrednost estradiola je omejena, saj polovica deklic s CPP nima povišanih vrednosti. Laboratorijske preiskave glede na klinično sliko lahko vključujejo še teste ščitnične funkcije in določanje vsebnosti 17-hidroksiprogesterona (17-OHP) ter človeškega horionskega gonadotropina (hCG), s katerimi se opredeli morebitne ostale vzroke prezgodnje pubertete; kongenitalno adrenalno hiperplazijo in človeški horionski gonadotropin izločujoči tumor.

- ULTRASONOGRAFIJA MEDENICE

Volumni jajčnikov in maternice so primerljivi s starostjo. Oblika maternice je v pre-pubertetnem obdobju ovalna, v pubertetnem pa hruškasta. Struktura jajčnikov je v pre-puberteti homogena/mikrocistična in multicistična/makrocistična/folikularna v puberteti. Debelina maternične sluznice je večja od 3 mm v pubertetnem obdobju (24). Sonografija trebušne votline in medenice izključi možne tumorje nadledvične žleze in jajčnikov pri bolnikih s prezgodnjo puberteto.

- MRI ČŽS

Nujno je opraviti MRI slikanje pri otrocih s CPP, zato da se izključi prisotnost lezij v ČŽS. Pri nekaterih bolnikih s prezgodnjo puberteto je CPP lahko edini prisotni znak intrakranialnega tumorja (29). Pri 4-19 % otrok s CPP lahko pričakujemo skrite intrakranialne lezije (11). Najpogostejša lezija ČŽS, ki se kaže s CPP, je hipotalamični hamartom, katerega odkrijemo le z MRI.

- DIFERENCIALNA DIAGNOSTIKA

Zlati standard, ki razlikuje med CPP in PPP, je GnRH test. Za CPP je značilno, da se gonadotropini odzovejo na GnRH, pri čemer se LH poviša bolj kot FSH. Pri PPP so ob povišanih spolnih hormonih gonadotropini nizki oz. zavrti.

- POZITIVNA DRUŽINSKA ZGODOVINA

Pozitivna družinska zgodovina lahko pomeni deden vzrok CPP. V pošteev pride genetska preiskava in potrditev genetskih sprememb, ki so vzrok za bolezen.

1.1.6 GENETSKO OZADJE CENTRALNE PREZGODNJE PUBERTETE

Kompleksne interakcije med genetskimi in okoljskimi dejavniki so tiste, ki izzovejo aktivacijo HHG osi in privedejo do začetka pubertete. Z morfološkimi preiskavami pogosto ne ugotovimo vzroka CPP, zato genetske analize omogočajo predvideti možen genetski vzrok za CPP. Analize vzročnih genov so dodaten pripomoček za diagnozo in primerno zdravljenje otrok s tem kliničnim stanjem, omogočajo pa tudi opredelitev pri družinski članih in možnost CPP pri mlajših sorojencih.

Z razvojem družinske in idiopatske CPP so povezani geni, ki jih prikazuje *Preglednica III*. Opisani so primeri bolnikov, ki imajo prisotne vzročne spremembe v genih *KISS1*, *KISS1R* ter *MKRN3* genih in so odgovorni za družinsko obliko CPP. Pogost genetski vzrok je pomanjkanje peptida odgovornega za inhibicijo reproduktivne osi, ki ga kodira gen *MKRN3* (30-35).

Po večini se družinska CPP deduje avtosomno dominantno z od spola neodvisno nepopolno penetranco. Drugače je v primeru *MKRN3* gena, kjer je alel, ki je podedovan od matere, metilirani oz. vtisnjen, zato se ne izraža. Izrazi se le alel podedovan po očetu (19, 33, 36, 37). Na izločanje GnRH delujejo inhibitorne, stimulatorne in permisivne mreže molekul, ki jih kodirajo geni *GABRA1*, *NPY-Y1R*, *LIN28B*, *TAC3* in *TACR3* (glej *Preglednico IV*). Vseeno v raziskavah do sedaj pri bolnikih s CPP v teh genih ter genih *GNRH1* in *GNRHR* niso našli vzročnih sprememb (19, 35, 38).

Preglednica III. Lastnosti genov in proteinov, ki jih ti geni kodirajo in so povezani z družinsko in idiopatsko CPP. (AD - avtosomno dominantno, PP - prezgodnja puberteta, dCPP - družinska CPP, iCPP - idiopatska CPP, nHH - normozmični prirojeni hipogonadizem) (povzeto po 3, 18, 20, 30, 35, 38, 40).

Gen	Lega	Dedovanje	Fenotip	Protein	Vloga proteina v PP
<i>KISS1</i>	1q32	AD	d/iCPP, nHH	Kispeptin 1	Aktivacija HHG osi
<i>KISS1R</i>	19p13.3	AD	d/iCPP, nHH	Receptor za kispeptin 1	
<i>MKRN3</i>	15q11.2	AD, maternalno vtisnjen gen	d/iCPP	"Makorin ring finger protein 3"	Inhibicija faktorjev, ki stimulirajo izločanje GnRH
<i>DLK1</i>	14q32.2	AD, maternalno vtisnjen gen	dCPP	"Delta like non-canonical Notch ligand "	Uravnavanje časovnega poteka pubertete

Preglednica IV. Lastnosti kandidatnih genov in proteinov ki jih ti kandidatni geni kodirajo in so povezani z družinsko in idiopatsko CPP.

Gen	Lega	Protein	Vloga proteina v CPP
<i>TAC3</i>	12q13.3	Nevrokinin B	Uravnavanje izločanja GnRH
<i>TAC3R</i>	4q24	Receptor nevrokinina B	
<i>TTF1</i>	9q34.13	Tiroidni transkripcijski faktor 1	Uravnavanje izražanja genov, ki so vključeni v stimulacijo in inhibicijo nevronske mreže GnRH
<i>EAP1</i>	14q24.3	Protein s povečanim izražanjem v puberteti	
<i>GABRA1</i>	5q34	$\alpha 1$ podenota γ -aminobutričnega kislinskega receptorja tipa A	Inhibitorni vpliv na GnRH nevrone
<i>NPY-Y1R</i>	4q32.2	Receptor podtipa Y1 neuropeptida Y	
<i>LIN28B</i>	6q16.3-q21	Homologni protein lin 28	Časovno-specifično izražena miRNA, ki nadzoruje časovni potek razvoja
<i>GNRHI</i>	8p21.2	Gonadoliberin	Delovanje na hipofizno izločanje gonadotropinov
<i>GNRHR</i>	4q13.2	Receptor gonadoliberina	

1.1.7 ZDRAVLJENJE

Glavna razloga za zdravljenje CPP sta ohranitev potenciala rasti do odrasle dobe in preprečitev psihosocialnih težav, ki lahko nastopijo kot posledica prezgodnjega pubertetnega razvoja. Deklice, ki so začele z zdravljenjem pred 6. letom starosti so v odrasli dobi višje v primerjavi s tistimi zdravljenimi po 8. letu starosti (41). Ključen je pravočasni začetek zdravljenja za doseg uspešnega izida. Odločitev je odvisna od posameznika glede na natančno klinično presojo in opazovalno dobo vsaj 6-12 mesecev, s katero se določi stopnjo napredovanja CPP (11).

Najpogosteje uporabljeni zdravili sta leuprolin acetat in triptorelin acetat. Gre za GnRH analoga, katere so razvili s kemijsko modifikacijo GnRH molekule, zato da so dosegli podaljšano zasedenost hipofiznega receptorja. Neprekinjena stimulacija hipofiznega GnRH receptorja izzove manjšo občutljivost receptorja na GnRH pulze, kar se kaže z manjšim izločanjem hipofiznih gonadotropinov. Zdravljenje ne odstrani že prisotnih pubertetnih sprememb, ampak le zavre napredovanje pubertetnega razvoja (11, 24, 41, 42).

Obstaja več različnih GnRH analogov z različnimi načini in pogostostjo uporabe. V Sloveniji je v uporabi triptorelin, ki se aplicira intramuskularno 1-krat na 3 mesece.

1.2 GENOMSKA DNA IN TEHNOLOGIJA ANALIZE DNA S SEKVENCIRANJEM NASLEDNJE GENERACIJE

Določitev nukleotidnega zaporedja človeškega genoma je ključna za iskanje vzroka nastanka številnih bolezni, še posebno to velja za dedno pogojene bolezni. Funkcionalna enota genoma so geni, ki kodirajo proteine. Kljub temu, da človeški genom sestavlja 19.000 genov, le-ti predstavljajo zgolj 1 – 2 % celotnega genoma. Preostali del genoma pripada nekodirajočemu delu (43, 44). Slednje veliko večje območje je posvečeno regulaciji genov - določa kdaj in kje se bodo geni izražali (44). Veliko genetskih sprememb, ki jih povezujemo s pogostimi boleznimi, se nahaja v območjih genoma izven kodirajoče regije, zato je pomembno, da se pri iskanju vzročne genetske spremembe ne omejimo zgolj na kodirajočo regijo (44). Podatkov, ki jih dobimo z določitvijo zaporedja celotnega genoma, je za obdelavo veliko, zato težko analiziramo vse naenkrat. V nadaljevanju poglavja bom predstavila, načine za obdelavo rezultatov pridobljenih z metodo določitve zaporedja celotnega genoma (angl. *Whole genome sequencing* – WGS). Prav tako bodo predstavljene in opisane vrste genetskih sprememb, ki jih s pristopom WGS lahko opredelimo. Poznavanje genetskih sprememb in njihove funkcijske vloge je pomembno za t. i. napovedne študije, s katerimi se lahko predvidi, na kakšen način bo določena genetska sprememba vplivala na fenotip in s tem predstavljala vzrok za določeno bolezen.

1.2.1 OSNOVE SEKVENCIRANJA NASLEDNJE GENERACIJE (angl. *NEXT GENERATION SEQUENCING* – NGS)

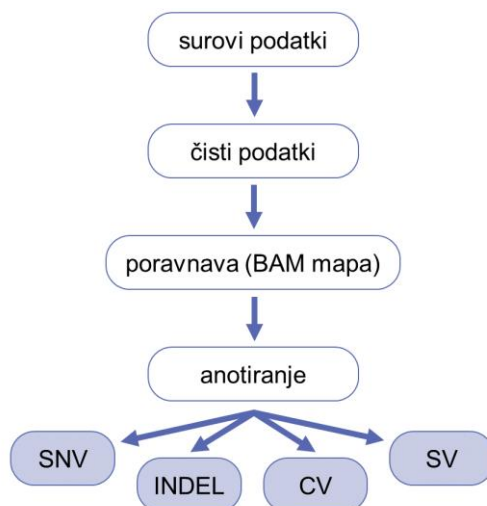
Sekvenciranje naslednje generacije je tehnologija, s katero se določi zaporedje DNA na masivno-paralelni način (45). Namesto pomnožitve in sekvenciranja ene cele molekule DNA naenkrat, je osnovni princip NGS najprej fragmentacija DNA in hkratno pomnoževanje posameznih fragmentov na pretočni celici, šele nato sledi sekvenciranje (46). Fragmenti se na pretočni celici pomnožijo vzporedno (paralelno) v masovnih količinah – naenkrat nastane več milijonov, tudi bilijonov kopij (45). Na voljo je več različic te tehnologije, ki so jih razvili proizvajalci: Illumina (San Diego, CA, ZDA), Roche (Basel, Švica) in Life Technologies (San Diego, CA, ZDA) (46). Različice NGS platform (in pripadajočih aparatov) se med seboj razlikujejo v vrstah analiz in uporabljenih kemijskih principih sekvenciranja in posledično delovnih protokolih ter v količini

dobljenih podatkov (45, 46). Čeprav obstaja več različic, je osnovni princip pri vseh enak in zajema tri glavne korake: pripravo knjižnice DNA, pomnoževanje in sekvenciranje (47).

Ena izmed NGS različic je metoda WGS, s katero se določi nukleotidno zaporedje celotnega genoma. Glavni koraki WGS s principi tehnologije, ki jo je razvilo podjetje Illumina, so opisani v poglavju Metode.

1.2.2 BIOINFORMATSKA OBDELAVA REZULTATOV PRIDOBLENIH Z PRISTOPOM SEKVENCIRANJA CELOTNEGA GENOMA

Potek obdelave rezultatov pridobljenih z WGS je za lažjo predstavbo shematsko prikazan na Sliki 1 (povzeto po 48).



Slika 1. Shema poteka obdelave rezultatov WGS (povzeto in prirejeno po (48)).

1.2.2.1 Surovi podatki (angl. *raw data*) in čisti podatki (angl. *clean data*)

Rezultat določitve zaporedja DNA, ki ga poda sekvenator so t. i. sekvenčni odčitki (angl. *sequence reads*) določenih dolžin s pripadajočim podatkom o kvaliteti sekvence. Takšne sekvence predstavljajo surove podatke (angl. *raw reads*). Da so rezultati sekvenciranja po bioinformatiki obdelavi pravilni in dovolj zanesljivi, je potrebno surove podatke, predno jih uporabimo za interpretacijo, filtrirati. Tako dobimo čiste odčitke (angl. *clean reads*). Filtriranje poteka v bralnih parih (angl. *read pair*). Hkrati se analizirata po dva sekvenčna odčitka skupaj, del zaporedja določenega v smeri naprej (angl. *forward*) in isti del

zaporedja določenega v obratni smeri (angl. *reverse*). Pri filtraciji se upošteva naslednja pravila:

- Bralni par se zavrže, če katerakoli od sekvenc vsebuje kontaminacijo z adapterjem.
- Bralni par se zavrže, če ima katerakoli od sekvenc določenih več kot 10 % negotovih nukleotidov.
- Bralni par se zavrže, če je v katerikoli od sekvenc delež nukleotidov z nizko kvaliteto večji od 50 %.

1.2.2.2 Poravnava

Po pridobitvi čistih odčitkov, sledi korak poravnave zaporedja (angl. *sequence alignment*) z algoritmom Burrows-Wheeler Aligner (BWA) (49). Poravnava zaporedja pomeni, da se dobljeno zaporedje naloži na referenčno zaporedje. To je zaporedje genoma ki je že bilo določeno; sestavljeno je iz zaporedij 13 zdravih posameznikov, ki imajo evropske prednike in je shranjeno v podatkovnih bazah (50). S tem, ko se dobljeno zaporedje poravnava z referenčnim, se prepozna preiskovančovo zaporedje, ki se z referenčnim prekriva. Sekvenčni odčitki se dovolj točno poravnajo na referenčno zaporedje, kadar se z njim ujemajo vsaj v 95 % (48, 51). Na ta način se zaporedju določijo pozicijske koordinate na genomu.

1.2.2.3 Vrste genetskih sprememb

V rezultatih sekvenciranja so prikazane tudi prisotne vrste genetskih sprememb skupaj s pripadajoči tipi, ki se nahajajo v kodirajočih regijah genoma. Vrste in tipi genetskih sprememb so zbrani v *Preglednici V*. Obstajajo štiri skupine sprememb: eno-nukleotidne spremembe (angl. *Single nucleotide variants - SNV*); spremembe insercija-delecija (angl. *Insertion-deletion - INDEL*); strukturne spremembe (angl. *Structural variants - SV*) in spremembe v številu kopij (angl. *Copy number variants - CNV*).

1. ENONUKLEOTIDNE SPREMEMBE (SNV)

V človeškem genomu je okoli 3,6 milijonov enonukleotidnih sprememb. Nad 95 % SNV, ki imajo visoko pojavnost v populaciji (alelna frekvenca v populaciji je nad 5 %), je opisanih v podatkovni bazi polimorfizmov (dbSNP, ExAc) (52, 53). Tipi SNV, ki se nahajajo v kodirajočih regijah so:

- Sinonimna enonukleotidna sprememba (angl. *synonymous SNV*). Sprememba enega nukleotida, ki ne spremeni aminokislinae.
- Drugačnosmiselna enonukleotidna sprememba (angl. *missense SNV*). Sprememba enega nukleotida, ki povzroči spremembo aminokislinae.
- Nesmiselna enonukleotidna sprememba (angl. *nonsense SNV*). To je ne sinonimna sprememba enega nukleotida, ki povzroči nastanek zgodnjega terminacijskega kodona na mestu spremembe.
- Izguba terminacijskega kodona (angl. *stoploss SNV*). Ne sinonimna sprememba enega nukleotida, ki vodi v izgubo terminacijskega kodona na mestu spremembe.
- Nepoznana enonukleotidna sprememba (angl. *unknown SNV*). V podatkovnih bazah o tej spremembi še ni znano, kakšen je vpliv spremenjenega nukleotida.

2. INSERCIJA-DELECIJA (INDEL)

V človeškem genomu je okoli 350.000 sprememb vrste insercija-delecija. Gre za vstavitve (insercije) ali/in izgubo (delecijo) zaporedja dolgega manj kot 50 bp. Sem spadajo manjše insercije, manjše delecije ter spremembe, kjer sočasno pride do insercije in delecije.

Posledica je lahko sprememba aminokislinae in sicer, kadar pride do insercije-delecije v kodirajočih regijah gena ali v spojitvenih območjih gena (angl. *splicing site*).

Spojitvena območja so zaporedja gena, ki se prepisujejo v pre-mRNA, in se nato po izrezu intronov, spojijo skupaj, da nastane zrela mRNA. V kodirajočih regijah obstaja več tipov sprememb insercija-delecija (glej *Preglednico V*). Kadar število vstavljenih ali/in izgubljenih baz ni večkratnik števila 3, lahko takšna insercija-delecija spremeni bralni okvir (angl. *frameshift INDEL – fs_ins/fs_del*), kar ima za posledico spremembo vseh nadaljnjih aminokislinae do ustavitve branja s prezgodnjim terminacijskim kodonom.

3. STRUKTURNE SPREMEMBE (SV)

O strukturalni spremembi v genomu govorimo takrat, kadar pride do izgube, vstavitve, podvojitve, obrata ali premestitve večjega dela zaporedja, dolžine med 1 000 in 3 milijoni baz. Strukturne spremembe so posledica kombinacije in so-vpliva bolezni ter variabilnosti znotraj posameznikov (48).

4. SPREMEMBE V ŠTEVILU KOPIJ (CNV)

Spremembe v številu kopij so tiste spremembe, ki povzročijo povečanje ali zmanjšanje števila kopij večjih fragmentov v genomu. Obstajata dva tipa takšnih sprememb; izguba števila kopij ali pridobitev števila kopij dela zaporedja. Spremenjeno število kopij določenih delov zaporedja je lahko vzrok za številne bolezni (48).

Preglednica V. Vrste genetskih sprememb in pripadajoči tipi v kodirajočih regijah genoma. (SNV – eno-nukleotidne spremembe, INDEL – spremembe insercija-delecija, SV – strukturne spremembe, CNV – spremembe v številu kopij, fs_ins/fs_del – insercija ali delecija baznih parov povzroči spremembo celotnega bralnega okvirja, non-fs_ins/non-fs_del – insercija ali delecija baznih parov ne povzroči spremembe celotnega bralnega okvirja).

Vrsta genetske spremembe	Tipi genetskih sprememb v kodirajočih regijah
SNV	sinonimne, drugačnosmiselne, pridobitev terminacijskega kodona, izguba terminacijskega kodona, nepoznane
INDEL	fs_ins/fs_del, non-fs_ins/non-fs_del, pridobitev terminacijskega kodona, izguba terminacijskega kodona, nepoznane
SV	izguba, vstavev, podvojitvev, obrat, premestitev dela zaporedja dolžine med 1 000 in 3 milijoni baz
CNV	izguba števila kopij delov zaporedja, pridobitev števila kopij delov zaporedja

1.2.2.4 Anotiranje

Po določitvi zaporedja celotnega genoma lahko dobimo več kot 4 milijone različnih genetskih sprememb (54). Da bi v tej množici našli tiste spremembe, ki so funkcionalno pomembne, so razvili računalniško orodje ANNOVAR (49). Orodje uporabi podatke iz podatkovnih baz, s katerimi anotira spremembe. Spremembam pripiše opise, ki so o njih zapisane v podatkovnih bazah (npr. v dbSNP, The 1000 Genomes Project in ostalih objavljenih bazah podatkov). Vsaka sprememba je torej po anotiranju opisana, te zbrane opise imenujemo genske anotacije. V njih najdemo podatke: o kromosomski regiji, v kateri se nahaja sprememba in o genski strukturi spremembe; klinično pomembnih spremembah in njihovi alelni frekvenci v populaciji; povejo nam ali sprememba vpliva na funkcijo proteina; izvemo osnovne informacije o spremembi (kakšna je globina pokritosti, tip DNA baze pred in po mutaciji, podatek o genotipu itd.). S tako obdelanimi podatki lahko nadalje iščemo kandidatne vzročne genetske spremembe (49).

1.2.2.5 Zagotavljanje zanesljivosti rezultatov WGS

Kvaliteto določitve zaporedja DNA spremljamo s Qphred vrednostjo, ki podaja kvaliteto baze (angl. *base quality*) določene v zaporedju in stopnjo napake zaporedja (angl. *sequence error rate - e*). Stopnjo napake zaporedja določimo na različne načine: dobljeno zaporedje lahko primerjamo z referenčnim in iščemo razlike med njima (pri tem računalnik upošteva možnost genetske spremembe); s pomočjo matematičnih modelov izračunamo verjetnost napake za vsako mesto v zaporedju ali pa z uporabo računalniških orodij ocenimo napako (55, 56). Najpogosteje za določitev stopnje napake zaporedja k vzorcu dodamo manjšo količino DNA virusa PhiX, ki služi kot referenca (57). Na podlagi pridobljene sekvence se določi stopnja napake.

Qphred izračunamo po naslednji enačbi (48):

Enačba 1: Izračun vrednosti Qphred:
$$Qphred = -10 \log_{10}(e)$$

Večja kot je stopnja napake nižja je Qphred vrednost. Razmerje med Qphred in e-vrednostjo prikazuje Preglednica VI.

Preglednica VI. Razmerje med Qphred vrednostjo in e-vrednostjo.

Qphred	stopnja napake zaporedja - e	stopnja pravilnosti zaporedja	Q-ocena
10	1/10	90 %	Q10
20	1/100	99 %	Q20
30	1/1.000	99,9 %	Q30
40	1/10.000	99,99 %	Q40

Sekvenator, kemija reaktantov in kvaliteta vzorca vplivajo na napako zaporedja in kvaliteto baze. Za porazdelitev napake zaporedja v WGS analizi veljata naslednji dve lastnosti:

- Stopnja napake se veča z večanjem števila sekvenčnih branj. Razlog je v slabem fluorescenčnem signalu fluorescenčnega označevalca, ki se po detekciji iz zaporedja ne odstrani v celoti.
- Prvih nekaj baz ima višjo stopnjo napake v primerjavi z ostalimi. Na začetku senzor sekvenatorja še ni dovolj občutljiv, da bi z gotovostjo lahko določil vrsto baze, zato je kvaliteta določanja na začetku nižja.

Za zanesljivo analizo je pri vseh bazah zahtevana kvaliteta večja od Q20, vsaj 80 % baz mora imeti kvaliteto Q30 ali večjo. Povprečna stopnja napake mora biti manjša od 0,1 %.

2 NAMEN DELA

V zadnjih 50. letih opazamo, da se spolni razvoj začenja vse bolj zgodaj (4,5). Družinska CPP je podedovana oblika prezgodnje pubertete. O njej govorimo, kadar je v družini več kot en član, ki je doživel prezgodnji pubertetni razvoj (18, 19, 20). Magistrska naloga je del mednarodnega projekta evropskega združenja za pediatrično endokrinologijo (ESPE Research Unit Grant), ki je usmerjen v opredelitev novih genetskih vzrokov družinske CPP.

Namen magistrske naloge je odkriti genetske spremembe v genomih članov štirih slovenskih družin s klinično potrjeno prisotno prezgodnjo puberteto in z nadaljnjo bioinformatično analizo opredeliti genetske vzroke bolezni.

Namen dela bomo dosegli preko naslednjih specifičnih ciljev:

- Izbranim preiskovancem iz štirih družin bomo z metodo sekvenciranja celotnega genoma določili nukleotidno zaporedje celotnega genoma.
- Z uporabo ustreznih analitskih orodij in algoritmov bomo izbrali kandidatne genetske vzroke za prezgodnjo puberteto. Osredotočili se bomo na potencialne vzročne gene za družinsko CPP, ki se deduje po materi.
- Kandidatne genetske spremembe bomo potrjevali s potrditveno metodo, sekvenciranjem po Sangerju.

S tem želimo potrditi naslednji hipotezi:

1. S sekvenciranjem celotnega genoma je mogoče opredeliti genetske vzroke družinske CPP.
2. V skupini slovenskih bolnikov so prisotni novi še neopisani genetski vzroki družinske CPP.

3 PREISKOVANCI IN METODE

3.1 PREISKOVANCI

V raziskavo smo vključili štiri slovenske družine, pri katerih je bila prisotnost prezgodnje pubertete klinično potrjena. Komisija za medicinsko etiko Republike Slovenije je predlagano raziskavo odobrila (številka odobritve: 29/06/14). Vsi preiskovanci so bili seznanjeni z namenom in potekom raziskave in so podpisali pisno privolitev.

V vsaki družini sta bili za prezgodnjo puberteto diagnosticirani mama in hčerka, oče je bil zdrav. Omenjene družine imajo več kot enega člana, ki je doživel prezgodnji pubertetni razvoj, zato obstaja velika verjetnost, da je prezgodnja puberteta v teh družinah dedne narave. Tri hčerke so bile zdravljene na Pediatrični kliniki Univerzitetnega kliničnega centra v Ljubljani, hčerka družine 2 zaradi prepozne obravnave ni bila zdravljena, prav tako niso bile zdravljene mame.

Pri vseh preiskovancih s prezgodnjo puberteto je bila predhodno že izključena prisotnost vzročne spremembe v genu *MKRN3*. Z metodo WGS je bilo obema staršema in štirim hčerkam iz štirih družin določeno nukleotidno zaporedje celotnega genoma. Potrditveno sekvenciranje po Sangerju smo izvedli na mami, očetu in dveh hčerkah družine 4 s prisotno prezgodnjo puberteto.

3.2 METODE

3.2.1 IZOLACIJA GENOMSKE DNA

Reagenti in raztopine

- Komercialno dostopen reagenčni komplet FlexiGene DNA Kit 250 (Qiagen), 100 % izopropanol, 70 % etanol.

Material in instrumenti

- Sterilne mikrocentrifugirne epruvete (1,5 mL), centrifugirke (50 mL), pipete (1000 µL, 100 µL, 10 µL), sterilni nastavki za pipete.
- Vodna kopel, centrifuga, vibracijski mešalnik.

Delovni postopek

Genomsko DNA (gDNA) smo izolirali iz periferne krvi preiskovancev z uporabo reagenčnega kompleta FlexiGene DNA Kit 250 (Qiagen). Delovni postopek smo izvedli po predpisanih navodilih proizvajalca kompleta (39).

3.2.2. SPEKTROFOTOMETRIČNO PREVERJANJE ČISTOSTI IZOLIRANE DNA

Reagenti in raztopine

- Destilirana voda, vzorec: izolirana gDNA.

Material in instrumenti

- Sterilne mikrocentrifugirne epruvete (1,5 mL), pipete (1000 μ L, 100 μ L, 10 μ L), sterilni nastavki za pipete, mikrovarčna kiveta (10 mm).
- Spektrofotometer (Lambda 25, Perkin Elmer, Massachusetts, ZDA).

Delovni postopek

Z aparatom Lambda 25 (Perkin Elmer, Massachusetts, ZDA) smo spektrofotometrično preverili čistost izolirane gDNA. Kiveto z raztopljeno gDNA je aparat presvetlil z ultravijolično svetlobo (UV), pri čemer so DNA baze in morebitno prisotni proteini absorbirali UV. DNA baze absorbirajo UV pri valovni dolžini 260 nm (A_{260}), a če so prisotni proteini opazimo absorbcijo tudi pri valovni dolžini 280 nm (A_{280}). Za pripravo knjižnic DNA smo uporabili vzorce gDNA, ki so imeli razmerje A_{260}/A_{280} med 1,8 in 2,0.

3.2.3 AGAROZNA ELEKTROFOREZA

Z agarozno elektroforezo smo ločili odseke gDNA in določili njihovo dolžino. Odseki nukleinskih kislin so nabiti negativno, zato so v električnem polju potovali proti pozitivno nabiti elektrodi, pri čemer je bila hitrost potovanja odvisna od velikosti odseka. Na ta način smo pred pripravo knjižnice za WGS preverili razgrajenost in kvaliteto izolirane gDNA ter prisotnost morebitne proteinske kontaminacije.

Reagenti in raztopine

- Agaroz (Sigma-Aldrich, St. Louis; Misuri, ZDA), 10x TBE pufer (1mol/L Tris HCl, 650 mmol/L borna kislina, 20 mmol/L EDTA, destilirana voda), 1x TBE pufer, barvilo SYBR Safe DNA Gel Stain (Invitrogen, Life Technologies, San Diego, CA, ZDA), označevalec za velikost odsekov DNA, 50 bp (Sigma – Aldrich, St. Louis, Misuri, ZDA), vzorec: izolirana gDNA.

Material in instrumenti

- Erlenmajerica (250 mL), merilni valj (100 mL), urno steklo, parafilm, pipeta (10 μ L) in nastavki za pipeto.
- Tehnica, mikrovalovna pečica, sistem za agarozno elektroforezo: nosilec in glavniček za gel, komora za opazovanje gela G:BOX.

Delovni postopek

Za pripravo 1 % agaroznega gela smo v erlenmajerico (250 mL) zatehtali 0,8 g agaroze, z merilnim valjem odmerili 80 mL 1x TBE pufru in dodali 8 μ L barvila SYBR Safe DNA Gel Stain. Stehtali smo vsebino erlenmajerice in si zapisali njeno težo. Nato smo erlenmajerico postavili v mikrovalovno pečico in segrevali dvakrat po 1 min (oz. dokler se niso raztopili vsi delčki agaroze). Ponovno smo stehtali vsebino erlenmajerice in dolili toliko destilirane vode, da je teža bila enaka tisti pred segrevanjem. Ko se je mešanica v erlenmajerici ohladila na 60 °C, smo še tekočo zlili v model z glavničkom in počakali, da se je strdila. Tako smo pripravili gel. V žepke gela smo vnesli 10 μ L označevalca velikosti odsekov in 7 μ L vzorca s produkti reakcije PCR. Elektroforezna ločba je tekla pri napetosti 90 V in je trajala 20 min. Gel z ločenimi odseki gDNA smo postavili v komoro za opazovanje gela G:BOX in ga s programom GeneSnap presvetlili z UV svetlobo, da smo dobili sliko ločenih odsekov gDNA.

3.2.4. FLUORIMETRIČNO MERJENJE KONCENTRACIJE DNA

Reagenti in raztopine

- Komercialno dostopen reagenčni komplet Qubit dsDNA BR (Life Technologies, San Diego, CA, ZDA), vzorec: izolirana gDNA.

Material in instrumenti

- Sterilne mikrocentrifugirne epruvete (1,5 mL), čiste in tanke PCR-mikrocentrifugirne epruvete (0,5 mL, Invitrogen, Life Technologies, San Diego, CA, ZDA), pipete (200 μ L, 10 μ L) in sterilni nastavki za pipete.
- Vibracijski mešalnik, mikrocentrifuga, fluorimeter Qubit 2.0 (Life Technologies, San Diego, CA, ZDA).

Delovni postopek

Koncentracijo izolirane dvojnovijačne (ds) gDNA smo fluorimetrično izmerili z uporabo reagenčnega kompleta Qubit dsDNA BR (Life Technologies) in fluorimetra Qubit 2.0 (Life Technologies). Rastopine smo pripravili po proizvajalčevih navodilih, prav tako smo po navodilih izmerili koncentracijo gDNA. Tekom obdelave vzorca z reagenčnim kompletom, se je med baze gDNA vrnilo interkalirajoče fluorescenčno barvilo. Večja kot je bila količina gDNA, več barvila se je vrnilo. Aparat je izmeril fluorescenco barvila, pri čemer je bila intenziteta fluorescence sorazmerna s koncentracijo ds gDNA v vzorcu. Za pripravo knjižnic DNA smo uporabili vsaj 1 μ g gDNA na vzorec (58).

3.2.5 PRIPRAVA DNA KNJIŽNICE IN SEKVENCIRANJE CELOTNEGA GENOMA

Pripravo DNA knjižnice in WGS so izvedli pri podjetju Novogene Co., Ltd. Poročilo o osnovni bioinformatiki analizi (kontrola kvalitete sekvenciranja, naleganje na referenčno sekvenco in identifikacija prisotnih genetskih sprememb) in o vseh delovnih postopkih, ki so jih pri tem uporabili, so priložili setu podatkov, ki smo jih prejeli za nadaljnjo analizo.

Knjižnice DNA so pripravili in sekvencirali z uporabo komercialno dostopnega reagenčnega kompleta TruSeq Nano DNA HT Sample preparation Kit (Illumina, San Diego, CA, ZDA) (59), pri čemer so se ravnali po priporočilih proizvajalca.

Z ultrazvočno napravo so najprej fragmentirali gDNA, da so nastali krajši odseki. Konce odsekov gDNA so popravili iz lepljivih v tope, jih fosforilirali in na 3' konce vezali adenine. Na konce odsekov so z reakcijo ligacije vezali z indeksi označene adapterje. Adapterji so bila kratka nukleotidna zaporedja, ki so se vezala na 5' ter 3' konce odsekov gDNA in so bili sestavljeni iz regije komplementarne začetnikom vezanih na pretočni celici, iz indeksa in vezavnega mesta za sekvenčni oligonukleotidni začetnik (48). Indeksi

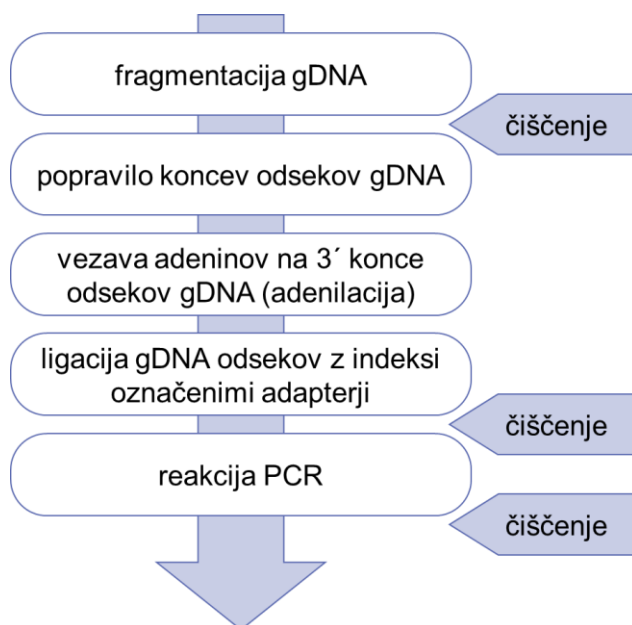
so bila edinstvena nukleotidna zaporedja dolga med 8-12 baznih parov (bp), ki so jih tekom priprave knjižnice dodali na odseke gDNA in jih na ta način označili. Sekvenator je indeksna zaporedja prepoznal, enake združil ter sestavil celotno zaporedje posamezne knjižnice DNA (46). Takšen način označevanja je omogočal združevanje ter hkratno sekvenciranje več različnih knjižnic DNA z vezanimi različnimi indeksi (46). Preko komplementarne regije so se adapterji vezani na odseke gDNA vezali na začetnike, ki so bili na površini pretočne celice v sekvenatorju.

Na koncu so odseke gDNA pomnožili z reakcijo PCR. Produkte PCR so očistili z AmpPure XP sistemom. Končne knjižnice DNA so analizirali s sistemom Agilent 2100, da so določili dolžine fragmentov v posamezni knjižnici. Končno koncentracijo posamezne DNA knjižnice so določili z metodo kvantitativnega PCR v realnem času (qPCR) (48). Rezultat priprave knjižnice DNA so bili torej odseki gDNA ligirani z indeksi označenimi adapterji (prikazuje *Slika 2*) (48).



Slika 2. Odsek gDNA ligiran z indeksi označenimi adapterji (prirejeno po (48)).

Za lažjo predstavo o poteku eksperimentalnega dela priprave knjižnice DNA za WGS smo oblikovali delovno shemo (glej *Slika 3*).



Slika 3. Delovna shema za pripravo knjižnice DNA (povzeto po 59).

3.2.5.1 Fragmentacija gDNA

Fragmentacijo gDNA so izvedli s postopkom sonifikacije, ki za rezanje ds gDNA uporabi ultrazvok. V pripravo za sonifikacijo so prenesli vzorce izolirane gDNA. Genomsko DNA so fragmentirali tako, da so sledili proizvajalčevim navodilom (59). Dobili so odseke gDNA s 3' in 5' lepljivimi konci.

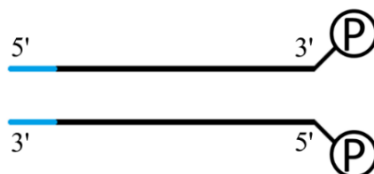
3.2.5.2 Čiščenje produktov fragmentacije gDNA, čiščenje odsekov gDNA ligiranih z indeksi označenimi adapterji in produktov reakcije PCR

Čiščenje produktov fragmentacije gDNA, čiščenje odsekov gDNA ligiranih z indeksi označenimi adapterji in čiščenje produktov reakcije PCR je potekalo po enakem delovnem postopku.

K vzorcem so dodali magnetne kroglice za čiščenje vzorca, na katere so se vezali želeni odseki in kroglice prenesli na magnetno stojalo. Ko se je tekočina v vzorcih zbistrla so odstranili in zavrgli ves supernatant. Sledilo je spiranje z etanolom. Tako so dobili očiščene vzorce.

3.2.5.3 Popravilo koncev odsekov gDNA

S popravilom koncev odsekov ds gDNA so lepljive konce, ki so jih dobili s fragmentacijo gDNA, pretvorili v tope konce. Dvojnovijačno gDNA so obdelali na naslednji način (glej *Slika 4*): encim eksonukleaza je s svojim 3' proti 5' delovanjem zapolnila 3' lepljivi konec, polimeraza pa je z 5' proti 3' aktivnostjo zapolnila 5' lepljivi konec. Na koncu je sledila še fosforilacija 3' konca ene verige in 5' konca druge verige.



Slika 4. Popravilo in fosforilacija koncev odsekov gDNA (prirejeno po (48)).

Sledilo je odstranjevanje velikih odsekov gDNA. K vzorcem so dodali magnetne kroglice za čiščenje vzorca, na katere so se vezali večji odseki gDNA in jih prenesli na magnetno stojalo. Ko se je tekočina v vzorcih zbistrila so obdržali supernatant z odstranjenimi velikimi odseki gDNA.

Odstranili so še majhne odseke gDNA. K vzorcem z odstranjenimi velikimi odseki gDNA so dodali magnetne kroglice za čiščenje vzorca, na katere so se vezali srednje veliki odseki gDNA in jih prenesli na magnetno stojalo. Ko se je tekočina v vzorcih zbistrila so odstranili ter zavrgli ves supernatant. Sledilo je spiranje z etanolom. Nato so k vzorcem dodali pufer za ponovno suspendiranje, jih ponovno prenesli na magnetno stojalo in počakali, da se je tekočina v vzorcih zbistrila. Obdržali so supernatant z odstranjenimi majhnimi odseki gDNA.

3.2.5.4 Vezava adeninov na 3' konce odsekov gDNA (adenilacija)

Na tope konce odsekov gDNA so dodali nukleotid adenin (A) (glej *Slika 5*). V naslednjem koraku je ligacija potekla tako, da se je nukleotid timin (T) na 3' koncu adapterja komplementarno vezal na dodani A nukleotid na koncih odsekov gDNA. Adenilacija je torej omogočila izvedbo reakcije ligacije z indeksi označenih adapterjev, s čimer so preprečili združevanje odsekov gDNA med reakcijo.



Slika 5. Primer ligacije odseka DNA z indeksi označenimi adapterji (prirejeno po (48)).

3.2.3.5 Ligacija z indeksi označenih adapterjev

Odseke gDNA so ligirali z indeksi označenimi adapterji (primer prikazuje *Slika 6*). Preko z indeksi označenih adapterjev so se odseki gDNA vezali na oligonukleotidna zaporedja, ki so bila vezana na površini pretočne celice v sekvenatorju.

3.2.5.6 Reakcija PCR

V tem koraku so specifično pomnožili le tiste odseke gDNA, ki so bili ligirani z indeksi označenimi adapterji na obeh koncih verig. Odseki gDNA, ki so bili ligirani z indeksi označenim adapterjem le na enem koncu verige, so kazali na neučinkovito reakcijo ligacije. Za sekvenciranje niso bili primerni, ker se na pretočni celici ne morejo pomnoževati.

K vzorcem so dodali mešanice PCR oligonukleotidnih začetnikov in mešanico za PCR. Ploščico so postavili v aparat za PCR. Vzorec gDNA so pomnožili pri naslednjih časovnih in temperaturnih pogojih:

- predgretje pokrova na 100 °C
- 3 min / 95 °C
- 8 ciklov: 20 s / 98 °C, 15 s / 60 °C, 30 s / 72 °C
- 5 min / 72 °C
- hlajenje / 4 °C

3.2.5.7 Določitev dolžine in kvantifikacija knjižnice DNA

Dolžino knjižnice DNA so določili na aparatu Agilent 2100 (60).

Na čip Agilent 2100 so nanegli gel, marker z vzorcem in lestvico ter ga vstavili v

bioanalizator. V aparatu je potekla elektroforezna ločba odsekov gDNA, pri čemer se je določila dolžina odsekov gDNA.

Knjižnico DNA so kvantificirali z metodo kvantitativnega PCR v realnem času (qPCR). Tekom reakcije PCR so se specifični oligonukleotidni začetniki vezali na indeksna zaporedja adapterjev ligiranih na odseke gDNA. Tako so lahko izmerili koncentracijo le tistih odsekov gDNA, ki so bili ligirani z indeksi označenimi adapterji. V fazi podaljševanja se je med nukleotide v zaporedju dsDNA vgradilo fluorescenčno barvilo in začelo oddajati svetlobo. Aparat za qPCR je po vsakem končanem ciklu reakcije izmeril fluorescenco oddane svetlobe, končni rezultat pa podal kot amplifikacijsko krivuljo (odvisnost intenzitete fluorescence od števila ciklov). Za vsak standard in vzorec so iz krivulje odčitali C_q vrednost (ta pove, pri katerem ciklu fluorescenca že značilno preseže fluorescenco ozadja). Nato so iz C_q vrednosti in znanih koncentracij standardov izrisali standardno krivuljo (odvisnost C_q od začetne koncentracije knjižnice DNA). Koncentracijo knjižnice DNA so odčitali iz krivulje glede na njeno C_q vrednost.

3.2.5.8 Dokončna priprava knjižnice DNA za WGS

Reagente in knjižnico DNA so predhodno pripravili z upoštevanjem vseh navodil proizvajalca, da so dobili učinkovite koncentracijske in volumske vrednosti.

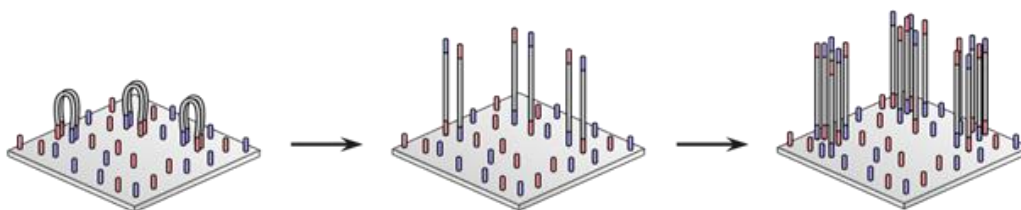
3.2.5.9 Metoda WGS

Knjižnica DNA je bila sekvencirana na aparatu HiSeq 4000 (Illumina, San Diego, CA, ZDA).

Pomnoževanje – premostitveni PCR (angl. *bridge PCR*)

Pomnoževanje je potekalo na površini pretočne celice v sekvenatorju, kjer sta bili vezani dve vrsti različnih oligonukleotidnih začetnikov. Vsaka izmed njiju je bila komplementarna enemu od adapterjev vezanih na odseke gDNA. Najprej so na pretočno celico nanесли enoverižno knjižnico DNA z vezanimi adapterji, ki so se povezali z ustreznimi oligonukleotidnimi začetniki. Prosti konci so se zvrili, da so dosegli drugega izmed vezanih začetnikov. Na ta način so se oblikovali mostovi enoverižnih odsekov gDNA, ki so imeli oba konca vezana na pretočni celici (glej *Slika 6*) (46, 47). Ko so na pretočno celico dodali reagente za PCR, so se enoverižni odseki gDNA pomnožili in zatem denaturirali. S ponavljanjem krogov pomnoževanja in denaturacije je nastalo okoli 1000

kopij vsakega odseka gDNA (46). Na koncu so ostale na pretočno celico vezane le verige, ki so potekale v smeri naprej, obratno-smerne pa so se odcepile in sprale (45).



Slika 6. Enoverižni odseki gDNA so se vezali s komplementarnimi oligonukleotidnimi začetniki, ki so bili vezani na površino pretočne celice in oblikovali mostove. Sledila je reakcija PCR in zatem denaturacija, da so nastali pomnoženi enoverižni odseki gDNA (prirejeno po (47)).

Sekvenciranje s sintezo (angl. *sequencing by synthesis – SBS*)

Ko so na pretočno celico dodali sekvenčni oligonukleotidni začetnik in se je ta vezal na adapter na odseku gDNA, se je začela sekvenčna reakcija. Sekvenciranje je potekalo v več ciklih, ki so bili sestavljeni iz treh faz. V prvi fazi so na pretočno celico dodali DNA polimerazo in posebno oblikovane nukleotide, ki so bili označeni s fluorescentnim barvilom (vsak izmed štirih nukleotidov s svojim) in so imeli vezano terminirajočo skupino (47). Encim je v novonastajajočo verigo vgradil nukleotid, nadaljnje podaljševanje verige pa je preprečila terminirajoča skupina na nukleotidu. (45). Vsi odseki gDNA so bili tako podaljšani za en nukleotid. V drugi fazi so se nevgrajeni nukleotidi sprali, vgrajeni pa so se identificirali s fluorescentno detekcijo. Po laserski osvetlitvi je fluorescentno barvilo oddalo svetlobo, kateri se je izmerila valovna dolžina značilna za posamezen nukleotid (45). V zadnji fazi cikla sta se fluorescentno barvilo in terminirajoča skupina kemično odcepila od vgrajenih nukleotidov. Cikli so se ponavljali dokler niso bila podaljšana in s tem določena celotna zaporedja odsekov gDNA (45).

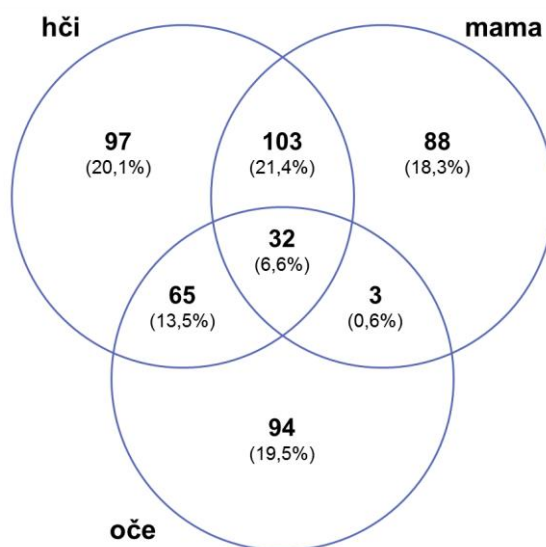
Na enak način se je pomnožilo in določilo še zaporedja obratno-smernih odsekov gDNA, ki so služili kot kontrola.

3.2.6 BIOINFORMATSKA ANALIZA OBDELANIH REZULTATOV PRIDOBLENIH S PRISTOPOM SEKVENCIRANJA CELOTNEGA GENOMA – ISKANJE VZROČNE GENETSKE SPREMEMBE V GENOMU

3.2.6.1 Potek iskanja kandidatnega gena za družinsko CPP v kodirajočih regijah genoma

S programskim orodjem Illumina VariantStudio v3.0 (Illumina, San Diego, CA, ZDA) smo obdelane rezultate pridobljene z WGS analizo filtrirali s postopki osnovne analize množic, ki jo demonstriramo z Vennovim diagramom (glej *Slika 7*). Naša naloga je bila poiskati spremembe v kodirajočih regijah genoma, ki se dedujejo iz bolnih staršev na otroka. Pri analiziranih primerih prezgodnje pubertete je bila močna indikacija maternalnega avtosomno dominantnega dedovanja. Oče je bil zdrav, mama in hči pa sta bili bolni, torej je hči podedovala okvarjen alel po mami. Iz genomov mame in hčerke smo izbrali tiste spremembe, ki so skupne obema, posledično smo skupnim spremembam mame in hčerke odstranili vse, ki so hkrati skupne zdravemu očetu. Torej smo naredili presek sprememb mame, hčerke in očeta ter preseku mame in hčerke odstranili del, ki je skupen mami, hčerki in očetu: $(MAMA \cap HČI) - (OČE \cap MAMA \cap HČI)$.

Pripravili smo shemo postopka filtriranja sprememb za posamezno družino, ki je prikazana na *Sliki 9*.



Slika 7. Primer Vennovega diagrama presekov sprememb mame, očeta in hčerke družine 4.

Spremembe smo imeli po bioinformatiki obdelavi rezultatov WGS urejene v skupine glede na to, katero vrsto genetske spremembe povzročajo. Po zgoraj opisanem premisleku, na kakšen način bomo iskali željene spremembe, smo v VariantStudio uvozili vse spremembe skupine SNV in skupine INDEL, ki pripadajo očetu, mami in hčerki. Predno smo se lotili filtriranja, smo podatke anotirali, kar pomeni, da smo vsem spremembam, ki so že zbrane in opisane v podatkovnih bazah, pripisali te opise. To je izvedlo orodje samo. Sledilo je filtriranje podatkov. V nadaljevanju so navedeni naslednji filtri, s katerimi smo zajemali spremembe glede na željene parametre:

SPREMEMBA (angl. *VARIANT*)

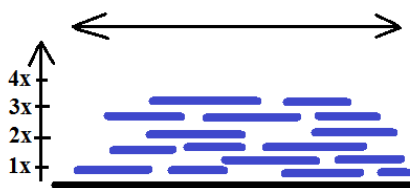
POSLEDICA (angl. *CONSEQUENCE*)

POGOSTOST V POPULACIJI (angl. *POPULATION FREQUENCY*)

DRUŽINSKA POGOJENOST (angl. *FAMILY BASED*)

1. Filter SPREMEMBA (angl. *VARIANT*)

S tem filtrom smo določili ustrezno globino pokritosti, ki jo izražamo kot 1x, 2x, 3x ... (glej *Slika 8*). Pove nam, kolikokrat v povprečju je sekvencirana vsaka baza v genomu. Izračunali smo jo po formuli: $C = N/G$ (N je število baz v vseh fragmentih, ki se nalegajo na referenčno zaporedje, G je dolžina genoma). Globina pokritosti nam pove o kvaliteti določanja zaporedja. Večja kot je, z večjo zanesljivostjo lahko trdimo, da se določena baza res nahaja na določenem mestu v zaporedju. Za klinično sliko je pomembna pokritost nad 10x.



Slika 8. Pokritost - leva puščica prikazuje globino pokritosti, zgornja puščica pa površinsko pokritost (prirejeno po (61)).

2. Filter POSLEDICA (angl. *CONSEQUENCE*)

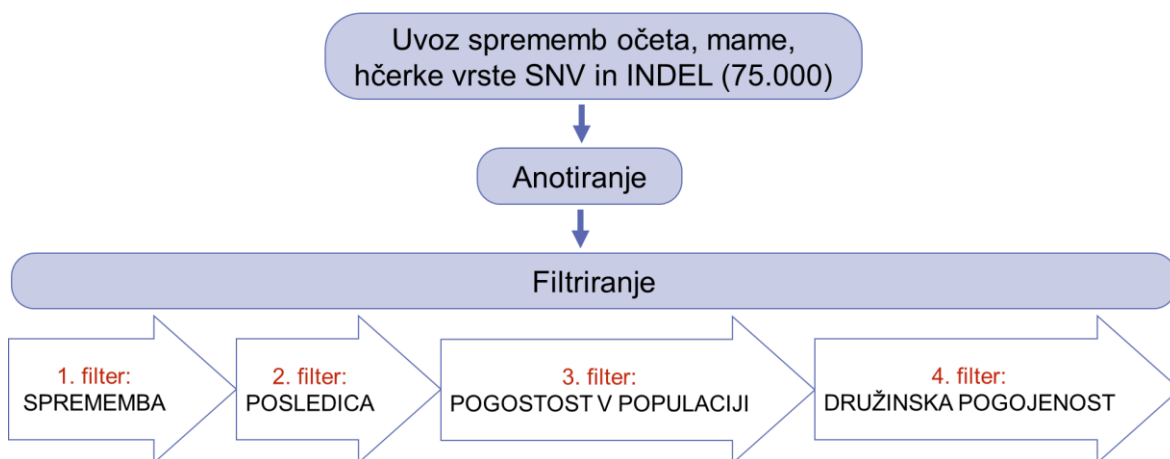
Tu imamo na voljo filtre, ki iščejo po različnih tipih genetskih sprememb. Izbrali smo tiste, ki vplivajo na funkcijo prepisane aminokisliline, torej: drugačnosmiselne, tiste ki spreminjajo bralni okvir in tiste ki kodirajo ali odstranijo terminacijski kodon.

3. Filter POGOSTOST V POPULACIJI (angl. *POPULATION FREQUENCY*)

Glede na znano pojavnost bolezni v populaciji smo določili skupno najvišjo frekvenco, s katero smo zajemali spremembe. Vemo, da je pojavnost prezgodnje pubertete 1 na 500, torej je bolne 0,2 % populacije. Če skupno najvišjo frekvenco nastavimo na 0,2 % pomeni, da zajemamo vse spremembe, ki se pojavljajo v 0,2 % in manj. Za večjo zanesljivost, da bomo res zajeli vse spremembe, ki povzročajo bolezen, smo frekvenco zvišali na 0,3 %.

4. Filter DRUŽINSKA POGOJENOST (angl. *FAMILY BASED*)

Označili smo način pridobitve spremembe oz. dedovanja, v našem primeru je to avtosomno-dominantno in označili člane družine, ki so bolni, v našem primeru mamo in hčerko. Ta filter nam je izločil vse očetove spremembe, ki so prav tako obstajale pri hčerki in mami.



Slika 9. Shema postopka filtracije sprememb pri posamezni družini s programskim orodjem Illumina VariantStudio v3.0.

Opisani postopek filtriranja smo ponovili pri vsaki družini posebej ter dobljene nabore sprememb shranili v Excelovo tabelo. S filtriranjem sprememb smo iz začetnega števila okoli 25.000 sprememb, kolikor jih je bilo prisotnih v posameznem vzorcu gDNA, prišli do skrčenega nabora okoli 100 sprememb za posamezno družino.

Sledilo je preverjanje ali v teh naborih obstajajo spremembe, ki so skupne vsem štirim družinam. Naredili smo presek naborov sprememb vseh družin: na spletu smo poiskali in odprli orodje Venny generator v2.1 (62), v ustrezna polja smo vnesli shranjeni nabor sprememb vsake družine in orodje nam je generiralo Vennov diagram.

Patogenost dobljenih skupnih sprememb smo ocenili z uporabo *in silico* napovednega algoritma Combined Annotation Dependent Depletion - CADD (63). Gre za računalniško orodje s pomočjo katerega kvantitativno ocenimo patogenost genetskih sprememb. Z uporabo algoritmov se genetskim spremembam pripiše ocene v številkah, ki so sorazmerne s patogenostjo, resnostjo bolezni in z vplivnostjo spremembe na določeno funkcijo (64). Na ta način razvrstimo spremembe, kar nam omogoči, da nadaljnjo raziskovanje usmerimo v potencialno najbolj škodljive.

Potencialno patogene spremembe, ki so skupne vsem družinam, smo eksperimentalno preverili s potrditveno metodo, in sicer s sekvenciranjem po Sangerju.

3.2.7 POTRJEVANJE S SEKVENCIRANJEM PO SANGERJU

Sekvenciranje po Sangerju je metoda za določitev nukleotidnega zaporedja izbranega odseka DNA. Gre za modificirano metodo PCR pomnoževanja DNA. Veriga in nanjo prilegajoči se oligonukleotidni začetnik se inkubirata skupaj z DNA polimerazo I v prisotnosti mešanice dNTP-jev in fluorescentno označenih ddNTP-jev (vsak izmed štirih različnih ddNTP-jev je označen z drugim fluorescentnim barvilom). Medtem ko se oligonukleotidna veriga pomnožuje, blokiramo vezavo naslednjega nukleotida, s tem da se na mesto dNTP-ja vgradi fluorescentno označen ddNTP, ki je končna baza na tej verigi. Ker ddNTP na 3' koncu nima hidroksilne skupine, se naslednji nukleotid ne more več vgraditi in pomnoževanje se zaključi na mestu vgradnje ddNTP-ja v nastajajoči DNA verigi. Po pomnožitvi imajo vsi odseki enak 5' konec in različen 3' konec, odvisno od tega na katerem mestu se je ddNTP vgradil v verigo. Odseke ločimo po velikosti s kapilarno elektroforezo, ki poteče v avtomatskem sekvenatorju. Odseki, ki se med seboj ločijo v dolžini enega nukleotida se razvrstijo po dolžini. Vsak odsek ima vgrajen končni ddNTP, ki ga zazna detektor fluorescence. Signal, ki ga aparat poda, so za štiri različna fluorescentna barvila ddNTP-jev izmerjene intenzitete oddane svetlobe pri štirih različnih valovnih dolžinah. Pri branju fluorescentnih signalov zaporednih odsekov dobimo sekvenco gDNA. Pomembno je ustrezno razmerje med ddNTP-ji in dNTP-ji, zato da dobimo primerno razporeditev različno dolgih odsekov, iz katerih določimo zaporedje. Za ddNTP derivate je to razmerje okoli 1:100 (65). Za zanesljivejšo določitev zaporedja je pomembno, da sekvenciranje izvedemo v obe smeri ločeno; v smeri naprej (angl. Forward) in v obratni smeri (angl. Reverse) z uporabo smernega (F) in obratno smernega

(R) oligonukleotidnega začetnika (65). Pravilnost določitve preverimo s skladnostjo dobljenih sekvenc. Na ta način lahko določimo zaporedje dolgo do okoli 1000 nukleotidov (66). Občasno se lahko pojavijo artefakti, ki so običajno posledica napak zaradi kemije sekvenčne reakcije oz. neustreznih oligonukleotidnih začetnikov (67).

Potrditveno metodo smo izvedli na družini 4. Za zanesljivo potrjevanje smo uporabili vzorce gDNA mame in očeta ter vzorca gDNA obeh bolnih hčerk. S Sangerjevo metodo sekvenciranja smo vzorcem gDNA določili nukleotidno zaporedje eksona 5 gena *ZNF717*.

3.2.7.1 Oblikovanje oligonukleotidnih začetnikov

V podatkovni bazi genomov ENSEMBL (68) smo za gen *ZNF717* poiskali referenčno zaporedje. Analizirana družina je imela genetsko spremembo opredeljeno v eksonu 5 tega gena, zato smo za oblikovanje smernih in protismernih oligonukleotidnih začetnikov v spletnem programu Primer 3 (v.0.4.0) uporabili zaporedje gena, ki je vključevalo ta del. S spletnim orodjem SNPCheck (69) smo v predlaganih oligonukleotidnih začetnikih preverili morebitno prisotnost polimorfizma posameznih nukleotidov (SNP). Izbrali smo takšen par oligonukleotidnih začetnikov, ki ni vseboval SNP-jev. Izbrana oligonukleotidna začetnika sta prikazana v *Preglednici VII*.

Preglednica VII. Oligonukleotidna začetnika z opisanimi lastnostmi, ki smo ju uporabili za potrditveno metodo.

Oligonukleotidni začetnik	Nukleotidno zaporedje 5' - 3'	Dolžina produktov (bp)	T _m - Temperatura prileganja (C°)
ZNF717_ex5_a_F	CCAGCCTCATTATCCATCAG	635	58.12
ZNF717_ex5_a_R	GTATGGTTTTTCCCAGCAT		58.76

3.2.7.2 Reakcija PCR

Reagenti in raztopine

- Komercialno dostopna reakcijska mešanica za reakcijo PCR Go Taq G2 Green Master Mix (Promega, San Luis Obispo, CA, ZDA), raztopini oligonukleotidnih začetnikov (F in R) s koncentracijo 100 µmol/L, voda primerna za reakcijo PCR vzorec: gDNA s koncentracijo 100 ng/µL

Material in instrumenti

- Reakcijske epruvete (200 μ L) in pokrovčki zanje, nosilec za vzorce s podstavkom, pipete (2 μ L, 10 μ L) in sterilni nastavki za pipete.
- Vibracijski mešalnik, centrifuga, aparat za reakcijo PCR ABI 9700 (Applied Biosystems, Foster City, CA, ZDA).

Delovni postopek

V reakcijsko epruveto (200 μ L) smo dali 12,5 μ L komercialno dostopne reakcijske mešanice za reakcijo PCR Go Taq G2 Green Master Mix (Promega, San Luis Obispo, CA, ZDA), 1 μ L raztopine oligonukleotidnih začetnikov F s koncentracijo 100 μ mol/L in 1 μ L raztopine oligonukleotidnih začetnikov R s koncentracijo 100 μ mol/L, 1 μ L vzorca gDNA s koncentracijo 100 mg/ μ L in 9,5 μ L vode primerne za reakcijo PCR.

Reagenčno mešanico smo premešali, centrifugirali 5 sekund ter jo prenesli v aparat za reakcijo PCR - PCR ABI 9700 (Applied Biosystems, Foster City, CA, ZDA). Vzorec gDNA smo pomnožili pri naslednjih časovnih in temperaturnih pogojih:

→ 2 min / 95 °C

→ 35 ciklov: 30 s / 95 °C, 30 s / 61 °C, 30 s / 72 °C

→ 4 min / 72 °C

→ hlajenje / 4 °C

3.2.7.3 Agarozna elektroforeza

Na agarozni elektroforezi smo ločili pomnožene odseke gDNA in jim določili dolžino. Na ta način smo preverili uspešnost pomnožitve odsekov gDNA. Odseki nukleinskih kislin so nabiti negativno, zato so v električnem polju potovali proti pozitivno nabiti elektrodi, pri čemer je bila hitrost potovanja odvisna od dolžine odseka.

Reagenti in raztopine

- Agaroz (Sigma-Aldrich, St. Louis; Misuri, ZDA), 10x TBE pufer (1mol/L Tris HCl, 650 mmol/L borna kislina, 20 mmol/L EDTA, destilirana voda), 1x TBE pufer, barvilo SYBR Safe DNA Gel Stain (Invitrogen, Life Technologies, San Diego, CA, ZDA),

označevalec za velikost odsekov DNA, 50 bp (Sigma – Aldrich, St. Louis, Misuri, ZDA), vzorec: produkti gDNA, ki jih dobimo z reakcijo PCR.

Material in instrumenti

- Erlenmajerica (250 mL), merilni valj (100 mL), urno steklo, parafilm, pipeta (10 μ L) in nastavki za pipeto.
- Tehnica, mikrovalovna pečica, sistem za agarozno elektroforezo: nosilec in glavniček za gel, komora za opazovanje gela G:BOX (Syngene).

Delovni postopek

Za pripravo 2 % agaroznega gela smo v erlenmajerico (250 mL) zatehtali 1,6 g agaroze, z merilnim valjem smo odmerili 80 mL 1x TBE pufra in dodali 8 μ L barvila SYBR Safe DNA Gel Stain. Stehtali smo vsebino erlenmajerice in si zapomnili njeno težo. Nato smo erlenmajerico postavili v mikrovalovno pečico in segrevali dvakrat po 1 min (oz. dokler se niso raztopili vsi delčki agaroze). Ponovno smo stehtali vsebino erlenmajerice in dolili toliko destilirane vode, da je teža bila enaka tisti pred segrevanjem. Ko se je mešanica v erlenmajerici ohladila na 60 °C, smo še tekočo zlili v model z glavničkom in počakali, da se je strdila. Tako smo pripravili gel. V žepke gela smo vnesli 10 μ L označevalca velikosti odsekov in 7 μ L vzorca s produkti reakcije PCR. Elektroforezna ločba je tekla pri napetosti 90 V in je trajala 20 min. Gel z ločenimi odseki gDNA smo postavili v komoro za opazovanje gela G:BOX in ga s programom GeneSnap presvetlili z UV svetlobo, da smo dobili sliko ločenih odsekov gDNA.

3.2.7.4 Sekvenčna reakcija

Reagenti in raztopine

- Raztopini oligonukleotidnih začetnikov (F in R) s koncentracijo 100 μ mol/L, mešanica encimov ExoSAP-IT (Affymetrix, Santa Clara, CA, ZDA), Big-Dye Terminator (Applied Biosystems, Foster City, CA, ZDA), natrijev acetat, etanol (100 % in 70 %), formamid, voda primerna za reakcijo PCR, vzorec: produkti reakcije PCR.

Material in instrumenti

- Sterilne mikrocentrifugirne epruvete (1,5 mL), pipete (2 μ L, 10 μ L, 20 μ L, 100 μ L, 1000 μ L) in sterilni nastavki za pipete, reakcijska ploščica (96 vdolbinic) MicroAmp® Fast Optical 96-Well Reaction Plate (Applied Biosystems, Foster City, CA, ZDA).
- Komora za sterilno delo, vibracijski mešalnik, centrifuga, aparat za reakcijo PCR GeneAmp PCR System 9700 (Applied biosystems, Foster City, CA, ZDA), sekvenator ABI 3500 Genetic Analyzer (Applied Biosystems, Foster City, CA, ZDA).

Delovni postopek

V epruveto smo k 2,5 μ L vzorca s produkti reakcije PCR dodali 1 μ L ExoSAP-IT mešanice (Affymetrix, Santa Clara, CA, ZDA), ki je vsebovala dva hidrolitična encima (eksonukleazo 1 in alkalno fosfatazo). S hidrolitičnim delovanjem encimov smo se znebili ne vgrajenih dNTP-jev, ne vgrajenih oligonukleotidnih začetnikov in ostalih ne željenih eno-verižnih produktov. Po tem, ko smo epruveto premešali in centrifugirali 5 sekund, smo epruveto postavili v aparat za reakcijo PCR ABI 9700 (Applied Biosystems, Foster City, CA, ZDA). Sledila je inkubacija na 37 °C 15 min in nato še inkubacija na 80 °C 15 min, s čimer smo ustavili delovanje encimov. Po inkubaciji je bila na vrsti sekvenčna reakcija. Izvedli smo jo v obe smeri ločeno; v smeri naprej (angl. Forward – F) in v obratni smeri (angl. Reverse – R) z uporabo oligonukleotidnih začetnikov F ali R. K očiščenim produktom reakcije PCR smo dodali 16,2 μ L mešanice za sekvenciranje BigDye Terminator (BigDye Reagent in 5x pufer za sekvenciranje) (Applied Biosystems, Foster City, CA, ZDA) in 0,3 μ L raztopine oligonukleotidnega začetnika F ali 0,3 μ L raztopine oligonukleotidnega začetnika R. Sledila je sekvenčna reakcija na aparatu za reakcijo PCR ABI 9700 (Applied Biosystems, Foster City, CA, ZDA) pri naslednjih časovnih in temperaturnih pogojih:

→ 1 min / 96 °C

→ 25 ciklov: 10 s / 96 °C, 5 s / 50 °C , 4 min / 60 °C

→ hlajenje / 4 °C

Po končani sekvenčni reakciji smo odstranili odvečne ne vgrajene ddNTP-je in oligonukleotidne začetnike tako, da smo k produktom sekvenčne reakcije dodali 2 μ L 3 mol/L natrijevega acetata in 50 μ L 100 % etanola. Epruveto smo premešali, jo inkubirali

na ledu 10 min ter centrifugirali pri 1500 g na 4 °C 45 min.

Po 45 min smo iz epruvete odstranili ves supernatant, dodali 70 µL 70 % etanola, premešali in centrifugirali pri 1500 g 15 min. Ponovno smo po centrifugiranju odstranili supernatant. V epruvetah je ostal oborjeni produkt sekvenčne reakcije, ki smo ga 10 min sušili na sobni temperaturi. Sledila je priprava oborjenih vzorcev za kapilarno elektroforezo.

Oborjene vzorce smo raztopili v 20 µL formamida in inkubirali na 95 °C 5 min. Po ohlavitvi na 10 °C smo pripravljene vzorce vstavili v sekvenator ABI 3500 Genetic Analyzer (applied Biosystem, Foster City, CA, ZDA). Rezultat je bil določeno nukleotidno zaporedje izbranega odseka gDNA.

3.2.7.5 Analiza rezultatov dobljenih s sekvenciranjem po Sangerju

Nukleotidno zaporedje gDNA je sekvenator določil do baze natančno. Prikazal ga je kot elektroferogram, na katerem so bile izrisane krivulje absorpcijskih vrhov vseh štirih baz. Zelene obarvane krivulje so predstavljale adenine (A), črne obarvane gvanine (G), modre citozine (C) in rdeče obarvane krivulje timine (T).

Da smo v sekvenciranem odseku gDNA preverili prisotnost genetske spremembe predhodno opredeljene z WGS, smo rezultat sekvenciranja po Sangerju preverili s programskim algoritmom BLAST (70). Ta je dobljeno zaporedje gena primerjal z referenčnim zaporedjem gena, ki je shranjeno v podatkovni bazi GenBank (št. referenčnega zaporedja gena *ZNF717*: NM_001128223.1) (71). Spremembe, ki smo jih prepoznali s sekvenciranjem po Sangerju, smo primerjali s spremembami predhodno opredeljenimi z WGS in na ta način potrdili oz. izključili njihovo prisotnost.

4 REZULTATI

4.1 KLINIČNE ZNAČILNOSTI PREISKOVANCEV

Preglednica VIII prikazuje 4 družine s pripadajočimi člani. Za vsakega člana družine imamo v preglednici podatke o starosti v času obravnave, o družinski anamnezi in klinični diagnozi. V četrti družini sta za prezgodnjo puberteto diagnosticirani obe hčerki, zato smo obravnavali obe, čeprav je bila v analizo WGS vključena samo ena. Hčerke 1. 3. in 4. družine so bile zdravljene z GnRH analogom – triptorelinom.

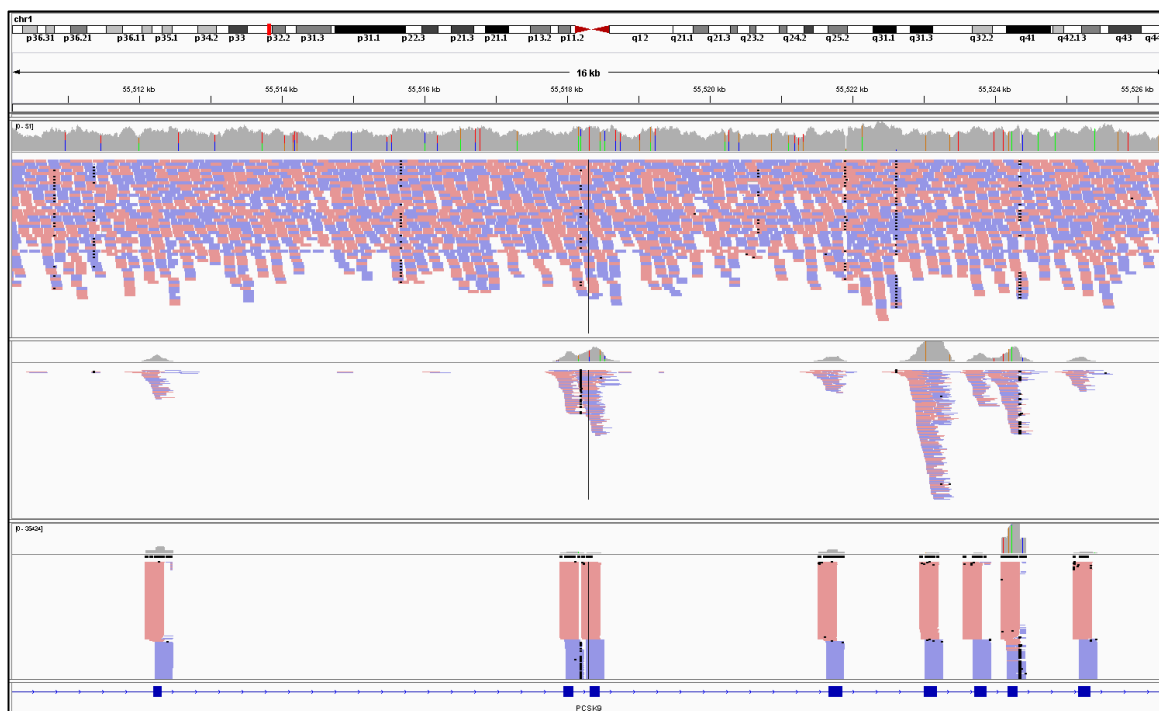
Preglednica VIII. Podatki o značilnostih preiskovancev (PP – prezgodnja puberteta, N – normalen pubertetni razvoj).

Družina	Preiskovanec	Starost	Družinska anamneza	Klinična diagnoza	Starost ob nastopu PP
1	hčerka 1	10	mama PP	PP	6,5
	mama 1	40	/	PP	menarha pri 10
	oče 1	39	/	N	/
2	hčerka 2	12	mama PP	PP	7,5
	mama 2	40	/	PP	menarha pri 9
	oče 2	40	/	N	/
3	hčerka 3	11	mama PP, oče zgodnejša puberteta	PP	6
	mama 3	39	/	PP	menarha pri 10
	oče 3	48	/	zgodnja puberteta	/
4	hčerka 4.1	10	sestra PP, mama PP	PP	8
	hčerka 4.2	9	sestra PP, mama PP	PP	6
	mama 4	38	/	PP	menarha pri 10
	oče 4	42	/	N	/

4.2 SEKVENCIRANJE CELOTNEGA GENOMA IN BIOINFORMATSKA ANALIZA

Analizo WGS smo izvedli na izbranih članih iz štirih družin s prisotno prezgodnjo puberteto, in sicer iz vsake družine pri eni hčerki in obeh starših (skupno na 12 preiskovancih).

Rezultati WGS analize so bili podani v obliki BAM mape (angl. *BAM file*). Gre za stisnjeno binarno verzijo SAM mape (angl. SAM file – Sequence Alignment/Map) (69). Obe obliki uporabljamo za prikaz poravnave sekvenc na referenčno zaporedje. BAM mapa je bila oblikovana z osnovnim algoritmom binarnega označevanja – vsaka sekvenca je bila po poravnavi na referenčno zaporedje označena z binarno oznako (angl. *bin index*) glede na njeno dolžino in mesto poravnave (69). Poravnane sekvence so bile označene z isto binarno oznako, če so bile enako dolge in so pripadale istemu mestu na referenčnem zaporedju. Iste binarne oznake so se naložile ena na drugo, da so oblikovale stolpec. Kadar se sekvenca ni skladala z referenčnim zaporedjem, je bil spremenjeni del zaporedja posebej označen. Na *Sliki 10* je prikazan izsek iz BAM mape z označenim mestom, kjer je bila v zaporedju določena genetska sprememba. Pokritost spremenjenega mesta je bila 48x, kar je za klinično interpretacijo rezultata dovolj zanesljiva vrednost.



Slika 10. Rezultat WGS analize - izsek iz BAM mape. Puščica označuje mesto genetske spremembe v sekvenci.

Rezultat bioinformatične obdelave rezultatov pridobljenih z metodo WGS je predstavljen v Prilogi 1.

Rezultat bioinformatične analize sprememb je prikazan z Vennovim diagramom (glej *Slika 11*). Vennov diagram gradijo 4 množice genov, v katerih so bile identificirane potencialno

vzročne genetske spremembe, vsaka od množice pripada eni od štirih družin. Genetske spremembe sodijo med vrste SNV in vrste INDEL. Množice se med seboj presekajo, kadar vsebujejo skupne gene. Predpostavili smo, da bo zaradi podobne klinične slike preiskovancev kandidatna vzročna genetska sprememba ležala pri vseh družinah v istem genu. Zato smo iskali gene, ki so skupne vsem štirim družinam. Pri tem smo identificirali en kandidatni gen *ZNF717*, ki je ustrezal kriterijem. Potencialno vzročne spremembe v tem genu so bile prisotne pri vseh štirih materah in hčerkah, kar je nakazovalo na možnost, da je gen vzrok družinske CPP. Ostali kandidatni geni niso bili prisotni pri vseh štirih družinah, zato smo jih izključili iz obravnave. V vseh štirih družinah smo z *in silico* algoritmom CADD ovrednotili genetske spremembe v genu *ZNF717* in s tem ocenili možnost za njihovo vzročnost. Rezultati so prikazani v *Preglednici IX*. Algoritem je vsako genetsko spremembo kvantitativno ocenil tako, da ji je dodelil surovo oceno in PHRED vrednost. Višja kot je bila surova ocena, večja je bila verjetnost, da je sprememba vzročna. Za prazno vrednost patogenosti smo izbrali PHRED vrednost 20, kar pomeni, da so genetske spremembe nad to vrednostjo sodile znotraj 1% najbolj vzročnih sprememb v človeškem genomu (64). V *Preglednici IX* so genetske spremembe, ki so presegale izbrano prazno vrednost in so bile ocenjene kot vzročne, obarvane rdeče. Takšne genetske spremembe so bile prisotne pri družinah 2, 3 in 4, nahajale so se na različnih pozicijah v zaporedju gena *ZNF717* in so kodirale različne aminokislino. Pri družinah 3 in 4 sta bili prisotni enaki spremembi, in sicer substitucija gvanina (G) v timin (T) (NM_001128223.1:c.644G>T/NP_001121695.1:p.Gly215Val) in NM_001128223.1:c.1255G>T/NP_001121695.1:p.Gly419Trp), pri družini 2 pa je bila prisotna substitucija citozina (C) v gvanin (G) (NM_001128223.1:c.746C>G/NP_001121695.1:p.Ser249Ter).

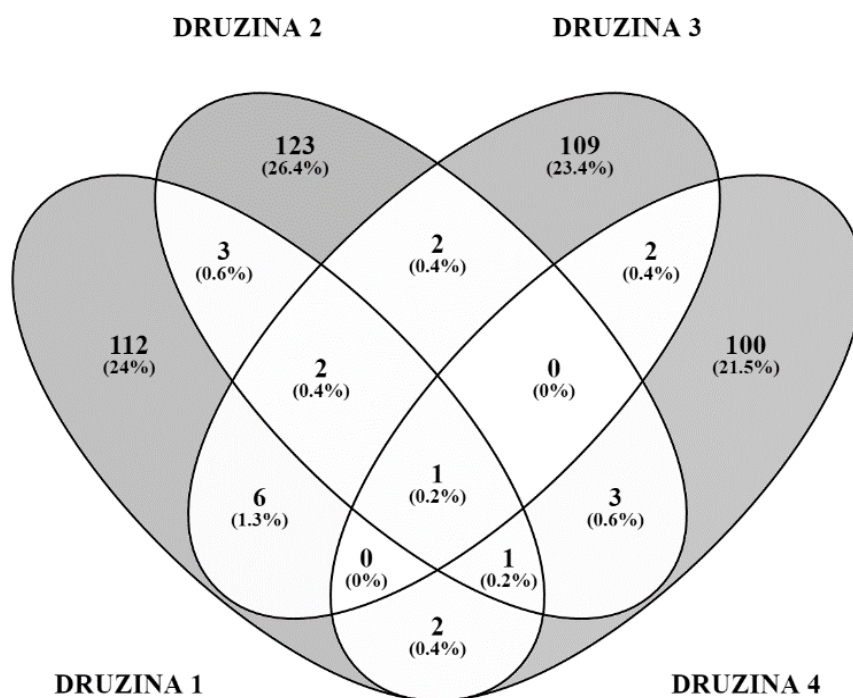
Družino z dvema obolelima hčerkama smo uporabili za dodatno opredelitev morebitne vzročnosti omenjenega gena, saj smo pričakovali prisotnost identične kandidatne genetske spremembe tudi pri sestri, pri kateri WGS ni bil opravljen. Posledično smo s sekvenciranjem po Sangerju potrjevali prisotnost morebitne vzročne genetske spremembe, ki smo jo opredelili tekom analize WGS, in sicer v eksonu 5, kjer je bila prisotna substitucija gvanina (G) v timin (T) (NM_001128223.1:c.1255G>T). Spremenjeni triplet je kodiral aminokislino triptofan namesto glicina (NP_001121695.1:p.Gly419Trp).

Preglednica IX. Z algoritmom CADD opredeljene genetske spremembe v genu ZNF717. Rdeče obarvane genetske spremembe so predvideno patogene in s tem potencialno vzročne za prezgodnjo puberteto (REF – normalno prisoten nukleotid v zaporedju gena, ALT – spremenjeni nukleotid v zaporedju gena, PHRED – kvantitativna ocena patogenosti, prazna vrednost za patogenost je 20).

Družina	Sprememba	REF	ALT	Surova ocena	PHRED
1	NM_001128223.1:c.460A>G NP_001121695.1:p.Met154Val	A	G	-1,227297	0,006
2	NM_001128223.1:c.2393A>G NP_001121695.1:p.Lys798Arg	A	G	-1,057872	0,012
	NM_001128223.1:c.1832G>T NP_001121695.1:p.Arg611Ile	G	T	4,316751	24
	NM_001128223.1:c.1783G>A NP_001121695.1:p.Glu595Lys	G	A	3,054623	22,4
	NM_001128223.1:c.1742A>G NP_001121695.1:p.His581Arg	A	G	2,637415	20,4
	NM_001128223.1:c.746C>G NP_001121695.1:p.Ser249Ter	C	G	7,281704	34
	NM_001128223.1:c.706T>G NP_001121695.1:p.Phe236Val	T	G	0,623781	8,32
	NM_001128223.1:c.644G>T NP_001121695.1:p.Gly215Val	G	T	3,374691	22,9
	NM_001128223.1:c.622A>G NP_001121695.1:p.Thr208Ala	A	G	0,797759	9,45
	NM_001128223.1:c.575A>G NP_001121695.1:p.His192Arg	A	G	-2,536191	0,001
	NM_001128223.1:c.548G>T NP_001121695.1:p.Cys183Phe	G	T	0,787812	9,389
NM_001128223.1:c.458G>A NP_001121695.1:p.Gly153Glu	G	A	2,044792	16,5	
NM_001128223.1:c.340C>A NP_001121695.1:p.Gln114Lys	C	A	-1,441611	0,003	

Preglednica X. Nadaljevanje Preglednice IX.

Družina	Sprememba	REF	ALT	Surova ocena	PHRED
3	NM_001128223.1:c.2703G>T NP_001121695.1:p.Glu901Asp	G	T	0,702242	8,842
	NM_001128223.1:c.2695G>A NP_001121695.1:p.Val899Ile	G	A	1,88339	15,49
	NM_001128223.1:c.644G>T NP_001121695.1:p.Gly215Val	G	T	3,374691	22,9
4	NM_001128223.1:c.2395A>G NP_001121695.1:p.Thr799Ala	A	G	-0,982703	0,017
	NM_001128223.1:c.2393A>G NP_001121695.1:p.Lys798Arg	A	G	-1,057872	0,012
	NM_001128223.1:c.2389G>C NP_001121695.1:p.Asp797His	G	C	-0,814364	0,04
	NM_001128223.1:c.1832G>T NP_001121695.1:p.Arg611Ile	G	T	4,316751	24
	NM_001128223.1:c.1316C>T NP_001121695.1:p.Thr439Ile	C	T	0,278697	5,485
	NM_001128223.1:c.1310G>A NP_001121695.1:p.Arg437Lys	G	A	-0,068078	1,979
	NM_001128223.1:c.1298G>A NP_001121695.1:p.Ser433Asn	G	A	-0,156256	1,348
	NM_001128223.1:c.1258G>C NP_001121695.1:p.Glu420Gln	G	C	3,406774	23
	NM_001128223.1:c.1255G>T NP_001121695.1:p.Gly419Trp	G	T	4,872661	24,9
	NM_001128223.1:c.1042T>G NP_001121695.1:p.Phe348Val	T	G	2,199065	17,5
	NM_001128223.1:c.973A>G NP_001121695.1:p.Ser325Gly	A	G	-0,684785	0,078
	NM_001128223.1:c.877A>G NP_001121695.1:p.Ile293Val	A	G	-0,764332	0,051
	NM_001128223.1:c.548G>C NP_001121695.1:p.Cys183Ser	G	C	-0,219568	0,998
NM_001128223.1:c.460A>G NP_001121695.1:p.Met154Val	A	G	-1,227297	0,006	

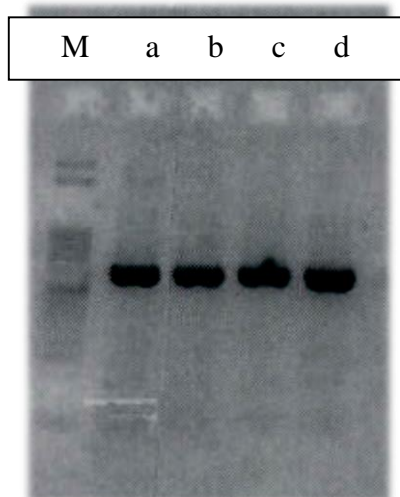


Slika 11. Rezultat bioinformatične analize rezultatov pridobljenih z metodo WGS je Vennov diagram, ki prikazuje preseke sprememb 4 družin.

4.3 POTRDI TVENA ANALIZA

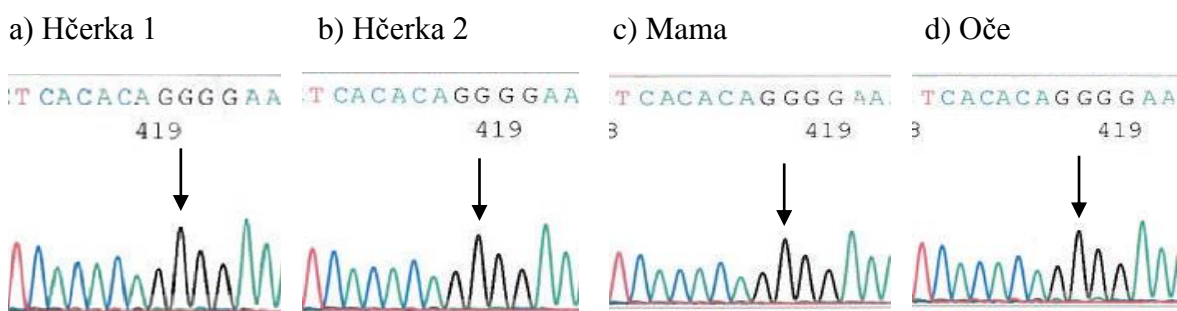
Potrjevanje prisotnosti kandidatnih sekvenčnih sprememb smo izvedli s sekvenciranjem po Sangerju na vzorcih gDNA preiskovancev četrte družine (rezultati so prikazani na *Sliki 12* in *Sliki 13*). V tej družini je bila poleg mame in ene hčerke za prezgodnjo puberteto diagnosticirana še druga hčerka, zato smo v potrditveno analizo vključili tudi njo. Takšna analiza ima zaradi dodatnega vzorca znotraj iste družine večjo moč in s tem večjo vplivnost na rezultat. Z omenjeno analizo smo želeli preveriti ali je sprememba predvidenega gena, ki je glede na bioinformatično analizo spremenjen pri vseh štirih družinah, pri njih zares prisotna. V ta namen je analiza ene od štirih družin dovolj.

Pred sekvenčno reakcijo smo vzorce pomnožili in odseke gDNA ločili z elektroforezo, da smo preverili uspešnost pomnožitve. Rezultat elektroforezne ločbe odsekov gDNA vzorcev mame, očeta in obeh hčerk prikazuje *Slika 12*.



Slika 12. Z elektroforezo ločeni odseki gDNA. V prvi koloni so ločeni odseki gDNA označevalca velikosti odsekov (M). Nato si iz leve proti desni vzorci gDNA četrte družine sledijo po naslednjem vrstnem redu: obe hčerki (a in b), mama (c) ter oče (d).

Zaporedje genoma določeno s sekvenčno reakcijo je prikazano kot elektroferogram (glej Slika 13). Pri obeh bolnih hčerkah in mami v rezultatih pridobljenih s potrditveno metodo nismo dokazali spremembe, ki smo jo predvideli z bioinformatiko analizo rezultatov WGS. V eksonu 5 gena *ZNF717* v kodonu 419 ni bilo prisotne substitucije G v T (NM_001128223.1:c.1255G>T/NP_001121695.1:p.Gly419Trp). Tako pri bolnih hčerkah in mami, kot tudi pri zdravem očetu nismo odkrili nobenih drugih genetskih sprememb v eksonu 5 gena *ZNF717*.

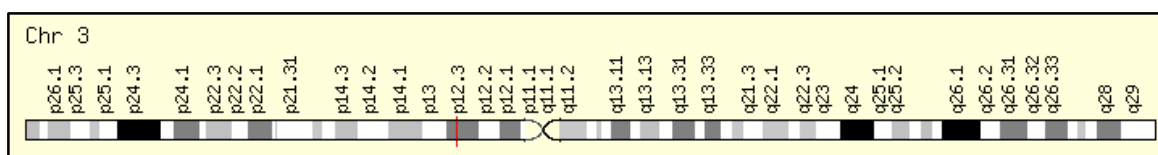


Slika 13. Rezultat potrditvene analize pri družini 4. Prikazani so 4 elektroferogrami (a, b, c in d), ki pripadajo četrti družini - hčerkama (a in b), mami (c) in očetu (d). Noben preiskovanec nima prisotne opredeljene genetske spremembe (NM_001128223.1:c.1255G>T/NP_001121695.1:p.Gly419Trp), ki smo jo preverjali s potrditveno metodo – sekvenciranje po Sangerju. Nukleotidi na elektroferogramu so prikazani z različnimi barvami, črni predstavljajo gvanin (G), zeleni adenin (A), modri citozin (C) in rdeči timin (T).

5 RAZPRAVA

V magistrski nalogi smo pri štirih družinah, ki so zajemale 4 mame in 5 hčerk s klinično diagnozo prezgodnje pubertete, želeli opredeliti genetske spremembe, ki bi bile vzrok za prezgodnjo puberteto. Z metodo WGS smo vsem staršem in štirim hčerkam določili zaporedje celotnega genoma. Pri tem smo pridobili velikansko količino podatkov, ki smo jih analizirali z bioinformatičnimi pristopi. V analizi smo preiskali kodirajoči del, bolj natančno genetske spremembe vrste SNV in INDEL, genoma preiskovancev in opredelili en kandidatni vzročni gen za prezgodnjo puberteto. Gre za gen *ZNF717*, ki se nahaja na krajši p-ročici kromosoma 3 (3p12.3), mesto nahajanja prikazuje *Slika 14* (72). Velik je 0,9 kbp (73). Izraža se predvsem v možganih, hipokampusu, talamusu, očesu, v pljučih, dojkah in tumorjih zarodnih celic (74). Kodira protein KRAB (angl. *Kruppel associated box*), ki sodi med transkripcijske faktorje. Zgrajen iz dveh domen (KRAB domene in C2H2 domene cinkovih prstov) (74). Proteini, ki nosijo motive cinkovih prstov, vežejo nukleinske kisline in kovinske ione (cink in železo). S tem uravnavajo gensko transkripcijo in translacijo, vplivajo na mRNA delovanje in preoblikujejo kromatin. Na ta način nadzorujejo in vodijo izražanje genov (74). Strukture cinkovih prstov, ki jih najdemo v produktu gena *ZNF717*, lahko najdemo tudi v produktu gena *MKRN3*, ki je dokazano genetski vzrok prezgodnje pubertete. Glede na to, da je protein, ki ga kodira kandidatni gen *ZNF717*, transkripcijski faktor in se izraža v celicah hipokampusa, obstaja možnost, da je vpleten v regulacijo aktivnosti HHG osi. Zato je bil ta gen dober kandidatni gen za nov genetski dejavnik, ki vpliva na prezgodnji razvoj pubertete.

S potrditvenim sekvenciranjem po Sangerju nismo potrdili, da bi bila sprememba (NM_001128223.1:c.1255G>T/NP_001121695.1:p.Gly419Trp) v genu *ZNF717* lahko vzrok za družinsko prezgodnjo puberteto, saj segregacija spremembe ni bila skladna s segregacijo klinične slike. S tem smo močno zmanjšali možnost, da bi bil ta gen odgovoren za razvoj prezgodnje pubertete pri analiziranih družinah.



Slika 14. Mesto nahajanja gena ZNF717 na kromosomu 3 (72).

5.1 IZVOR NAPAK PRI SEKVENCIRANJU CELOTNEGA GENOMA IN BIOINFORMATSKI ANALIZI REZULTATOV

Napake so lahko lažno pozitivne ali lažno negativne. Pri lažno pozitivni napaki pri sekvenciranju naslednje generacije opredelimo spremembo, ki dejansko ni prisotna v preiskovančevem genomu. Lažno negativna napaka pa predstavlja spremembo, ki je dejansko prisotna v preiskovančevem genomu, a s sekvenciranjem ni bila določena in zato neprepoznana. Z bioinformatško analizo smo v rezultatih sekvenciranja s pristopom WGS kot možni vzrok prezgodnje pubertete opredelili genetske spremembe v genu *ZNF717*. Pri vseh štirih družinah - pri 4 mamah in njihovih hčerkah je imel omenjeni gen prisotne kandidatne vzročne spremembe. Dejansko prisotnost spremembe v tem genu smo preverjali pri četrti družini, v kateri so prezgodnjo puberteto doživeli mama in dve hčerki. S potrditveno analizo nismo dokazali te spremembe pri preiskovankah s prezgodnjo puberteto. Gre torej za lažno pozitivno spremembo, ki je bila določena z WGS metodo, ni pa bila potrjena s potrditveno metodo. Vzroki, da z WGS metodo dobimo napačne rezultate, so različni in so podrobneje opisani v nadaljevanju.

5.1.1 NAPAKE V OBDELAVI PODATKOV PRIDOBLENIH S SEKVENCIRANJEM CELOTNEGA GENOMA

Surovi podatki pridobljeni s sekvenciranjem po principu WGS so kratka zaporedja, ki jih v celotno zaporedje genoma sestavimo z računalniškim postopkom. Z uporabo različnih algoritmov kratka zaporedja poravnamo na referenčni genom in jim določimo genomske koordinate. Nato z računalniškim orodjem (npr. GATK ali VCMM) določimo neuskkljenosti med referenčnim in preiskovančevim genomom. Pomembno je zmanjšati napake v začetnem delu analiznega postopka, torej pri določitvi koordinat in pri preverjanju usklajenosti med referenčnim in preiskovančevim genomom, saj v nasprotnem primeru celotna nadaljnja interpretacija temelji na napačnih predpostavkah. Napake se lahko pojavijo že v poravnavi zaporedja na referenčni genom, kateri naj ne bi bil v celoti pokrit - obstajalo naj bi več lokusov, ki nimajo določenega zaporedja; v referenčnem genomu naj bi manjkalo čez 160 000 000 baz (75). Po poravnavi zaporedja na referenčni genom sledi anotiranje genetskih sprememb prisotnih v zaporedju, pri tem podatke za anotacijo pridobimo iz različnih podatkovnih baz. Gre za najpomembnejši del obdelave

rezultatov WGS, ki pa vseeno ni brez pomanjkljivosti. Podatkovne baze niso povsem zanesljive. V njih na primer obstajajo zapisane genetske spremembe, ki so navedene kot vzročne, ker pa ne dosegajo dovolj visoke penetrance (t.j. stopnje izražanja), je njihova povezava z boleznijo vprašljiva (76). Razlog za negotove in tudi napačne objave je pomanjkljiv nadzor nad objavljanjem genetskih podatkov, predvsem je pomanjkljivo preverjanje podatkov pred vključitvijo v bazo (75, 77), tudi ni postavljenih doslednih smernic, ki bi dovolj natančno razlikovale med vzročnimi in ne-vzročnimi spremembami (78). Podatkovne baze prav tako ne vsebujejo opisov zelo redkih in patogenih genetskih sprememb (50).

5.1.2 SLABŠA OBČUTLJIVOST IN TOČNOST SEKVENCIRANJA CELOTNEGA GENOMA

Obe, analitična točnost in analitična občutljivost metode, določata kakšna bosta pravilnost in zaznavnost sprememb v zaporedju genoma. Analitična točnost je sposobnost metode, da bazo v zaporedju določi pravilno (79). Analitična občutljivost pa pove, kolikšen delež baz lahko zaznamo v zaporedju (79).

Vse kar vpliva na točnost vpliva tudi na občutljivost in obratno. Vpliv ima več spremenljivk, te so: izbrana tehnologija sekvenciranja, način priprave vzorca, poravnava, kvaliteta določitve baz, globina pokritosti, genomska regija, vrste genetskih sprememb in bioinformatični pristopi v analizi sprememb (77).

Verjetnost, da pravilno zaznamo vse spremembe prisotne v zaporedju in s tem občutljivost in točnost so manjši, kadar so baze določene z nizko kvaliteto (to pomeni, da je v zaporedju visoka stopnja napake) in kadar je zaporedje neenakomerno pokrito. Kvaliteta branja je odvisna od tega, katero tehnologijo sekvenciranja uporabimo, na enakomernost pokritosti zaporedja pa vpliva še priprava vzorca, npr. kemijska obdelava z GC bogatih regij izboljša pomnoževanje DNA teh regij, in s tem poveča enakomernost pokritosti (80). Pokritost za WGS metodo je standardizirana glede na vrsto genetske spremembe, ki jo iščemo in glede na naravo preiskovane bolezni. Če želimo zajeti vse spremembe vrste SNV in INDEL je zahtevana 50x povprečna pokritost (80, 81).

Sekvenčni odčitki, ki jih dobimo z NGS sekvenciranjem, so določeni z neko stopnjo napake. Večkrat kot so določeni bolj smo lahko prepričani v pravilnost njihovih določitve. Težava nastopi, kadar se napaka ne loči od spremembe. Takrat je rešitev ustrezno povečanje števila identičnih sekvenčnih odčitkov na predvidenem mestu spremembe v

povezavi s stopnjo napake – poveča se globino pokritosti (80). Ker je na ta način morebitna sprememba določena z večjo gotovostjo, se zmanjša možnost lažno pozitivnih rezultatov. Kadar želimo opredeliti spremembe, ki so vzročne za dedne bolezni običajno naredimo kompromis med številom vzorcev in pokritostjo. Delež lažno pozitivnih sprememb se kljub nižji pokritosti zmanjša, kadar podatke več vzorcev – mame, očeta in otroka, kombiniramo med seboj (80). V tem primeru je priporočljivo uporabiti enako pokritost za vsakega analiziranega člana, zato da so morebitno prisotne spremembe pri vseh določene z enako gotovostjo in so rezultati primerljivi. S tem zmanjšamo možnost napačno pozitivnih in napačno negativnih rezultatov (80).

Vseeno pa tudi, če zmanjšamo vplive spremenljivk (npr. s pripravo vzorca) in optimalno prilagodimo parametre v bioinformatiki analizi rezultatov (npr. globino pokritosti), ne moremo doseči 100 % občutljivosti in točnosti WGS. Zaradi tehničnih lastnosti WGS sta občutljivost in točnost metode okoli 97,5 % (77, 82). Aparat HiSeq 4000, na katerem se je izvedla analiza WGS, je določil 84-92 % sprememb vrst SNV, INDEL, CNV in SV z 99,9 % točnostjo, pri čemer je bila povprečna pokritost sekvenciranja 37x, kar je v našem primeru, ko smo kombinirali podatke večih vzorcev, dovolj (48).

5.1.3 KOLIČINA PODATKOV

Obstajajo primeri podedovanih bolezni, ki pri preiskovancih niso razrešeni, ker vzročnih genetskih sprememb niso iskali v nekodirajočih regijah genoma (50). Z analitičnim pristopom, ki dovolj učinkovito razlikuje med vzročnimi in benignimi genetskimi spremembami, lahko izkoristimo prednost določitve celotnega genoma pri identifikaciji genetskih sprememb odgovornih za bolezni. Celovita preiskava in opredelitev genetskih sprememb, ki niso omejene zgolj na kodirajočo regijo genoma, je mogoča z metodo WGS. Z njo lahko identificiramo genetske spremembe prisotne v nekodirajočih regijah genoma, ki bi jih drugače s tarčno usmerjenimi metodami (npr. sekvenciranje celotnega eksoma, angl. *whole exome sequencing* – WES) lahko zgrešili. Prav tako je z WGS pristopom pri dovolj visoki pokritosti (vsaj 150x (83)) mogoča identifikacija vseh sprememb vrste CNV, ki obstajajo v človeškem genomu (84). Količina podatkov, ki jih dobimo z WGS metodo pa je tista, ki predstavlja težavo, ko iščemo genetske spremembe pomembne za medicinsko genetiko v vseh regijah genoma. Določitev zaporedja enega genoma, ki je omejena na 30x globino pokritosti in 660 milijonov branj, nam da količino podatkov, ki je enaka 100 GB (75). Znotraj več kot 4 milijonov vseh sprememb, ki so v heterogenem človeškem genomu

normalno prisotne, je potrebno najti eno ali nekaj sprememb, ki so potencialno vzročne za preiskovano bolezen (54). Tolikšna količina podatkov je težavna za biomedicinsko obravnavo. Za analize celotnih genomov je potrebna dovolj zmogljiva strojna in programska oprema ter čas, da lahko obdelamo toliko podatkov naenkrat. Ko bo genomsko sekvenciranje postalo klinično dovolj pomembno, bo potrebna oprema verjetno tudi komercialno dostopnejša.

5.1.4 KOMPLEKSNOŠT GENOMA

Zaporedje človeškega genoma v 50-69 % gradijo ponavljajoča zaporedja (77).

Ponavljajoča zaporedja so transpozoni (deli zaporedja, ki se znotraj genoma prestavljajo iz ene lokacije na drugo), nizko kompleksna območja (kot so homopolimeri) in psevdogeni (ne funkcionalne kopije zaporedja zelo podobne normalnim genom, a se ne izražajo).

Obstajajo tudi paralogni geni (nastanejo s podvojitvijo dela zaporedja znotraj genoma), ki so povezani z boleznimi, nekateri od njih tudi z ekscitacijskimi nepravilnostmi v živčnem sistemu (možna povezava s prezgodnjo puberteto) (77). Ključnega pomena v kliničnem sekvenciranju je točnost določitve zaporedja (77). Ponavljajoča narava genoma povzroča težave v poravnavi zaporedja na referenčnega in ima vplive na točnost določitve zaporedja. Ponavljajoči del zaporedja se z enako verjetnostjo poravnava z več mesti na referenčnem zaporedju. Če je v preiskovančevem genomu prisotna sprememba v eni in ne v več kopijah ponavljajočega zaporedja, se lahko napačno naloži na referenčni genom. Napačna poravnava vodi v napačno pozitivne ali napačno negativne določitve sprememb v zaporedju in s tem v klinično napačne odločitve.

5.1.5 NARAVA DEDNE BOLEZNI

Pri avtosomno-dominantno podedovanih boleznih je za patološki fenotip odgovorna le ena izmed sprememb gena povezanega z dedno boleznijo. Iskanja dednih patogenih genetskih sprememb, torej takšnih, ki se dedujejo iz staršev na otroka, se je potrebno lotiti s filtriranjem sprememb, s čimer se znebimo tudi *de novo* pridobljenih sprememb. Osnovna predpostavka pri dednih boleznih je, da si vsi bolni znotraj družine delijo enake genetske spremembe. Cilj filtriranja je prepoznati genetske spremembe, ki so skupne bolnemu staršu in bolnemu otroku, saj so potemtakem potencialno vzročne ter zavreči množico ostalih, ki so hkrati prisotne v zdravem staršu in bolnem otroku, saj so potemtakem nepovezane z dedno boleznijo. Analitični pristop, ki temelji na primerjavi genomov znotraj družine

omogoča, da zavržemo benigne spremembe, ki so v opisih lahko patogene in da obdržimo tiste, ki so v opisih sicer benigne, a v resnici lahko patološke. Takšen osnovni pristop v analizi dednih bolezni lahko zmanjša število kandidatnih genov iz tisočih na enega (50).

Dominantne genetske spremembe so lahko posledica tudi *de novo* mutacij. V tem primeru ne moremo uporabiti takšnega načina filtriranja, ampak se moramo osredotočiti na spremembe, ki so prisotne samo pri bolniku ne pa pri njegovih starših.

Obstajajo pa bolezni, ki so genetsko heterogene, ločimo alelno in lokusno heterogenost. Alelna heterogenost pomeni, da različne genetske spremembe istega lokusa podobno vplivajo na fenotip in povzročijo isto bolezen; lokusna heterogenost pa pomeni, da različne genetske spremembe genov na različnih lokusih povzročijo isto bolezen. S prej opisanim analitskim pristopom lahko pri lokusno heterogenih dednih boleznih dobimo napačno pozitivne rezultate (50). V primeru lokusno heterogene dedne bolezni si potencialno patogene spremembe preiskovanci delijo le po naključju in v resnici so spremembe benigne.

Lažen rezultat, ki smo ga pridobili z bioinformatično analizo rezultatov WGS, je torej v tej raziskavi lahko odraz naslednjih predvidenih napak in predpostavk:

- Napaka je lahko analitskega izvora. Zaradi tehničnih lastnosti je WGS metoda nagnjena k pomanjkljivi detekciji sprememb. Lahko je prišlo do izgube sprememb v začetni obdelavi rezultatov WGS; primer je napaka poravnave zaporedja na referenčni genom, ker je le-to nepopolno.
- Ponavljajoča območja genoma so nagnjena k napačni poravnavi. Napačna poravnava in netočni podatki v podatkovnih bazah vodijo v napačno in nepopolno anotacijo sprememb prisotnih v preiskovanem genomu, zaradi česar smo lahko izgubili del sprememb; med njimi morda tudi tisto, ki je vzročna za dedno prezgodnjo puberteto.
- Slabost bioinformatične analize je zagotovo v dejstvu, da nekodirajočih regij genoma nismo interpretirali. Te regije predstavljajo vsaj 98 % genoma (44). Glede na dejstvo, da nekodirajoča regija regulira gene kodirajoče regije, je neinterpretiran ostal pomemben del genoma, v katerem se lahko nahajajo spremembe s spremenjenim izrezovanjem in z regulacijo genov vzročnih za prezgodnjo puberteto.

- Če predpostavimo, da je preiskovana oblika dedne prezgodnje pubertete lokusno heterogena, pomeni, da jo povzročijo genetske spremembe, ki jih najdemo na različnih lokusih. V tem primeru so skupne in torej potencialno patogene spremembe gena, ki si ga delijo vse štiri družine, skupne le po naključju in v resnici niso vzrok za dedno prezgodnjo puberteto. Za dedno prezgodnjo puberteto, ki je lokusno heterogena, je lahko v vsaki družini odgovoren drug gen in v njem druga vzročna sprememba.
- Prisotnost prezgodnje pubertete znotraj družine še ni dokaz, da gre za dedno obliko bolezni. Pozitivna družinska zgodovina lahko pomeni podedovano družinsko CPP, a je potrebna potrditev genetskih sprememb v preiskovancih. Potencialno vzročnih genetskih sprememb v preiskovanih družinah nismo potrdili, zato lahko obstaja tudi možnost, da ne gre za dedno obliko ampak za pridobljeno obliko prezgodnje pubertete (npr. posledica okoljskih dejavnikov ali manj verjetno *de novo* mutacij pri hčerkah). Za razjasnitev so potrebne nadaljnje genetske preiskave.

5.2 POMEN POTRJEVANJA REZULTATOV SEKVENCIRANJA NASLEDNJE GENERACIJE

V laboratorijski diagnostiki dednih bolezni so rezultati genetskih testiranj izrednega pomena. Na njih temeljijo pomembne klinične odločitve, ki se ne dotikajo samo bolnika, ampak širše družine. Zato je pravilnost rezultatov in interpretacija rezultatov ključna za pravilen izvid. Izguba katerekoli z boleznijo povezane spremembe ali izdaja napačno pozitivnega rezultata, bi lahko imela pomembne zdravstvene posledice za preiskovance. Testi za sekvenciranje so visoko zanesljivi takrat, ko dosegajo minimalno globino pokritosti nad 100x za spremembe, ki imajo alelno frekvenco nad 40 % (82). Kadar ti kriteriji niso doseženi, morajo biti vzorci analizirani še s sekundarno oz. potrditveno metodo, s sekvenciranjem po Sangerju. Visoka občutljivost in specifičnost diagnostičnih testov zagotavljata zanesljivost rezultatov. V splošnem sekvenciranja naslednje generacije, kamor spada tudi WGS, dosegajo občutljivost okoli 97,5 %, medtem ko je pri Sangerju 100 % (82). Vse NGS metode so v primerjavi s sekvenciranjem po Sangerju nagnjene k napačnim rezultatom (82). Vzrok za to je nižja analitična občutljivost NGS metod, ki pomeni manjšo zaznavnost baz, zato v analizi izgubimo del sprememb, med njimi lahko tudi tiste povezane z boleznijo. Preden v NGS analizi določimo zaporedje, ga poravnamo

na referenčno zaporedje. V genomu obstajajo deli zaporedja - območja s homopolimeri in psevdogeni ter ostale ponavljajoče se regije - ki so nagnjena k napačni poravnavi. Napačna poravnava povzroči, da z NGS metodo zaznamo manj sprememb. Rezultate pridobljene z NGS metodo lahko v nadaljnji bioinformatiki analizi prilagodimo s parametri (npr. uporabimo višje vrednosti za kvaliteto baz), da odstranimo napačno pozitivne spremembe. Vendar takšen pristop pomeni izgubo določenega števila sprememb, lahko tudi tistih, ki so klinično pomembne (82). Potrjevanje s Sangerjevo analizo je nujno potrebno zato, da obdržimo visoko občutljivost in specifičnost genetskega diagnostičnega testiranja, s čimer zagotovimo točnost določitve celotnega zaporedja.

5.3 NADALJNI KORAKI ZA OPREDELITEV GENETSKEGA VZROKA

Na koncu raziskave ostaja genetski vzrok prezgodnje pubertete v analiziranih družinah še vedno nepoznan. Za razjasnitev so potrebne dodatne preiskave, predlogov za nadaljno delo je več in so opisani v nadaljevanju.

5.3.1 ANALIZA GENETSKIH SPREMOMB VRST CNV IN SV

V kodirajoči regiji genoma genetskih sprememb vrste CNV in SV še nismo analizirali. V človeškem genomu je 4,5-15 % sprememb vrste CNV, kar pomeni, da jih najdemo v povprečju 12 v vsakem posamezniku (85, 86). Te spremembe združijo ali razdružijo gene, vplivajo na količino in nivo izražanja genov. Spremembe v številu kopij prispevajo k razvoju bolezni kot so Alzheimerjeva bolezen, shizofrenija, rak dojke, debelost ter vplivajo na razvoj človeka (85). Sprememb vrste SV je v genomu 1,2 % in naj bi bili vzrok za nevrološke oz. nevrokognitivne motnje ter razvoj in napredovanje raka (87). Zaradi povezave teh vrst sprememb z boleznimi je potrebno ponoviti analizo na spremembah vrste CNV in SV. Cilj analize bi bil enak cilju, ki je bil določen v analizi sprememb SNV in INDEL, torej najti in potrditi gene s spremembami vrste CNV oz. SV, ki so skupni vsem analiziranim družinam.

5.3.2 INTERPRETACIJA NEKODIRAJOČIH PODROČIJ GENOMA

V sklopu magistrske naloge smo določili nukleotidno zaporedje celotnega genoma, interpretirali pa smo le spremembe v kodirajočih zaporedjih. Tako je potrebno tudi nekodirajoča področja genoma interpretirati na enak način kot smo interpretirali kodirajoča. V tej analizi bi bilo potrebno opredeliti in potrditi spremembe, ki se pri vseh družinah nahajajo v istih intronih. Povprečni gen, ki kodira proteine, je zgrajen iz osmih do desetih eksonov in dvajsetkrat daljših intronov (88). Intronska zaporedja so pomembna, ker predstavljajo spojitelna območja in vezavna mesta za regulatorne proteine; kodirajo regulatorje transkripcije in RNA gene; introni, ki se po izrezu iz pre-mRNA molekule ne razgradijo oblikujejo funkcionalne regulatorne RNA molekule (malo jedrno RNA in mikro RNA) in preko različnega izrezovanja ali neizrezovanja lahko regulirajo izražanje genov (88). Prav zaradi teh funkcij intronov so spremembe v njih dokazano povezane z boleznimi. Spremembe v intronih lahko oblikujejo nova spojitelna območja, zaradi česar je motena prepoznavna običajnih spojitelvenih območij (88). Spremembe v transkripcijskih regulatornih motivih in RNA genih, ki jih kodirajo spremenjena intronska zaporedja, motijo regulacijo izražanja genov (88).

5.3.3 PREISKAVA CELOTNEGA GENOMA S CILJEM OPREDELITVE GENETSKIH SPREMEMB PO POSAMEZNI DRUŽINI

V primeru, da ne najdemo skupnih genov in/ali intronov s spremembami pri vseh družinah, bi analizo lahko osredotočili na posamezno družino. Preiskali bi celotne genome preiskovancev posamezne družine. Opredelili in potrdili bi spremembe v genih in/ali intronih, ki so hkrati skupne mami in hčerki, niso pa skupne očetu: (MAMA U HČI) – (OČE U MAMA U HČI).

5.3.4 PREISKAVA CELOTNEGA GENOMA S CILJEM OPREDELITVE GENETSKIH SPREMEMB SKUPNIH VSEM HČERKAM

Če se skupnih genov in/ali intronov s spremembami pri mami in hčerki posamezne družine ne najde, bi lahko iskali gene in/ali introne s spremembami, ki so skupne vsem hčerkam.

Prisotnost takšnih genetskih sprememb bi pomenilo, da niso bile podedovane, ampak so bile pridobljene *de novo* zaradi vplivov okolja ali epigenetskih sprememb.

5.3.5 EPIGENETSKE PREISKAVE

Kadar se pri vseh družinah, po posamezni družini ali pri vseh hčerkah ne bi našlo skupnih genov in/ali intronov, bi se lahko izvedle epigenetske preiskave. Epigenom je fizična struktura genoma z vsemi kemijskimi modifikacijami, ki ga preoblikujejo. Različne prostorske oblike genoma različno določajo aktivnost genov in različno vplivajo na regulacijo izražanja genov (89). Med epigenetske spremembe sodijo DNA metilacija, modifikacija histonov, organizacija kromatina (nukleosom) in vplivi delovanja nekodirajočih RNA molekul (89). Analizirali bi vsako hčerko posebej z epigenetskimi tehnikami kot sta bisulfitno sekvenciranje celotnega genoma (angl. *Whole genome bisulfite sequencing* - *WGBS*) ter kombinacija kromatinske imunoprecipitacije in sekvenciranja naslednje generacije (angl. *Chromatin immunoprecipitation with sequencing* – *ChIP-Seq*). Z uporabo teh tehnik bi omogočili opazovanje regulacije genoma v bolezenskem stanju (90). Z njimi bi raziskali spremembe v interakcijah med DNA, RNA in proteini, identificirali spremenjena vezavna mesta za transkripcijske faktorje, opazovali spremembe v strukturi kromatina in histonske modifikacije ter določili spremenjena vezavna mesta za proteine (90).

6 SKLEP

Rezultate in ugotovitve magistrske naloge lahko sklenemo v naslednjih točkah:

- Izbranim preiskovancem iz štirih družin smo z metodo sekvenciranja celotnega genoma določili nukleotidno zaporedje celotnega genoma.
- Z uporabo ustreznih analitskih orodij in algoritmov smo izbrali kandidatne genetske vzroke za prezgodnjo puberteto. Osredotočili smo se na potencialne vzročne gene za družinsko CPP, ki se deduje po materi.
- Za vsako družino smo oblikovali nabor približno 100 sprememb kodirajoče regije genoma, natančneje vrste SNV in INDEL, ki so skupne mami in hčerki. Pri vseh štirih družinah smo znotraj teh naborov odkrili spremembe,

ki se nahajajo v istemu genu, to je genu *ZNF717*. Z algoritmom CADD smo za družine 2, 3 in 4 predvideli patogene genetske spremembe v tem genu.

- S potrditvenim sekvenciranjem po Sangerju smo sorojenkama in staršema družine 4 določili nukleotidno zaporedje eksona 5 gena *ZNF717*. Na podlagi rezultatov smo ovrgli možnost, da je kandidatna sprememba v genu *ZNF717* dejansko vzrok za prezgodnjo puberteto.
- Zaradi lažno pozitivnega rezultata, ki smo ga pridobili z bioinformatično analizo rezultatov sekvenciranja celotnega genoma, smo razpravljali o možnih izvorih napak ter predvideli možne vzroke za napačen rezultat.
- Pristop sekvenciranja celotnega genoma omogoča opredelitev sprememb v kodirajočih in nekodirajočih področjih, vendar ni optimalen. Metoda ni 100 % točna in občutljiva. Ni nujno, da se določijo vse baze pravilno in ni nujno, da se zaznajo vse baze v zaporedju, kar pomeni, da obstaja možnost analitske napake.

Z raziskavo nismo uspeli potrditi zastavljenih hipotez. S pristopom WGS smo pri preiskovanih družinah prepoznali nov kandidatni gen, ki bi lahko bil vzrok za prezgodnjo puberteto, vendar potrditvena analiza tega ni potrdila. Sekvenciranje celotnega genoma s pristopom WGS ni brez napak, zato je potrjevanje WGS rezultatov s Sangerjem ključno za točno genetsko diagnostiko. Pri preiskovancih genoma nismo interpretirali v celoti; spremembe vrste CNV in SV kodirajoče regije ter nekodirajoča regija še vedno ostajajo neanalizirane. Možno je, da je vzrok bolezni katera izmed sprememb teh vrst. Seveda je vzrok bolezni pri preiskovanih družinah lahko heterogen in imajo preiskovane družine različne genetske vzroke za CPP in ne enakega, kar smo predpostavili v tej raziskavi. Bolezen pri preiskovanih družinah pa je lahko tudi epigenetsko pogojena ali povzročena z okoljskimi dejavniki. Zato so za opredelitev genetskega vzroka prezgodnje pubertete pri preiskovanih družinah potrebne nadaljnje raziskave.

7 LITERATURA

1. Kaplowitz PB. Precocious Puberty. Medscape: Drugs and Disease, Pediatrics: General Medicine. 2017. <http://emedicine.medscape.com/article/924002-overview>. (Dostopno: marec, 2017).
2. Burtis CA, Ashwood ER, Bruns DE, Sawyer BG. Fundamentals of Clinical Chemistry. Sixth edition. Philadelphia, Saunders, 2008: 780-812.
3. Abreu AP, Kaiser BU. Pubertal development and regulation. Lancet Diabetes Endocrinol. 2016; 4(3): 254–264.
4. Parent A, Teilmann G, Juul A, Skakkeabek EN, Toppari J, Bourguignon J. The Timing of Normal Puberty and the Age Limits of Sexual Precocity: Variations around the World, Secular Trends, and Changes after Migration. Endocrine reviews. 2003; 24(5): 668-693.
5. Euling YS, Herman-Giddens ME, Lee AP, Selevan GS, Juul A, Sørensen AT, Dunkel L, Himes HJ, Teilmann G, Swan SH. Examination of US Puberty-Timing Data from 1940 to 1994 for Secular Trends: Panel Findings. Pediatrics. 2008; 121: 172-191.
6. Kaplowitz BP, Oberfield SE. Reexamination of the Age limit for Defining When Puberty is precocious in girls in the United States: implications for evaluation and treatment. Drug and Therapeutics and Executive Committees of the Lawson Wilkins Pediatric Endocrine Society. Pediatrics. 1999; 104(4): 936-941.
7. Parent AS, Franssen D, Fudvoye J, Gerard A, Bourguignon JP. Developmental variations in environmental influences including endocrine disruptors on pubertal timing and neuroendocrine control: Revision on human observations and mechanistic insight from rodents. Front Neuroendocrinol. 2015; 38: 12-36.
8. Lakshman R, Forouhi N, Luben R, Bingham S, Khaw K, Wareham N, Ong KK. Association between age at menarche and risk of diabetes in adults: results from the EPIC-Norfolk cohort study. Diabetologia 2008; 51:781-786.
9. Lakshman R, Forouhi NG, Sharp SJ, Luben R, Bingham SA, Khaw KT, Wareham NJ, Ong KK. Early age at menarche associated with cardiovascular disease and mortality. J Clin Endocrinol Metab. 2009; 94:4953-4960.

10. He C, Zhang C, Hunter DJ, Hu FB. Age at menarche and risk of type 2 diabetes: results from two large prospective cohorts. *Am J Epidemiol.* 2009; 171:334-344.
11. Partsch C, Heger S, Sippell GW. Management and outcome of central precocious puberty. *Clinical Endocrinology.* 2002; 56: 129-148.
12. Abreu AP, Dauber A, Macedo DB, Noel SD, Brito VN, Gill JC, Cukier P, Thompson IR, Navarro VM, Gagliardi PC, Rodrigues T, Kochi C, Longui CA, Beckers D, de Zegher F, Montenegro LR, Mendonca BB, Carroll RS, Hirschhorn JN, Latronico AC, Kaiser UB. Central Precocious Puberty caused by mutations in Imprinted gene MKRN3. *The new England journal of Medicine.* 2013; 368(26): 2467-2475.
13. Houk CP, Kunselman AR, Lee PA. Adequacy of a single unstimulated luteinizing hormone level to diagnose central precocious puberty in girls. *Pediatrics.* 2009; 123(6): 1059-63.
14. Cesario K, Hughes AL. Precocious Puberty: A comprehensive Review of Literature. *Jognn.* 2007; 36(3): 263-274.
15. Teilmann G, Pedersen CB, Jensen TK, Skakkebaek NE, Juul A. Prevalence and incidence of precocious pubertal development in Denmark: an epidemiologic study based on national registres. *Pediatrics.* 2005; 116(6): 1323-1328.
16. Palmert RM, Boepple AP. Variation in the Timing of Puberty: Clinical Spectrum and Genetic Investigation. *J Clin. Endocrinol Metab.* 2001; 86(6): 2364-2368.
17. Elks CE. Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. *Nat Genet.* 2010; 42:1077-1085.
18. Macedo DB, Brito NV, Latronico AC. New causes of central precocious Puberty: The role of genetic factors. *Neuroendocrinology.* 2014; 100: 1-8.
19. Macedo BD, Silveira LF, Bessa DS, Brito VN, Latronico AC. Sexual precocity - Genetic bases of Central Precocious Puberty and Autonomous Gonadal Activation. *Genetic Issues.* 2016; 29: 50-71.
20. Carel JC, Leger J. Precocious Puberty. *N Engl J Med.* 2008; 358: 2366-2377.
21. World Health Organization. Antiretroviral Therapy for HIV Infection in Infants and Children: Towards Universal Access: Recommendations for a Public Health Approach. Geneva, 2010 Revision: ANNEX H, SEXUAL MATURITY RATING (TANNER STAGING) IN ADOLESCENTS.

22. Marshall VA, Tanner JM. Variations in pattern of pubertal changes in girls. *Arch Dis Child*. 1969; 44(235): 291-303.
23. Baypai A, Nenon PSN. Contemporary issues in precocious puberty. *Indian J Endocrinol Metab*. 2011; 15(3): 172-179.
24. Marshall VA, Tanner JM. Variations in pattern of pubertal changes in boys. *Arch Dis Child*. 1970; 45(239): 13-23.
25. Goede J, Hack WW, Sijstermans K, van der Voort-Doedens LM, Van der Ploeg T, Meij-de Vries A, Delemarre-van de Waal HA. Normative Values for Testicular Volume Measured by Ultrasonography in a Normal population from Infancy to Adolescence. *Horm Res Paediatr*. 2011; 76(1): 56-64.
26. Mughal MA, Hassan N, Ahmed A. Bone Age Assessment Methods: A Critical Review. *Pak J Med Sci*. 2014; 30(1): 211-215.
27. Galluzzi F, Salti R, Bindi G, Pasquini E, La Cauza C. Adult height comparison between boys and girls with precocious puberty after long-term gonadotrophin-releasing hormone analogue therapy. *Acta Paediatrica*. 1998; 87: 521-527.
28. Neely EK, Hintz RL, Wilson DM, Lee PA, Gautier T, Argente J, Stene M. Normal ranges for immunochemiluminometric gonadotropin assays. *Journal of Pediatrics*. 1995; 127: 40-46.
29. Tremblay L, Frigon JY. Precocious puberty in adolescent girls: A biomarker of later psychosocial adjustment problems. *Child Psychiatry Hum Dev*. 2005; 36:73–94.
30. Grandone A, Cantelmi G, Cirillo G, Marzuillo P, Luongo C, del Giudice EM, Perrone L. A case of familial central precocious puberty caused by a novel mutation in the makorin RING finger protein 3 gene. *BMC Endocrine Disorders*. 2015; 15: 60-64.
31. De Vries L, Gat-Yablonski G, Dror N, Singer A, Phillip M. A novel MKRN3 missense mutation causing familial precocious puberty. *Human Reproduction*. 2014; 29(12): 2838-2843.
32. Settas N, Dacou-Voutetakis C, Karantza M, Kanaka-Gantenbein C, Chrousos GP, Voutetakis A. Central Precocious Puberty in a Girl and Early Puberty in Her Brother Caused by a Novel Mutation in the MKRN3 Gene. *J Clin Endocrinol Metab*. 2014; 99(4): 647-651.

33. Abreau AP, Dauber A, Macedo DB, Noel SD, Brito VN, Gill JC, Cukier P, Thompson IR, Navarro VM, Gagliardi PC, Rodrigues T, Kochi C, Longui CA, Beckers D, de Zegher F, Montenegro LR, Mendonca BB, Carroll RS, Hirschhorn JN, Latronico AC, Kaiser UB. Central Precocious Puberty Caused by Mutations in the Imprinted Gene of MKRN3. *N Engl J Med*. 2013; 368(26).
34. Schreiner F, Gohlke B, Hamm M, Korsch E, Woelfle J. MKRN Mutations in familial Central Precocious Puberty. *Horm Res Paediatr* 2014; 82:122–126.
35. Hwang JS. The genes associated with gonadotropin-releasing hormone-dependent precocious puberty. *Korean J Pediatr*. 2012; 55(1): 6-10.
36. Simon D, Ba I, Mekhail N, Ecosse E, Paulsen A, Zenaty D, Houang M, Jesuran Perelroizen M, de Filippo GP, Salerno M, Simonin G, Reynaud R, Carel JC, Leger J, de Roux N. Mutation in the maternally imprinted gene MKRN3 are common in familial central precocious puberty. *European Journal of Endocrinology*. 2016; 174(1): 1-8.
37. De Vries L, Kauschansky, Shohat M, Phillip M. Familial central precocious puberty suggests autosomal dominant inheritance. *J Clin Endocrinol Metab*. 2004; 89: 1794–1800.
38. Shin Y. An update on genetic causes of central precocious puberty. *Annals of Pediatric Endocrinology & Metabolism*. 2016; 21: 66-69.
39. FlexiGene DNA Handbook for purification of DNA from human whole blood, buffy coat, cultured cells, 2010.
40. Dauber A, Cunha-Silva M, Macedo DB, Brito VN, Abreu AP, Roberts SA, Montenegro LR, Andrew M, Kirby A, Weirauch MT, Labilloy G, Bessa DS, Carroll RS, Jacobs DC, Chappell PE, Mendonca BB, Haig D, Kaiser UB, Latronico AC. Paternally Inherited DLK1 Deletion Associated With Familial Central Precocious Puberty. *JCEM*. 2017, 102(5): 1557-1567.
41. Fuqua SJ. Treatment and Outcomes of Precocious Puberty: An Update. *J Clin Endocrinol Metab*. 2013; 98(6): 2198-2207.
42. Chemaitilly W, Trivin C, Adan L, Gall V, Sainte-Rose C, Brauner R. Central precocious puberty: clinical and laboratory features. *Clinical Endocrinology*. 2001; 54: 289–294.

43. Kellis M, Wold B, Snyder MP, Bernstein BE, Kundaje A, Marinov GK, Ward LD, Birney E, Crawford GE, Dekker J, Dunham I, Elnitski LL, Farnham PG, Feingold EA, Gerstein M, Giddings MC, Gilbert DM, Gingeras TR, Green ED, Guigo R, Hubbard T, Kent J, Lieb JD, Mayers MJ, Pazin MJ, Ren B, Stamatoyannopoulos JA, Weng Z, White KP, Hardison RS. Defining functional DNA elements in the human genome. *PNAS*. 2014, 111(17): 6131-6138.
44. Chi KR. The dark side of the human genome. *Nature*. 2016, 538: 275-277.
45. Moorthie S, Mattocks CJ, Wright CF. Review of massively parallel DNA sequencing technologies. *Hugo J*. 2011, 5(1-4): 1-12.
46. https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf. (Dostopno: 8.11. 2017).
47. <http://www.atdbio.com/content/58/Next-generation-sequencing>. (Dostopno: 8.11. 2017).
48. Novogene poročilo izvedbe priprave knjižnice DNA in WGS. file:///E:/Primary_Report.html. (Dostopno: 17.8. 2017).
49. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform[J]. *Bioinformatics*. 2009, 25(14): 1754-1760.
50. Royer-Bertrand B, Rivolta C. Whole genome sequencing as a means to assess pathogenic mutations in medical genetics and cancer. *Cell. Mol. Life Sci*. 2015, 72: 1463-1471.
51. Lee D, Hormozdiary F, Xin H, Hach F, Mutlu O, Alkan C. Fast and accurate mapping of Complete Genomic reads. *Methods*. 2014, 1-8.
52. dbSNP. <https://www.ncbi.nlm.nih.gov/SNP/>. (Dostopno: 8.11. 2017).
53. ExAc. <http://exac.broadinstitute.org/>. (Dostopno: 8. 11. 2017).
54. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*. 2015, 526: 68-74.
55. Zhu X, Wang J, Peng P, Shete S. Empirical estimation of sequencing error rates using smoothing splines. *BMC Bioinformatics*. 2016, 17:177-188.
56. Richterich P. Estimation of Errors in "Raw" DNA Sequences: A Validation Study. *Genome Res*. 1998, 8(3): 251-259.

57. <https://www.illumina.com/content/dam/illumina-marketing/documents/products/technotes/hiseq-phix-control-v3-technical-note.pdf>. (Dostopno: 23.9. 2017).
58. User Guided: Qubit® ds DNA BR Assay Kits. <https://www.thermofisher.com/order/catalog/product/Q32850?SID=srch-srp-Q32850>. (Dostopno: 17.8.2017).
59. Illumina TrueSeq® Nano DNA Library Prep Reference Guide. https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_truseq/truseqnanodna/truseq-nano-dna-library-prep-guide-15041110-d.pdf. (Dostopno: 17.8. 2017).
60. Aparat Agilent 2100. <http://www.agilent.com/en/promotions/2100-bioanalyzer>, Agilent DNA Kit Guide. http://www.agilent.com/cs/library/usermanuals/public/G2938-90321_SensitivityDNA_KG_EN.pdf. (Dostopno: 9. 11. 2017).
61. <http://www.metagenomics.wiki/pdf/definition/coverage-read-depth>. (Dostopno: 13.8. 2017).
62. Oliveros CJ. BioinfoGP. Venny 2.1. 2015. <http://bioinfogp.cnb.csic.es/tools/venny/>. (Dostopno: 12.5. 2017).
63. CADD algoritm. <http://cadd.gs.washington.edu/score>. (Dostopno: 9.11. 2017).
64. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet.* 2014, 46(3): 310-315.
65. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* 1977, 74(12): 5643-5467.
66. Shendure J, Ji H. Next-generation DNA sequencing . *Nat Biotechnol.* 2008, 26(10): 1135-1145.
67. <https://www.gatc-biotech.com/en/expertise/sanger-sequencing.html>. (Dostopno: 18.9. 2017).
68. Ensembl Genome Browser. <http://www.ensembl.org./index.html>. (Dostopno: 15.9. 2017).

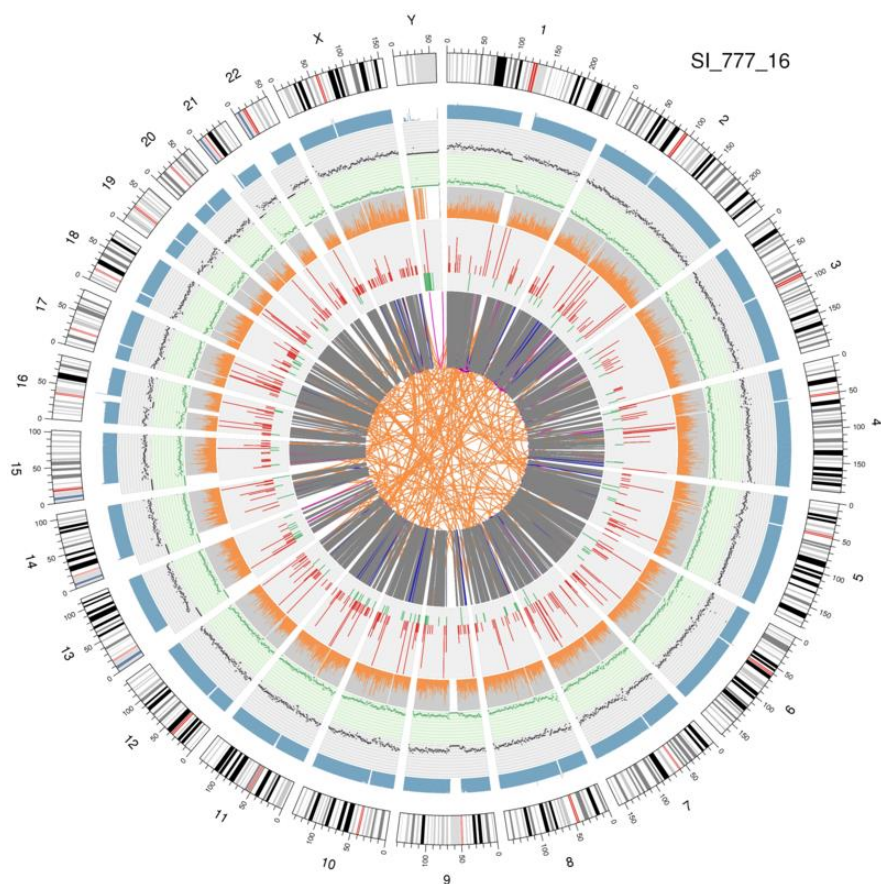
69. BAM mapa. <https://samtools.github.io/hts-specs/SAMv1.pdf>. (Dostopno: 25.9. 2017).
70. BLAST (Basic Local Alignment Search Tool).
<http://blast.ncbi.nlm.nih.gov/Blast.cgi>. (Dostopno: 14. 8. 2017).
71. Podatkovna baza GenBank.
https://www.ncbi.nlm.nih.gov/nucore/NM_001128223.1. (Dostopno: 14.8. 2017)
72. Gen *ZNF717*. <http://www.genecards.org/cgi-bin/carddisp.pl?gene=ZNF717>.
(Dostopno: 10.5. 2017.)
73. Gen *ZNF717*. <http://www.genecards.org/cgi-bin/carddisp.pl?gene=ZNF717>.
(Dostopno: 10.11. 2017).
74. Gen *ZNF717*.
<https://www.ncbi.nlm.nih.gov/ieeb/research/acembly/av.cgi?db=human&c=Gene&l=ZNF717>. (Dostopno: 10.5. 2017).
75. Winkler EC, Wiemann S. Findings made in gene panel to whole genome sequencing: data, knowledge, ethics – and consequences? *Expert Rev Mol Diagn.* 2016, 16(12):1259-1270.
76. Dorschner DO, Amendola LM, Turner EH, Robertson PD, Shirts BH, Gallego CJ, Bennett RL, Jones KL, Tokita MJ, Bennett JT, Kim JH, Rosenthal EA, Kim DS; Tabor HK, Bamshad MJ, Motulsky AG, Scott CR, Pirtchard CC, Walsh T, Burke W, Raskind WH, Byers P, Hisama FM, Nickerson DA, Jarvik GP. Actionable, Pathogenic Incidental Findings in 1,000 Participants' Exomes. *Am j Hum Genet.* 2013, 93(4): 631-640.
77. Goldfeder RL, Priest JR, Zook JM, Grove ME, Waggott D, Wheeler MT, Salit M, Ashley EA. Medical implications of technical accuracy in genome sequencing. *Genome Medicine.* 2016, 8(24).
78. Dewey FE, Grove ME, Pan C. Clinical Interpretation and Implications of Whole Genome Sequencing. *JAMA.* 2014, 311(10): 985-987.
79. Chin ELH, Silva da C, Hegde M. Assessment of clinical analytical sensitivity and specificity of next-generation sequencing for detection of simple and complex mutations. *BMC Genet.* 2013, 14(6): 1-13.

80. Sims D, Sudberg I, Illot NE, Heger A, Ponting CP. Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics*. 2014, 15: 121-132.
81. Variation detection sensitivity and biases in whole genome and exome sequencing. *BMC Bioinformatics*. 2014, 15(1): 247-258.
82. Mu W, Lu H, Chen J, Li S, Elliot AM. Sanger Confirmation is Required to Achieve Optimal Sensitivity and Specificity in Next Generation Sequencing Panel Testing. *The Journal of Molecular Diagnostics*. 2016, 18(6): 923-932.
83. Glusman G, Severson A, Dhankani V, Robinson M, Farrah T, Mauldin DE, Stittrich SA, Roach JC, Brunkow ME, Bodian DL, Vockley JG, Schmulevich I, Niederhuber JE, Hood L. Identification of copy number variants in whole-genome data using Reference Coverage Profiles. *Front Genet*. 2015, 6:45-58.
84. Meienberg J, Bruggmann R, Oexle K, Matyas G. Clinical sequencing: is WGS the better WES? *Hum Genet*. 2016, 135: 359-362.
85. Li W, Olivier M. Current analysis platforms and methods for detecting copy number variants. *Physiol Genomics*. 2013, 45(1): 1-16.
86. Zarrei M, MacDonald MR, Merico D, Scherer SW. A copy number variation map of the human genome. *Nature Reviews Genetics*. 2015, 16:172-183.
87. Tattini L, Aurizio D' R, Magi A. Detection of Genomic Structural Variants from Next Generation Sequencing Data. *Front Bioeng Biotechnol*. 2015, 3:92-100.
88. Vaz-Drago R, Custodio N, Carmo-Fonseca M. Deep intronic mutations and human disease. *Hum Genet*. 2017.
89. Allis CD, Jenuwein T. The molecular hallmarks of epigenetic control. *Nature Reviews Genetics*. 2016, 17:487-500.
90. Illumina Epigenetic Analysis Techniques.
<https://www.illumina.com/techniques/popular-applications/epigenetics.html>.
(Dostopno: 15.11. 2017).

8 PRILOGA

Priloga 1: Rezultat, ki ga dobimo po bioinformatički obdelavi podatkov pridobljenih z metodo WGS, je krog sestavljen iz 7 obročev (glej *Slika 15*). Krog prikazuje podatke o človeškem genomu, in sicer če ga gledamo od zunaj proti notranjosti nam o genomu poda naslednje informacije:

- 1_ Na prvem obroču (najbolj zunanji) so podatki o kromosomih.
- 2_ Na drugem obroču so izrisani histogrami, ki prikazujejo pokritost zaporedja gDNA, ki smo jo zajeli z metodo WGS. En histogram pomeni povprečno pokritost območja velikega 0,5 Mbp.
- 3_ Tretji obroč nam pove, kakšna je gostota in razpršenost genetskih sprememb vrste INDEL . Ena črna pika predstavlja eno spremembo INDEL v razponu 1 Mbp.
- 4_ Četrti obroč, nam pove kakšna je gostota in razpršenost genetskih sprememb vrste SNV. Ena zelena pika predstavlja eno spremembo SNV v razponu 1 Mbp.
- 5_ Na petem obroču vidimo histograme, ki prikažejo razmerje med homozigotnimi SNV in heterozigotnimi SNV. En histogram predstavlja območje veliko 0,5 Mbp.
- 6_ Šesti obroč prikazuje prisotne genetske spremembe vrste CNV, rdeča pomeni pridobitev in zelena izgubo števila kopij delov zaporedja.
- 7_ Na sedmem obroču (najbolj notranji) vidimo interference genetskih sprememb vrste SV v eksonih in spojivnih območjih zaporedja gDNA. Premestitve so prikazane z oranžno barvo, insercije z zeleno, delecije s sivo, podvojitve z rožnato in obrati modro barvo.



Slika 15. Rezultat bioinformatične obdelave podatkov pridobljenih z metodo WGS (48).