

**Univerzita Karlova v Praze**

**Přírodovědecká fakulta**

Studijní program: Speciální chemicko-biologické obory

Studijní obor: Molekulární biologie (BMOBIBO)



**Tereza Kmochová**

Současné metody a možnosti výzkumu vzácných geneticky podmíněných  
onemocnění

Characterisation of rare genetic diseases using genomic techniques

Bakalářská práce

Školitel: Ing. Kateřina Hodaňová, PhD.

Praha, 2015



**Poděkování a prohlášení:**

Chtěla bych poděkovat své školitelce, Ing. Kateřině Hodaňové, PhD a pracovníkům z Ústavu dědičných metabolických poruch 1. LF UK za odborné vedení, cenné rady a pomoc při sepsování této bakalářské práce. Zároveň chci poděkovat celé své rodině za podporu.

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně, a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, 15. 5. 2015

Podpis: .....

Tereza Kmochová

## **Abstrakt**

V této bakalářské práci je popsán význam studia vzácných geneticky podmíněných onemocnění a jsou zhodnoceny současné metody a možnosti určení jejich genetických příčin.

Úvod se zabývá definicí vzácných onemocnění a důvody, proč je studujeme. V současnosti je známo přibližně 8000 nemocí, které jsou způsobeny funkčně závažnými mutacemi v jednom či několika genech. Klinický projev geneticky podmíněného onemocnění a výsledky běžných funkčních a biochemických vyšetření neumožňují ve většině případů přesné určení diagnózy. Neznalost diagnózy a příčin nemoci snižuje kvalitu života pacientů a jejich rodin. Určení genetických příčin umožňuje přesnou klasifikaci studovaných případů a je základním východiskem pro kvalifikované genetické poradenství a prevenci vzácných nemocí v rodinách.

Další kapitoly popisují metody analýzy DNA (sekvenování DNA různými metodami), možnosti, jak postupně najít genetickou příčinu nemoci a ukazují základní přístupy ke studiu funkčních dopadů mutací.

Praktické využití těchto metod je ukázáno na případě dvou sourozenců s opožděným psychomotorickým vývojem, zvláštní kvalitou vlasů (*pili torti*), poruchou sluchu a hyperurikemickou nefropatií, u kterých genetická analýza odhalila jako příčinu nemoci závažnou mutaci v genu *RMND1*.

**Klíčová slova:** vzácná onemocnění, analýza DNA, opožděný psychomotorický vývoj, porucha sluchu, hyperurikemická nefropatie, *pili torti*, *RMND1*

## **Abstract**

In this thesis the importance of studying genetically determined rare diseases is described and current methods and options to determine their genetic causes are evaluated.

The introduction describes the definition of rare diseases and the reasons why we study them. Currently it is known about 8,000 diseases which are caused by functionally serious mutations in one or more genes. The clinical symptoms of genetically caused rare diseases and the results of common functional and biochemical tests do not allow the exact diagnosis in most cases. The absence of the diagnosis and the causes of diseases reduces the quality of life of patients and their families. The determination of genetic causes allows accurate classification of studied cases and it is a base for qualified genetic counseling and prevention of rare diseases in the families.

The next chapters describe the methods of DNA analysis (DNA sequencing using different methods), the possibilities of finding the genetic cause of a disease and show the basic approaches to the study of functional impact of the mutations.

The practical use of these methods is shown in the case of two siblings with delayed psychomotoric development, special hair quality (*pili torti*), hearing impairment and hyperuricemic nephropathy in which the genetic analysis revealed the mutation in the gene *RMND1* as the cause of the disease.

**Key words:** rare diseases, DNA analysis, delayed psychomotoric development, hearing impairment, hyperuricemic nephropathy, *pili torti*, *RMND1*

## **Seznam zkratk:**

CNV – duplikace (Copy Number Variations)

dbSNP – databáze jednonukleotidových polymorfismů (The Single Nucleotide Polymorphism database)

DNA – deoxyribonukleová kyselina (deoxyribonucleic acid)

ExAC – databáze „The Exome Aggregation Consortium“

ESP – „Exome Sequencing Project“

IGV – počítačový program Integrative Genome Viewer

kDa – kilodalton; atomová hmotnostní jednotka

NCBI – Národní centrum pro biotechnologické informace (The National Center for Biotechnology Information)

NHLBI – „National Heart, Lung and Blood Institute“

NGS – Nové metody sekvenování (New Generation Sequencing)

OMIM – databáze „Online Mendelian Inheritance in Man“

PCR – polymerázová řetězová reakce (Polymerase Chain Reaction)

RMND1 – mitochondriální protein (Required for Meiotic Nuclear Division 1 homolog (*S. cerevisiae*))

*RMND1* – gen (Required for Meiotic Nuclear Division 1 homolog (*S. cerevisiae*))

SNP – jednonukleotidový polymorfismus (Single Nucleotide Polymorphism)

SNV – jednonukleotidové varianty (Single Nucleotide Variants)

STR – mikrosatelity (Short Tandem Repeats)

## OBSAH

1. Úvod .....	1
2. Definice vzácných geneticky podmíněných onemocnění a jejich prevalence .....	2
3. Význam studia vzácných onemocnění .....	3
4. Příčiny vzniku vzácných onemocnění .....	4
4.1. Mutace .....	4
4.2. Mutační rychlost .....	5
5. Hledání genetických příčin vzácných onemocnění .....	7
5.1. Počátky molekulární biologie .....	7
5.2. Molekulární medicína .....	8
6. Nástroje studia příčin vzácných geneticky podmíněných onemocnění .....	9
6.1. Klinická charakterizace .....	9
6.2. Biochemická genetika .....	10
7. Genetické mapování a poziční klonování .....	11
8. Mapování a sekvence lidského genomu .....	12
8.1. Sekvenování genomů a vznik databází .....	12
8.2. Nové metody sekvenování DNA .....	13
8.3. Cílené sekvenování .....	14
8.4. Exomové sekvenování .....	14
8.5. Genetická variabilita člověka .....	14
9. Praktická část .....	15
9.1. Objasnění genetické příčiny hyperurikemické nefropatie u dvou sourozenců .....	15
10. Diskuse .....	21
11. Závěr .....	22
Seznam použité literatury.....	23

## 1. Úvod

Vzácná onemocnění jsou skupinou přibližně 8000 různých nemocí, jejichž společnou charakteristikou je vzácnost výskytu v populaci. Tato onemocnění jsou ze své definice jednotlivě vzácná a populačně specifická. Jako skupina jsou však vzácná onemocnění relativně častá a postihují přibližně 10% populace. Většina vzácných onemocnění vede k závažnému zdravotnímu postižení, které se projevuje převážně v dětském věku nebo dospívání. Závažná zdravotní postižení způsobená onemocněním v 8% případech vedou až ke smrti pacienta (<http://www.vzacnenemoci.cz/sites/default/files/narodni-strategie.pdf>). S prodlužující se délkou života se zvyšuje i počet vzácných nemocí, které se projevují až v dospělosti či stáří. Určení příčiny nemoci je v každém jednotlivém případě základním východiskem pro správnou léčbu, kvalifikované genetické poradenství a prevenci v rodině. Určení diagnózy je též významným faktorem pro získání odpovídající sociální a zdravotnické podpory a začlenění pacienta do společnosti. Omezená znalost a zkušenost s jednotlivými typy vzácných onemocnění, různorodost jejich klinických projevů a malá specifická běžně prováděných funkčních a biochemických vyšetření bohužel ve většině případů neumožňují přesné určení diagnózy. Neznalost příčin nemoci tak dlouhodobě snižuje kvalitu života mnoha pacientů a jejich rodin a představuje závažný zdravotnický a sociální problém. Tento neuspokojivý stav může částečně změnit větší informovanost odborné i laické veřejnosti o problematice vzácných onemocnění, intenzivní výzkum vzácných onemocnění a zejména využívání nových diagnostických metod založených na analýze genetické informace postižených jedinců a jejich rodin.

Ve své práci se proto snažím definovat skupinu vzácných onemocnění, zdůraznit význam studia jednotlivých nemocí, přinést přehled příčin jejich vzniku a demonstrovat nové možnosti a metody jejich efektivní diagnostiky, charakterizace a léčby.

V praktické části své práce prezentuji své vlastní zkušenosti na případu vzácného geneticky podmíněného onemocnění v jedné španělské rodině. K odhalení kauzální mutace, která je příčinou onemocnění v této rodině jsem využila současných metod a možností analýzy DNA.



## **2. Definice vzácných geneticky podmíněných onemocnění a jejich prevalence**

Jako vzácné onemocnění definujeme nemoci s chronickým, progresivním a degenerativním projevem, které postihují malé procento populace. Onemocnění jsou většinou způsobena vrozenou či získanou změnou genetické informace a projevují se změnou na fenotypu a normálních biologických funkcích. Příznaky onemocnění se objevují nejčastěji v dětském věku a jsou většinou fatální. Počet pacientů postižených se vzácným onemocněním v Evropské Unii je méně než jedna osoba na 2000 osob a ve Spojených státech amerických je postižena méně než jedna osoba na 1250 osob (Schieppati *et al.*, 2008). Počty pacientů se u různých typů nemocí výrazně liší. U těch nejvzácnějších onemocnění může být postiženo jen pár desítek lidí na celém světě ([www.eurordis.org/IMG/pdf/princeps\\_document-EN.pdf](http://www.eurordis.org/IMG/pdf/princeps_document-EN.pdf)). Celkový počet pacientů postižených vzácným onemocněním je odhadován přibližně na 30 milionů v Evropě, 25 milionů v Severní Americe (Schieppati *et al.*, 2008) a celosvětově je postiženo až 10% populace. Jako skupina jsou proto vzácná onemocnění relativně častá.

### 3. Význam studia vzácných onemocnění

Vzácná onemocnění představují významný zdravotnický a sociální problém. U většiny pacientů se onemocnění projevují závažným zdravotním postižením, jehož příčiny nejsou ošetřujícím lékařům, pacientům a jejich rodinám bezprostředně známy. Ve snaze nalézt příčiny nemoci jsou pacienti opakovaně hospitalizováni a vyšetřováni řadou specialistů. Nadměrná lékařská péče představuje pro pacienta a jeho rodinu značnou fyzickou, psychickou, ekonomickou i časovou zátěž. Pokud toto úsilí nevede dlouhodobě k určení diagnózy či zlepšení zdravotního stavu, ztrácí pacient i jeho rodina důvěru v systém zdravotní péče a vyhledává pomoc na jiných místech. Neznalost příčin nemoci vede často i k nevhodným způsobům léčby a brání zajištění odpovídající sociální a zdravotnické podpory a začlenění pacienta do společnosti. Neznalost příčin nemoci též limituje možnosti genetického poradenství, brání prevenci onemocnění formou prenatální či preimplantační diagnostiky a významně ovlivňuje reprodukční rozhodování rodičů či pacientů. To vše pak ve svém důsledku vede ke stigmatizaci a významnému snižování kvality života pacientů a jejich bezprostředního okolí. Nezanedbatelné nejsou ani náklady, které společnost vydává na nadměrnou či nesprávnou zdravotní péči o tuto skupinu pacientů.

Jedinou cestou k určení příčin vzácných onemocnění je intenzivní studium jednotlivých případů a výzkum jednotlivých skupin nemocí na národních či mezinárodních úrovních. Za tímto účelem jsou v řadě států budovány a podporovány specializované diagnostické a výzkumné laboratoře a je připravována a prováděna řada strategických opatření na národní úrovni (např. Národní strategie pro vzácná onemocnění v České republice na léta 2010-2020; [http://svp-vzacnaonemocneni.cz/portal/wpcontent/uploads/narodni\\_strategie\\_pro\\_vzacna\\_onemocneni.pdf](http://svp-vzacnaonemocneni.cz/portal/wpcontent/uploads/narodni_strategie_pro_vzacna_onemocneni.pdf)) i mezinárodní úrovni (např. Orphanet (<http://www.orpha.net>), National Plans and Strategies on Rare Diseases in Europe, EUROPLAN) (Taruscio *et al.*, 2010; Taruscio *et al.*, 2012). Diagnostika jednotlivých případů a studium vzácných onemocnění má též významný přesah do základního biomedicínského výzkumu a poznávání biologie člověka. Objasnění příčin studované nemoci odhaluje funkce lidských genů, odkrývá nové biochemické a regulační dráhy, objasňuje základní biologické, buněčné a molekulární příčiny nemocí a definuje potenciální terapeutické cíle. Tyto znalosti, které nelze získat jinak, než studiem vzácných nemocí, tak mají bezprostřední přesah k pochopení příčin, mechanismům vzniku i léčby a prevence populačně častých komplexních nemocí.

#### 4. Příčiny vzniku vzácných onemocnění

Přesná příčina vzniku vzácného onemocnění není ve většině případů známá. Příčiny onemocnění jsou ve většině případů geneticky podmíněné, mohou být však způsobeny vlivem vnějšího prostředí nebo kombinací těchto faktorů.

Zhruba 80% vzácných onemocnění má jednoznačně genetické příčiny (Dodge *et al.*, 2010). Geneticky podmíněná onemocnění vznikají v důsledku přítomnosti jedné nebo více funkčně závažných mutací v genomu postiženého jedince (kauzální mutace). V závislosti na rozsahu mutací rozeznáváme genomové, chromozomové, genové a regulační mutace. Genomové mutace zahrnují změny počtu chromozomů (polyploidie, aneuploidie). Chromozomové mutace mění strukturu jednotlivých chromozomů (translokace, rozsáhlé delece či duplikace – „copy number variations, CNVs“). Genové mutace ovlivňují strukturu a sekvenci a mají vliv na funkci jednoho genu (delece, inserce, duplikace, inverze, bodové mutace, apod.) Regulační mutace jsou lokalizovány v oblastech genomu, které ovlivňují správnou funkci genomu či genu (promotorové oblasti, enhancery, silencery, oblasti epigenetických modifikací).

Negenetické příčiny vzácných onemocnění mohou vzniknout během početí nebo vývoje či během poporodní adaptace postiženého jedince. V průběhu těhotenství může být vážným problémem imunologická inkompatibilita dítěte a matky, způsob života matky, působení faktorů vnějšího prostředí nebo sociálních faktorů. Předčasný porod a poporodní adaptace (např. poruchy acidobazické rovnováhy, asfyxie, novorozenecká žloutenka) též mohou být příčinou vzácného onemocnění. Další příčiny mohou vzniknout během časného vývoje dítěte, kdy je dítě vystaveno nějaké infekci (bakterie, viry), otravě toxickou látkou (působení nějakých léků), úrazu (trauma) nebo dětskému zneužívání (šikana). I neúplná výživa dítěte (malnutrice) může být významným faktorem, který se projeví ve vzácné onemocnění.

##### 4.1 Mutace

Geneticky podmíněná onemocnění vznikají v důsledku závažných mutací v lidském genomu, které významně ovlivňují správnou funkci organismu.

Mutace jsou trvalé změny v genetické informaci jedince a jsou důsledkem přirozené mutability lidského genomu během evoluce lidstva a života jedince. Mutace vznikají působením různých faktorů - fyzikálních faktorů (ionizující záření, ultrafialové záření), chemických faktorů (planární aromatické sloučeniny, silné oxidanty, bojové látky – yperit, apod.), biologických

faktorů (virové infekce, transpozibilní elementy, retrotranspozony) nebo při chybách v replikačním a reparačním procesu DNA. K významným změnám chromozomálních struktur dochází též při mitóze nebo meióze (chromozomální aberace, translokace, rozsáhlé delecce či amplifikace, rekombinace). K mutacím dochází jak v somatických, tak i v zárodečných buňkách každého jedince (Griffiths *et al.*, 2000). Vznik mutací je základním zdrojem genetické variability člověka a jsou jedním z nástrojů evolučního vývoje člověka. Řada z nově vzniklých genetických variant (alel) mění bezprostředně po svém vzniku vlastnosti svého nositele (buňky nebo organismu). Tyto nově nabyté vlastnosti nemusí mít na svého nositele bezprostřední nebo zásadní vliv (neutrální mutace). Některé mutace však mohou být evolučně výhodné (např. odolnost proti infekčním nemocem). Jiné mutace mohou být neslučitelné se životem (letální mutace) nebo zhoršují biologickou zdatnost nositele a projevují se formou nemoci. Míra evoluční prospěšnosti či neúspěšnosti jednotlivých mutací se odráží v jejich populačních frekvencích. Evolučně výhodné mutace postupně dosahují v populacích vysokých frekvencí. Naopak letální mutace a mutace zhoršující biologickou zdatnost nositele jsou v populaci vzácné, případně jejich frekvence zůstává trvale nízká. Frekvence těchto vzácných variant (alel) se však může významně změnit v některých populacích mechanismem genetického driftu (Lande, 1976), např. efekt hrdla lahve, kdy nastane prudký pokles jedinců v populaci na méně než 50% a dojde ke snížení genetické variability, která s postupem času zase roste (Hawks *et al.*, 2000) nebo efekt zakladatele, kdy vznikne zcela nová populace z velmi malé skupiny jedinců (Templeton, 1980).

## 4.2 Mutační rychlost

Rychlost vzniku mutací v lidském genomu byla a je předmětem řady úvah a studií. J. B. S. Haldane pronesl hypotézu, že mužská zárodečná linie (spermatogeneze) je více mutagenní než ženská zárodečná linie (oogeneze) (Haldane, 1947). Mutační rychlost je specifická pro různé populace a odvíjí se od řady různých parametrů. Hlavním důkazem mutační specifity je, že barva pleti člověka závisí na síle ultrafialového záření. Frekvence této mutace (5'-TCC-3' → 5'-TTC-3') v evropské populaci vzrostla o 50% od doby rozdělení Evropanů od Asiatů před 50 tisíci lety a dnes je nejčastější somatickou mutací způsobující maligní melanom (Harris, 2015). Mutační rychlost je rozdílná mezi živočišnými druhy (Kumar *et al.*, 2002; Taylor *et al.*, 2006), populacemi (Fu *et al.*, 2014; Harris, 2015), rodinami, jednotlivci, pohlavím (Haldane, 1947), buněčnými typy (Tomlinson *et al.*, 1996), jednotlivými chromozomy (McVean *et al.*, 1997), rekombinačními místy (Lercher *et al.*, 2002), různými

genomovými oblastmi (Wolfe *et al.*, 1989), geny a nukleotidy (Nachman *et al.*, 2000), a také závisí na stáří organismu (Kong *et al.*, 2012; Mathieson *et al.*, 2014). Mezi druhové studie, studie mezi historicky vzdálenými jedinci, studie v rodinách (*de novo* varianty) studie v jednotlivých populacích, studie v buněčných klonách apod. ukazují, že mutační rychlost jednonukleotidových variant SNV („single nucleotide variants“) se u člověka pohybuje v rozmezí  $0,96-2,17 \times 10^{-8}$  nukleotidů/genom (Campbell *et al.*, 2013). Na základě těchto studií můžeme odhadnout frekvenci vzniku cca 70 *de novo* mutací na každého nově narozeného jedince (Keightley, 2012). Ohromný počet nových mutací vznikl díky exponenciálnímu populačnímu růstu, ke kterému došlo během posledních 10 tisíc let. Obrovské rozdíly v úrovních genetické variability mezi jednotlivými populacemi jsou důsledkem postupné kolonizace planety a osídlování řady oblastí během posledních 700 let. Lidská genetická variabilita je původní a sdílená, ale obsahuje mnoho nových a vzácných mutací, které jsou spojeny se vznikem a existencí řad Mendelovských a komplexních onemocnění.

## 5. Hledání genetických příčin vzácných onemocnění

Genetické příčiny onemocnění v moderním lidstvu vznikly postupně se stavbou a evolucí lidského genomu. Nejstarší lidské alely se objevily před několika miliony lety, současně s postupným vývojem lidského druhu na africkém kontinentu (Cavalli-Sforza *et al.*, 2003). Zhruba 90% celé genetické variability v dnešní populaci lidí je složeno a sdíleno z těchto původních genetických polymorfismů (McClellan *et al.*, 2010).

Neobvyklé projevy vzácných onemocnění, přirozená lidská zvědavost a touha po poznání vždy vedly ke snahám o objasnění jejich příčin. Ve starověkém Egyptě a Mezopotámii lidé začali více chápat, že každá nemoc má svou příčinu a snažili se určit diagnózu, prognózu, etiologii a nalézt terapii. Díky pečlivému pozorování a záznamům byly již v Talmudu a Bibli popsány příznaky nadměrného krvácení - hemofilie. V 10. století arabský vědec Albucasis poprvé identifikoval dědičnou povahu hemofilie, když si všiml souvislosti v úmrtí několika příbuzných chlapců, kteří všichni podleli vykrvácení po menším zranění. (<https://www.hemophilia.org/Bleeding-Disorders/History-of-Bleeding-Disorders>). Lékaři si začali uvědomovat, že se řada onemocnění dědí a přenáší do dalších generací, a že musí existovat nějaký mechanismus, jak se faktory způsobující onemocnění předávají od rodičů svým potomkům. V 19. století moderní medicína začala dostávat určitý řád, když lékaři začali vytvářet a využívat mnohem systematictější postupy analýz podle pacientových příznaků. Kvalita popisu diagnóz se postupně zlepšovala s tím, jak se měnila dostupná technika, která umožňovala lékařům lépe a podrobněji zkoumat a porovnávat klinické projevy nemocí s patologickými a chemickými změnami ve tkáních a tělních tekutinách pacientů. Postupně se začaly využívat nové fyzikálně – chemické, makroskopické a mikroskopické metody, které umožnily podrobnější studium buněk, jejich struktur a vlastností. Díky těmto novým metodám vznikaly největší objevy, které daly základy molekulární biologie a genetickému oboru.

### 5.1 Počátky molekulární biologie

Darwinova teorie o přírodním výběru (Darwin, 1859) byla částečně a nezávisle na sobě vysvětlena Mendelovským modelem dědičnosti (Mendel *et al.*, 1965). Friedrich Miescher v roce 1869 poprvé izoloval směs DNA, jejíž význam neznal, ale nazval ji nuklein (Dahm, 2005). Walter Flemming studoval buněčné jádro a objevil strukturu chromozomů – chromatiny a popsal buněčné dělení – mitózu (Flemming, 1879). Nebyl ale obeznámen s Mendelovými zákony, tudíž si svá pozorování nespojoval s genetickou dědičností.

Znovuobjevení Mendelových zákonů (Bateson, 1901), lepší porozumění a pochopení základních buněčných funkcí dovedlo Waltera Suttona k Boveri-Suttonově teorii, která popisuje, že na chromozomech organismů jsou uloženy dědičné faktory podle Mendelových zákonů (Sutton, 1902). Tuto domněnku potvrdil Thomas Hunt Morgan, díky pokusům s křížením octomilek a následným potvrzením teorie, že dědičné faktory neboli geny jsou skutečně umístěny na chromozomech (Morgan, 1911). Využívání krystalografických technik umožnilo popsat první struktury proteinů a určit vztah mezi strukturou DNA a základním mechanismem uchování a dědičností genetické informace. Oswald Avery a jeho experiment potvrdil, že genetická informace je uložena v molekule DNA a nikoli v proteinech, jak se dříve předpokládalo (Avery *et al.*, 1944). To vedlo k objevu dvoušroubovice DNA Francisem Crickem a Jamesem Watsonem (Watson *et al.*, 1953) a definici Centrálního dogmatu molekulární biologie Francisem Crickem v roce 1958 (Crick, 1970). Na konci padesátých let se mnoho významných vědců na celém světě snažilo o navržení elegantní metody translace DNA do proteinů podle dogmatu. Jako prvním se to podařilo Marshallu Nirenbergovi a jeho kolegům, kdy se jim podařilo rozluštit genetický kód (Nirenberg *et al.*, 1966). Počátky sekvenování DNA začaly v roce 1977, díky objevům a následným kombinováním různých metod – Sangerova dideoxy metoda (Sanger *et al.*, 1977) a Maxam-Gilbertova metoda (Maxam *et al.*, 1977). Do roku 1983 existoval jediný způsob, namnožení určitého úseku DNA pomocí klonování v bakteriálním plasmidu, než Kary Mullis objevil metodu polymerázové řetězové reakce (PCR), která umožňuje amplifikovat téměř jakýkoliv úsek DNA (Mullis *et al.*, 1994). Objev PCR byl zásadním zlomem pro vývoj sekvenování DNA a pro nalezení různých genetických variabilit člověka na úrovni jednotlivých nukleotidů.

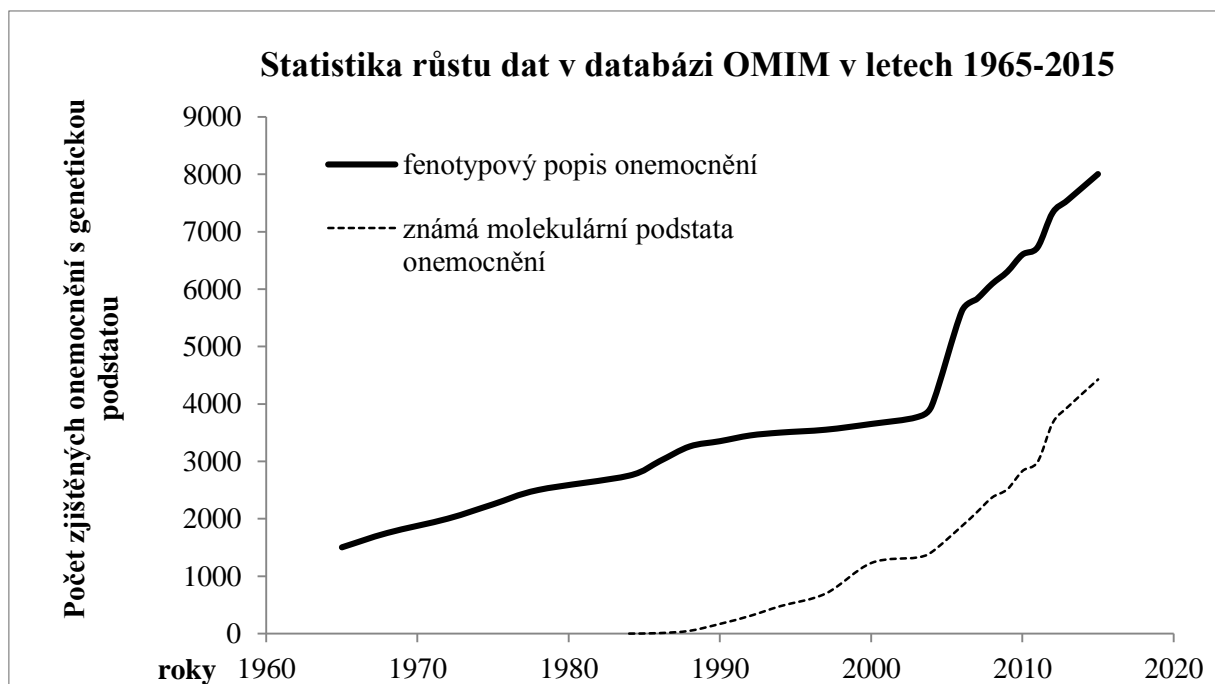
## **5.2 Molekulární medicína**

Pochopení základních fyzikálních, biologických, biochemických a molekulárních procesů a mechanismů v buňce umožnilo začít identifikovat molekulární a genetické příčiny nemocí. V roce 1949 Linus Pauling jako první prokázal příčinu vzácného onemocnění (srpkovitá anémie) na molekulární úrovni (Pauling *et al.*, 1949). Vznikl tak obor molekulární medicíny, která využívá různé biochemické, genetické imunologické a biotechnologické nástroje pro zkoumání příčin onemocnění, která jsou primárně způsobená chemickými změnami ovlivňující strukturu, funkci a vlastnosti molekul (Hunt *et al.*, 1958).

## 6. Nástroje studia příčin vzácných geneticky podmíněných onemocnění

### 6.1 Klinická charakterizace

Pro určování genetických příčin vzácných onemocnění je zásadní podrobná klinická charakterizace onemocnění popsáním jeho znaků (fenotypů), průběhu a jeho možného dalšího vývoje (progrese). Klinický popis se poté publikuje ve formě článků pro odbornou veřejnost. Abstrakta, klíčová slova a autoři těchto vědeckých článků jsou vyhledatelní pomocí řady databází, např. PubMed, Google Scholar, Scopus, atd. Stejně tak jsou vyhledatelné i články, které původní informace citují. Zásadní pro shromáždění všech charakteristik vzácných genetických onemocnění bylo vytvoření databáze Online Mendelian Inheritance in Man (OMIM), do níž stále přibývají další případy (viz Obr. 1). Databáze OMIM shromažďuje všechna dosud známá geneticky způsobená onemocnění s jejich popisem fenotypu a navádí přímo na relevantní geny v lidském genomu a umožňuje další výzkum a poskytuje nástroje pro další genomické analýzy kauzálního genu (Hamosh *et al.*, 2005).



**Obr. 1. Statistika růstu dat v databázi OMIM v letech 1965-2015** (zdroj dat: Ústav dědičných metabolických poruch, 1. LF UK)



## 6.2 Biochemická genetika

Mimo prostého klinického popisu se pro studium onemocnění začaly postupně používat nové analytické metody. Analýzy probíhaly na anatomické úrovni (orgány, tkáně), mikroskopické úrovni (buňky a buněčné struktury) a biochemické úrovni. Tyto metody umožnily klasifikaci onemocnění do různých skupin – například onemocnění charakteristická hromaděním nějakého abnormálního materiálu v buňkách a tkáních, organelová onemocnění charakterizovaná absencí či změnami velikosti organel, např. peroxizomů nebo např. lyzozomální onemocnění či onemocnění spojená se změnami v zastoupení nebo koncentracích metabolitů v tělních tekutinách a tkáních. Toto poznání, možnost charakterizace střídaných látek a nástroje molekulární biologie umožnily vznik a rozvoj biochemické genetiky. Biochemická genetika umožňuje studovat genetickou příčinu biochemické poruchy tak, že nejprve určíme povahu abnormálně se vyskytující látky. Pokud je vyskytující se látka protein, můžeme dnes na základě znalosti sekvence proteinu pomocí programu BLASTX vyhledat v databázi přímo jeho genovou sekvenci a v té následně hledat metodou sekvenování DNA kauzální mutaci. Pokud je onou látkou nějaký metabolit, pokoušíme se určit příslušný biochemický defekt a následně izolovat enzym nebo protein, který je odpovědný za metabolismus této látky. Na základě znalosti proteinu můžeme opět vyhledat v databázi příslušnou genovou sekvenci a v té následně hledat metodou sekvenování DNA kauzální mutaci. Postupy biochemické genetiky a funkčního klonování identifikovaly mutace na genech způsobující onemocnění na úrovni metabolitů, proteinů nebo enzymů, např. fenylalanin hydroxylázu, způsobující fenylketonurii (Kwok *et al.*, 1985). Taková onemocnění patří do skupiny dědičných metabolických poruch, ve které je cca 800 dosud objevených metabolických onemocnění (Fernandes *et al.*, 2006).

## 7. Genetické mapování a poziční klonování

Pokud nebyla známá biochemická příčina onemocnění, začal se využívat další způsob pro zjištění příčin onemocnění – poziční klonování. Tento postup vycházel z konceptu genetické vazby definované T. H. Morganem (Morgan *et al.*, 1925). Základní snahou postupu bylo nalézt vztah mezi dědičností studované nemoci a dědičností nějakého dalšího faktoru. Historicky to byl nejprve vztah mezi spoludědičností nemoci a jiného fenotypu, např. dědičností jiné nemoci nebo pohlaví, barvy očí či krevní srážlivosti. Později to byl vztah dědičností nemoci s abnormálním počtem nebo strukturou chromozomů či s polymorfními krevními skupinami a proteinovými markery. Klíčový byl objev polymorfismu délkových restrikčních fragmentů a možnosti jejich využití pro vytvoření genetické mapy člověka a mapování genů pomocí vazebné analýzy (Botstein *et al.*, 1980). Genetické mapování je založeno na sledování a srovnávání dědičností jednotlivých polymorfních alel analyzovaných genetických markerů z nemocných a zdravých příbuzných jedinců. Nalezení alel, které jsou děděny spolu se studovanou nemocí, spolu se znalostí jejich pozice v genetické mapě následně přímo určují lokalizaci a pozici zmutovaného genu v genomu. Tento postup obecně vyžaduje vyšetření dostatečného počtu genetických markerů a zejména shromáždění velkého množství rodin a jejich příslušníků nebo relativně velkého počtu pacientů se stejným onemocněním, což je u konkrétního vzácného onemocnění limitujícím faktorem. Metoda pozičního klonování byla poměrně úspěšná pro identifikaci několika set případů (zhruba 1% všech studovaných onemocnění touto metodou), ale byla časově a finančně náročná a omezená historicky limitovanou znalostí sekvence lidského genomu.

## 8. Mapování a sekvence lidského genomu

Koncept genetického mapování a možnost jeho univerzálního využití pro odhalování genetických příčin nemocí byl impulzem pro vyhledávání nových typů genetických markerů, např. mikrosatelity („short tandem repeats, STRs“ (Jeffreys *et al.*, 1985)) a k postupnému budování genetických a fyzických map lidského genomu. Tento vývoj nakonec umožnil postupnou sekvenaci celého lidského genomu. V průběhu procesu mapování a sekvenování lidského genomu postupně narůstal počet osekvenovaných DNA od různých jedinců. Porovnáváním sekvencí jednotlivých lidí byly nalézány genetické varianty na úrovni jednotlivých nukleotidů - jednonukleotidové polymorfismy („single-nucleotide polymorphism, SNP“) (Sachidanandam *et al.*, 2001). To vedlo k vytvoření SNP databáze a haplotypové mapy lidského genomu HapMap (Gibbs *et al.*, 2003), která umožnila porovnávat SNP a jejich původní genetické varianty v lidském genomu (Altshuler *et al.*, 2000). Bylo též navrženo, že SNPs mohou být využívány jako markery pro poziční klonování a že řada geneticky podmíněných onemocnění může být přímo způsobena těmito SNPs (Collins *et al.*, 1999).

### 8.1 Sekvenování genomů a vznik databází

Vůbec první kompletní osekvenování celé DNA proběhlo u bakteriofágu  $\phi$ X174 (Sanger *et al.*, 1977). Rychle se zvyšující počet osekvenovaných DNA různých organismů vyžadoval nové nástroje pro uchovávání a analýzu DNA sekvencí. To vedlo k potřebě založení „The National Center for Biotechnology Information“ (NCBI) a ke vzniku databáze sekvencí GenBank, která dnes obsahuje veřejně dostupné sekvence nukleotidů pro téměř 260 000 formálně popsanych druhů (Benson *et al.*, 2013). Nejvýznamnějšími novými bioinformatickými nástroji se stalo vzájemné porovnávání a analýza sekvencí pomocí počítačových programů CLUSTAL (Higgins *et al.*, 1988), BLAST (Altschul *et al.*, 1990), FASTA (Pearson, 1990). Spojení všech dosud osekvenovaných DNA určilo referenční sekvenci lidského genomu v RefSeq databázi („The Genome Reference Consortium“ (Pruitt *et al.*, 2005)). Zvyšující se znalost sekvence lidského genomu postupně umožňovala jeho lepší funkční anotaci. Vznikly genomové prohlížeče, např. ENSEMBL (Hubbard *et al.*, 2002) nebo UCSC genome browser (Kent *et al.*, 2002), umožňující online vyhledávání a analýzu sekvence lidského genomu. Objevila se možnost sekvence pouze určitých (kódujících) oblastí DNA a vznikly prokaryotické cDNA knihovny pro kolekci naklonovaných fragmentů DNA - cDNA („complementary DNA“) určitého eukaryotického organismu. Pro automatické a paralelní sledování genové exprese nebo genotypování SNPs

byly vytvořeny vysokokapacitní systémy – DNA čipy neboli DNA microarray (Schena *et al.*, 1995).

## 8.2 Nové metody sekvenování DNA

Současně s rozvojem znalostí sekvence a funkce lidského genomu se rozvíjely také nové přístupy sekvenování DNA. Cílem bylo vytvořit postupy a technologie, které by významně zvětšily kapacitu tradičního Sangerova sekvenování (např. mnohokapilární systémy (Karger *et al.*, 2009)), nebo umožnily zcela jiným způsobem rychlé a levné sekvenování genomů nebo jejich částí u velkého počtu lidí („Next-Generation Sequencing, NGS“). První úspěšně zvládnutou NGS technologií bylo pyrosekvenování (Ronaghi *et al.*, 1996) komerčně rozvíjené firmou 454 a posléze Roche (Rothberg *et al.*, 2008). Současně s technologií pyrosekvenování se rozvíjely technologie sekvenování pomocí ligace fluorescenčně značených oligonukleotidů (Shendure *et al.*, 2005), později SOLiD („Sequencing by Oligonucleotide Ligation and Detection“; Life Sciences), nebo pomocí fluorescenčních reverzibilních dideoxy-terminátorů (Bentley *et al.*, 2008) (sekvenační přístroje Solexa, později Illumina). Možnosti NGS sekvenování dokumentovaly první projekty úspěšné sekvenace genomů jednotlivých osobností, např. Craiga Ventera (Sangerovo sekvenování; Celera Genomics) (Levy *et al.*, 2007) nebo Jamese Watsona (pyrosekvenování; 454) (Wheeler *et al.*, 2008) a sekvence cytogeneticky normálních leukemických buněk pomocí technologie Illumina (Ley *et al.*, 2008).

## 8.3 Cílené sekvenování

Rozsáhlému uplatnění NGS technologií zprvu bránila technologická náročnost, relativně vysoká cena a zejména malá schopnost analýzy a interpretace obrovského množství dat. Tato skutečnost vedla k rozvoji technologií, které umožňovaly sekvenovat pouze vybrané oblasti genomu (cílené sekvenování). Pro získání DNA oblastí, které byly předmětem zájmu, bylo využito možnosti obohacení původního vzorku DNA pomocí PCR, pomocí DNA čipů „SeqCap EZ System“ (Roche NimbleGen) nebo pomocí roztoku oligonukleotidových prób či pomocí komplementárních RNA „The SureSelect Target Enrichment“ (Agilent). Tento vývoj vedl k rozvoji technologie exomového sekvenování.

## 8.4 Exomové sekvenování

Exomové sekvenování umožňuje mnohem efektivnější způsob určování genetických variant v genomu jedince a je zatím nejvhodnější způsob pro studium příčin vzácných geneticky podmíněných onemocnění. Lidský exom tvoří malou oblast lidského genomu, zhruba 1%, což je asi 180 000 exonů (Ng *et al.*, 2009), proto je exomové sekvenování snazší, rychlejší, levnější a umožňuje dosáhnout vyššího pokrytí. V exomu se nachází až 85% závažných mutací, která způsobují vzácná onemocnění (Choi *et al.*, 2009). Potenciál exomového sekvenování pro odhalení genetických příčin vzácných nemocí byl prvně demonstrován na případě vzácného geneticky podmíněného onemocnění s neznámou příčinou, na Millerově syndromu. Osekvenovaly se exomy čtyř postižených jedinců (dva sourozenci a dva navzájem si nepříbuzní pacienti) a s pomocí mapování v HapMap, filtrování podle jednotlivých genetických modelů a porovnávání variant v SNP databázích se určil kandidátní gen *DHODH*, ve kterém se vyskytovala kauzální mutace způsobující Millerův syndrom. Následné Sangerovo sekvenování potvrdilo kauzální mutaci v tomto genu (Ng *et al.*, 2010). V následujících letech se exomové sekvenování stalo univerzální metodou pro odhalování genetických příčin vzácných, ale i komplexních onemocnění.

## 8.5 Genetická variabilita člověka

Rozšíření technologií NGS a možnost sekvence vybraných sad genů nebo genomových oblastí, sekvence exomů či rozsáhlé sekvenční studie celých genomů velkého počtu lidí a populací postupně odkrývají rozsah genetické variability člověka. Díky projektům, jako například The 1000 Genomes Project (The Genomes Project, 2010), The Genome 10K Project (Haussler *et al.*, 2009); Exome Variant Server, NHLBI GO Exome Sequencing Project, ESP (<http://evs.gs.washington.edu/EVS/>); The Exome Aggregation Consortium, ExAC (<http://exac.broadinstitute.org/>) a The Single Nucleotide Polymorphism Database dbSNP (<http://www.ncbi.nlm.nih.gov/snp/>) známe desítky miliónů genetických variant a jejich frekvence v různých genech a populacích. Znalost mutability jednotlivých genů a znalost populačních frekvencí jednotlivých variant je jedním ze základních nástrojů pro hodnocení případné kauzality genetických variant nalézáných ve vzorcích DNA pacientů se vzácným onemocněním. Je zřejmé, že vzácná onemocnění budou spojena pravděpodobně s mutacemi v genech, které jsou sekvenčně konzervované a s variantami, které jsou unikátní nebo populačně vzácné.

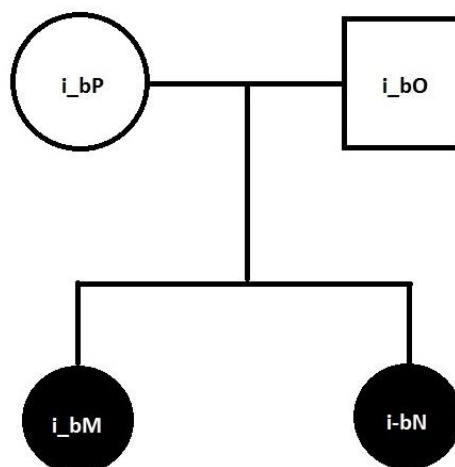
## 9. Praktická část

V praktické části své práce prezentuji své vlastní zkušenosti s využitím současných možností a metod analýzy genomu v diagnostice geneticky podmíněného vzácného onemocnění na případu jedné španělské rodiny.

Úkolem bylo zjistit, zda se mé výsledky budou shodovat s výsledky již vyřešeného případu, jehož diagnostiku a příčinu onemocnění zjistil tým z Ústavu dědičných metabolických poruch 1. LF UK. Před začátkem této praktické části jsem však příčinu onemocnění neznala.

### 9.1 Objasnění genetické příčiny hyperurikemické nefropatie u dvou sourozenců

V rámci praktické části této práce jsem měla za úkol zjistit genetickou příčinu vzácného onemocnění ve španělské rodině. Jednalo se o rodinu zdravých rodičů, ve které obě děti (sestry, 4 a 7 let) mají klinické příznaky opožděného psychomotorického vývoje, zvláštní kvalitu vlasů (*pili torti*), poruchy sluchu, snížené funkce ledvin, hypourické hyperurikémie (snížené vylučování kyseliny močové v krvi a zvýšená koncentrace kyseliny močové v séru), anémie a hyperkalémie. Oba rodiče jsou zdraví, pouze otec má mírný projev hyperurikémie s normální funkcí ledvin. Rodiče si nejsou vědomi jakékoliv vzájemné příbuznosti. Každý člen této rodiny je zobrazen na rodokmenu (viz Obr. 2) s příslušnými laboratorními kódy vzorků - i\_bM, i\_bN, i-bP, i\_bO.



Obr. 2. Rodokmen postižené rodiny

Klíčová slova charakterizující klinické projevy („*hypouricosuria hyperuricemia, anemia, nephropathy, hyperkalemia*“) jsem zadala do databáze OMIM. Na základě klinických příznaků jsem našla možnou diagnózu hyperurikemické nefropatie, familiární juvenilie 1 a 2, způsobené mutacemi v genu pro renin (*REN*) a uromodulin (*UMOD*) (viz Obr. 3).

Search: 'hypouricosuria hyperuricemia anemia nephropathy hyperkalemia'  
Results: 1 - 10 of 1,054 | [Show 100](#) | [Download As](#) ▾ | [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [Next](#) [Last](#)

1 : # [613092](#). [HYPERURICEMIC NEPHROPATHY, FAMILIAL JUVENILE, 2; HNFJ2](#)  
Cytogenetic location: [1q32.1](#)  
Matching terms: hyperuricemia, nephropathy, anemia

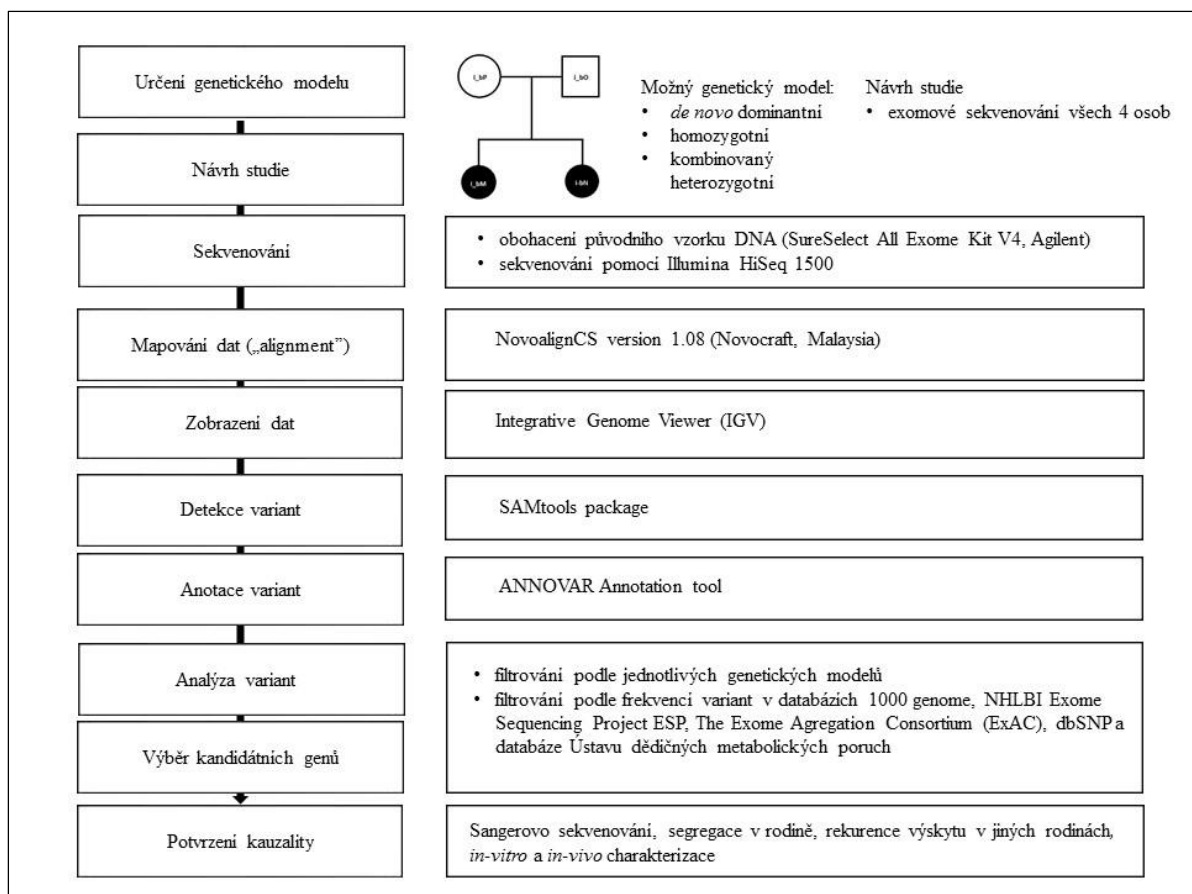
2 : # [162000](#). [HYPERURICEMIC NEPHROPATHY, FAMILIAL JUVENILE, 1; HNFJ1](#)  
Cytogenetic location: [16p12.3](#)  
Matching terms: hyperuricemia, nephropathy

**Obr. 3. Možné diagnózy podle zadaných klinických příznaků** (převzato z databáze OMIM, <http://www.omim.org/>)

Na základě této informace bylo u pacientů provedeno cílené sekvenování genů *REN* a *UMOD*, které ale neodhalilo žádné kauzální mutace. U všech členů rodiny proto bylo provedeno exomové sekvenování. Hledání kauzální mutace proběhlo pomocí standardního postupu (viz Obr. 4). DNA byla obohacena pomocí kitu „SureSelect Human All Exon V4“ (Agilent). Exomové sekvenování celé rodiny bylo provedeno na sekvenátoru Illumina HiSeq 1500. Získaná data byla mapována na referenčním genomu pomocí programu NovoalignCS version 1.08 (Novocraft, Malaysia). Byly nalezeny odchylky od referenční sekvence a tyto odchylky byly poté anotovány programem SAMtools package (Li *et al.*, 2009). Po mapování dat byly získány anotace genetických variant pomocí programu ANNOVAR Annotation tool (Wang *et al.*, 2010). Oproti referenční sekvenci bylo ve všech čtyřech analyzovaných vzorcích nalezeno celkem 141 547 genetických variant. Na základě rodokmenu jsem vyhledávala varianty, které splňují model pozorované dědičnosti a vzácně se vyskytují (<1%) v populačních databázích 1000 genomes, ESP, ExAC a v databázi Ústavu dědičných metabolických poruch. Vyhledávala jsem nejdříve varianty splňující podmínky *de novo* dominantního modelu (varianty, které jsou přítomny heterozygotně u obou dětí, ale ne u rodičů), poté jsem hledala varianty splňující

podmínky homozygotního modelu (rodiče heterozygoté a děti homozygoté pro jednu variantu) a nakonec varianty splňující podmínky kombinovaného heterozygotního modelu (každý z rodičů je heterozygot pro jinou mutaci v jednom genu a obě děti jsou kombinovaní heterozygoté pro obě mutace). Výsledky filtrování jsou v Tabulce 1. V anotační tabulce jsem nastavila filtrování pro heterozygotní model s autozomálně dominantními a autozomálně recesivními variantami. Pro další krok filtrace jsem vybrala všechny varianty splňující model kombinovaných heterozygotů (i\_bM,i\_bN,i\_bP a i\_bM,i\_bN,i\_bO, dle rodokmenu). Tyto kritéria splňovalo 18 266 variant. V dalším kroku jsem vybrala pouze kódující varianty, celkem jich zůstalo 2985. Dále jsem nastavila data kontrolních skupin, tzn. vyhledávání pouze těch variant, které nejsou přítomny v našich kontrolních skupinách, zbylo 257 variant. Jako poslední krok jsem přefiltrovala klinicky relevantní varianty a varianty, u kterých byl vliv mutace na funkci proteinu predikovaný jako patogenní. Zůstaly mi 3 varianty, z nichž dvě varianty představují závažnou mutaci v genu *RMND1* (viz Tab. 1). Podle uvedených dat (viz Tab. 2) jsem zkontrolovala výskyt obou závažných mutací v genu *RMND1* pomocí programu Integrative Genomic Viewer (IGV) (Robinson *et al.*, 2011) (viz Obr. 5 a 6). Poté jsem vyhledala dostupné informace o genu *RMND1* a porovnála jsem vztah mezi kauzální mutací v genu *RMND1* a projevenými fenotypy pacientů.





**Obr. 4. Schéma standardního postupu hledání kauzální mutace způsobující onemocnění (upraveno dle (Kmoch *et al.*, 2013))**

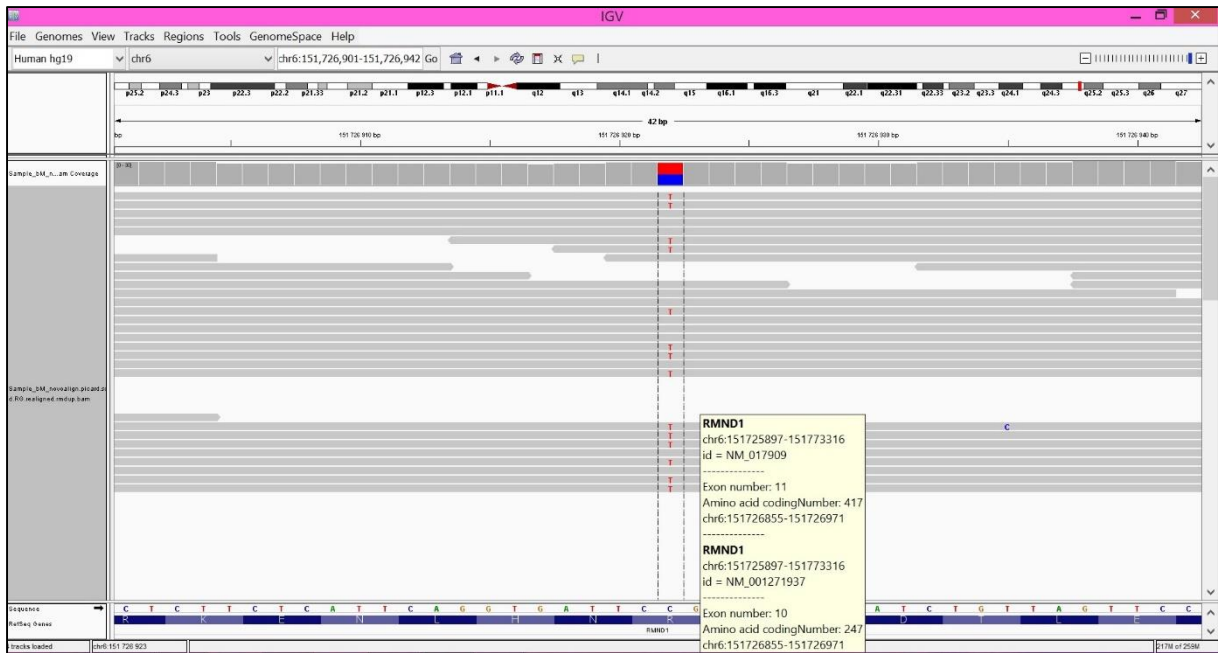
**Tab. 1. Počet variant v průběhu filtrování dat u příslušného genetického modelu**

	<b>heterozygotní model</b>	<b>homozygotní model</b>	<b><i>de novo</i> dominantní model</b>
počet variant celkem	141 547	141 547	141 547
model kombinovaných heterozygotů/homozygotů/ <i>de novo</i> variant	18 266	1862	129
kódující varianty	2985	237	19
varianty nepřítomné v kontrolní skupině	257	6	19
klinicky relevantní varianty	<b>2</b>	0	0

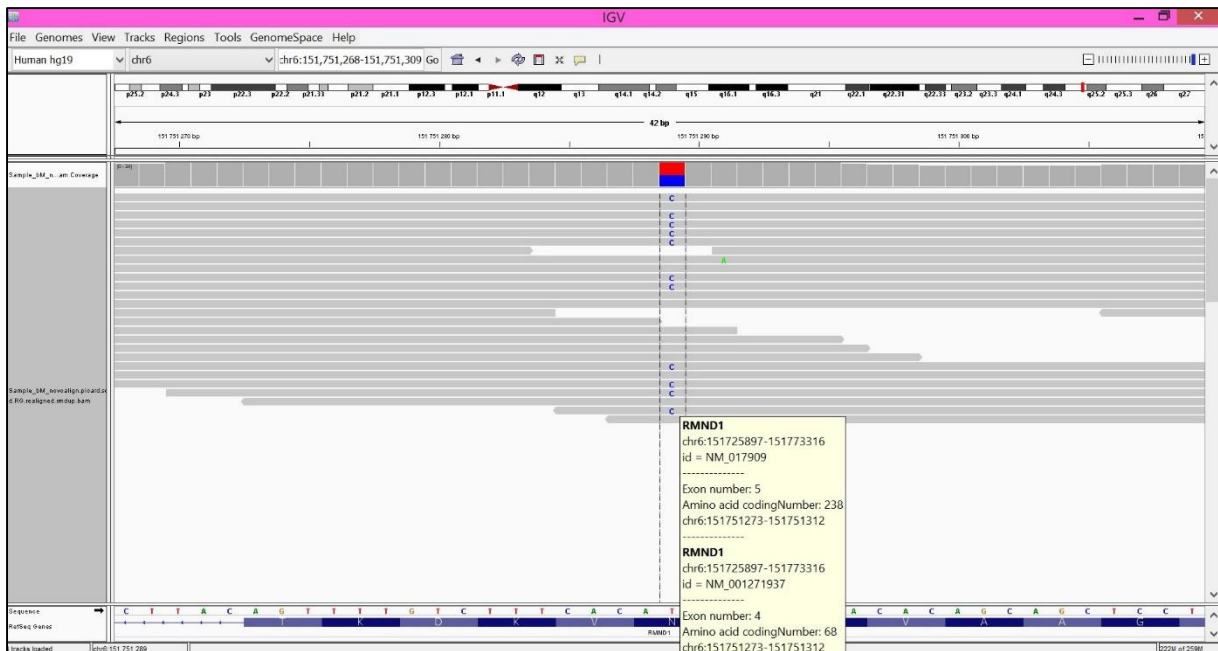
**Tab. 2. Výsledek přísného filtrování všech možných kauzálních variant v analyzovaných exomech**

chrom	<b>chr6</b>	<b>chr6</b>
start	<b>151726921</b>	<b>151751288</b>
end	<b>151726922</b>	<b>151751289</b>
ref	<b>C</b>	<b>T</b>
alt	<b>T</b>	<b>C</b>
rs_ids	<b>rs397515421</b>	<b>rs144972972</b>
clinvar_sig	<b>pathogenic</b>	<b>pathogenic</b>
gene	<b><i>RMND1</i></b>	<b><i>RMND1</i></b>
aa_change	<b>R206Q</b>	<b>N238S</b>
aaf_exac_all	<b>8,24E-06</b>	<b>0,000173</b>
variant_samples	<b>i_bM,i_bN,i_bP</b>	<b>i_bM,i_bN,i_bO</b>
HET_samples	<b>i_bM,i_bN,i_bP</b>	<b>i_bM,i_bN,i_bO</b>

(vysvětlivky: **chrom** – chromozom; **start** – poziční začátek úseku DNA s kauzální mutací; **end** – poziční konec úseku DNA s kauzální mutací; **ref** – nukleotidová báze podle referenční sekvence; **alt** – alternativní nukleotidová báze; **rs\_ids** – identifikátor varianty v databázi dbSNP; **clinvar\_sig** – klinicky relevantní skóre varianty podle databáze ClinVar; **gene** – gen; **aa\_change** – předpokládaný efekt varianty na aminokyselinovou sekvenci příslušného proteinu; **aaf\_exac\_all** – frekvence výskytu v databázi ExAC; **variant\_samples** – laboratorní kód analyzovaných vzorků; **HET\_samples** – vzorky, ve kterých byla nalezená varianta přítomná v heterozygotní formě)



Obr. 5. Grafická vizualizace a přesná lokalizace první varianty, kauzální mutace chr6:151726921 C→T v genu *RMND1* (Integrative Genomic Viewer)



Obr. 6. Grafická vizualizace a přesná lokalizace druhé varianty, kauzální mutace chr6:151751288 T→C v genu *RMND1* (Integrative Genomic Viewer)

## 10. Diskuse

Cílem praktické části bylo zjistit příčinu vzácného onemocnění s již popsanými klinickými projevy. S využitím všech dostupných metod analýzy DNA, popsaných výše v této práci jsem postupně našla dvě patogenní varianty s kauzální mutací genu *RMND1* vyskytujícího se na chromozomu 6. Obě varianty s mutací byly následně zpětně ověřeny a potvrzeny klasickou metodou Sangerova sekvenování. Po finálním potvrzení jsem vyhledala všechny dostupné informace o genu *RMND1*, protože mi nebyly známy jeho skutečné funkce. Rešerší dostupné literatury jsem zjistila, že tento gen kóduje integrální membránový mitochondriální protein RMND1, který se sbaluje do velkého proteinového komplexu o velikosti 240 kDa. Tento komplex pomáhá při translaci 13 polypeptidů, které jsou důležitými podjednotkami komplexů oxidativní fosforylace (Janer *et al.*, 2012). Mitochondriální onemocnění jsou nejčastěji způsobeny mutacemi v mitochondriální DNA nebo genových produktech, které jsou transportovány do mitochondrie z cytosolu a nejčastěji ovlivňují mitochondriální metabolismus, který zajišťuje správnou funkci oxidativní fosforylace ve vnitřní mitochondriální membráně. Oxidativní fosforylace je mitochondriální metabolická dráha vyrábějící energii ve formě ATP pro celý metabolismus aerobního organismu. Poruchy oxidativní fosforylace a mitochondriální poruchy se projevují velmi rozmanitými klinickými projevy, ovlivňující jednotlivé tkáně nebo komplexnější struktury, např. centrální a periferní nervovou soustavu, srdce, svaly, játra ledviny, zrak a sluch. Mutace v genu *RMND1* celkově ovlivňují mitochondriální translaci a funkci, které se posléze projevují ojedinělými klinickými fenotypy, např. encefalopatií (Garcia-Diaz *et al.*, 2012) nebo neonatální laktátovou acidózou, selháním ledvin v dětském věku, hluchotou a multiorgánovým postižením (Taylor *et al.*, 2014; Janer *et al.*, 2015). Tyto popsané fenotypy jsou způsobeny závažnými mutacemi genu *RMND1*, které ovlivňují mitochondriální protein RMND1 a dokazují, že familiární nefropatie dětského věku ve studované rodině je pravděpodobně způsobena nalezenými mutacemi *RMND1*.

## 11. Závěr

Ve své práci jsem se pokusila podat přehled metod a možností výzkumu, které byly historicky používány pro objasnění a charakterizaci příčin geneticky podmíněných vzácných onemocnění. Dále popisuji aktuální (moderní) postupy genetické analýzy, význam a možnosti studia a případ, ve kterém využití těchto nových metod objasnilo příčinu nemoci u dvou sourozenců s hyperurikemickou nefropatií.

Zvyšující se podvědomí o vzácných onemocněních, znalost struktury a funkce lidského genomu, výše popsané metody, sekvenační analýzy DNA, systematické shromažďování sekvenačních dat a možnost jejich porovnávání umožnilo postupně určit genetickou příčinu několika tisíc onemocnění (Botstein *et al.*, 2003). Možnost přímého porovnávání genomové sekvence postiženého jedince, případně jeho rodiny s genetickou variabilitou populace se dnes stává univerzálním nástrojem pro hledání genetických příčin všech vzácných nemocí. Tento přístup není tak limitován, jako v minulosti předběžnou znalostí biochemické podstaty, ani velikostí vyšetřované rodiny nebo rozsáhlého souboru pacientů. Exomové, případně genomové sekvenování budou stále častějším základním diagnostickým testem pro pacienty a rodiny se vzácným onemocněním. Míra klinické využitelnosti těchto přístupů se bude stále zvyšovat s tím, jak porostou naše schopnosti správně interpretovat genetickou informaci jedince ve vztahu k příčinám a prognóze onemocnění. Základním předpokladem správné interpretace genomových dat jedince bude existence platforem a databází, které umožní odpovědné, dobrovolné a bezpečné sdílení genomových a klinických dat na celosvětové úrovni. Řada takových aktivit se v současnosti rozvíjí (Mailman *et al.*, 2007; Byrne *et al.*, 2012). Rozvoji tohoto přístupu však zatím brání možné zneužití znalosti genetické informace jedince.

## Seznam použité literatury

(\* sekundární citace)

- Altshuler D., Pollara V.J., Cowles C.R., Van Etten W.J., Baldwin J., Linton L. and Lander E.S. (2000). "An SNP map of the human genome generated by reduced representation shotgun sequencing." *Nature* **407**(6803): 513-516.
- Altschul S.F., Gish W., Miller W., Myers E.W. and Lipman D.J. (1990). "Basic local alignment search tool." *Journal of Molecular Biology* **215**(3): 403-410.
- Avery O.T., MacLeod C.M. and McCarty M. (1944). "Studies on the chemical nature of the substance inducing transformation of pneumococcal types induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III." *The Journal of Experimental Medicine* **79**(2): 137-158.
- Bateson W. (1901). "Experiments in plant hybridization by Gregor Mendel." *Journal of the Royal Horticultural Society* **24**(1): 1-32.
- Benson D.A., Cavanaugh M., Clark K., Karsch-Mizrachi I., Lipman D.J., Ostell J. and Sayers E.W. (2013). "GenBank." *Nucleic Acids Research* **41**(D1): D36-D42.
- Bentley D.R., Balasubramanian S., Swerdlow H.P., Smith G.P., Milton J., Brown C.G., Hall K.P., Evers D.J., Barnes C.L., Bignell H.R., Boutell J.M., Bryant J., Carter R.J., Keira Cheetham R., Cox A.J., Ellis D.J., Flatbush M.R., Gormley N.A., Humphray S.J., Irving L.J., Karbelashvili M.S., Kirk S.M., Li H., Liu X., Maisinger K.S., Murray L.J., Obradovic B., Ost T., Parkinson M.L., Pratt M.R., Rasolonjatovo I.M.J., Reed M.T., Rigatti R., Rodighiero C., Ross M.T., Sabot A., Sankar S.V., Scally A., Schroth G.P., Smith M.E., Smith V.P., Spiridou A., Torrance P.E., Tzonev S.S., Vermaas E.H., Walter K., Wu X., Zhang L., Alam M.D., Anastasi C., Aniebo I.C., Bailey D.M.D., Bancarz I.R., Banerjee S., Barbour S.G., Baybayan P.A., Benoit V.A., Benson K.F., Bevis C., Black P.J., Boodhun A., Brennan J.S., Bridgham J.A., Brown R.C., Brown A.A., Buermann D.H., Bundu A.A., Burrows J.C., Carter N.P., Castillo N., Chiara E. Catenazzi M., Chang S., Neil Cooley R., Crake N.R., Dada O.O., Diakoumakos K.D., Dominguez-Fernandez B., Earnshaw D.J., Egbujor U.C., Elmore D.W., Echin S.S., Ewan M.R., Fedurco M., Fraser L.J., Fuentes Fajardo K.V., Scott Furey W., George D., Gietzen K.J., Goddard C.P., Golda G.S., Granieri P.A., Green D.E., Gustafson D.L., Hansen N.F., Harnish K., Haudenschild C.D., Heyer N.I., Hims M.M., Ho J.T., Horgan A.M., Hoschler K., Hurwitz S., Ivanov D.V., Johnson M.Q., James T., Huw Jones T.A., Kang G.-D., Kerelska T.H., Kersey A.D., Khrebtukova I., Kindwall A.P., Kingsbury Z., Kokko-Gonzales P.I., Kumar A., Laurent M.A., Lawley C.T., Lee S.E., Lee X., Liao A.K., Loch J.A., Lok M., Luo S., Mammen R.M., Martin J.W., McCauley P.G., McNitt P., Mehta P., Moon K.W., Mullens J.W., Newington T., Ning Z., Ling Ng B., Novo S.M., O'Neill M.J., Osborne M.A., Osnowski A., Ostadan O., Paraschos L.L., Pickering L., Pike A.C., Pike A.C., Chris Pinkard D., Pliskin D.P., Podhasky J., Quijano V.J., Raczy C., Rae V.H., Rawlings S.R., Chiva Rodriguez A., Roe P.M., Rogers J., Rogert Bacigalupo M.C., Romanov N., Romieu A., Roth R.K., Rourke N.J., Ruediger S.T., Rusman E., Sanches-Kuiper R.M., Schenker M.R., Seoane J.M., Shaw R.J., Shiver M.K., Short S.W., Sizto N.L., Sluis J.P., Smith M.A., Ernest Sohna J., Spence E.J., Stevens K., Sutton N., Szajkowski L., Tregidgo C.L., Turcatti G., vandeVondele S., Verhovsky Y., Virk S.M., Wakelin S., Walcott G.C., Wang J., Worsley G.J., Yan J., Yau L., Zuerlein M., Rogers J., Mullikin J.C., Hurles M.E., McCooke N.J., West J.S., Oaks F.L., Lundberg P.L., Klenerman D., Durbin R. and Smith A.J. (2008). "Accurate whole human genome sequencing using reversible terminator chemistry." *Nature* **456**(7218): 53-59.
- Botstein D. and Risch N. (2003). "Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease." *Nature Genetics* **33**: 228-237.
- Botstein D., White R.L., Skolnick M. and Davis R.W. (1980). "Construction of a genetic linkage map in man using restriction fragment length polymorphisms." *American Journal of Human Genetics* **32**(3): 314-331.

- Byrne M., Fokkema I.F., Lancaster O., Adamusiak T., Ahonen-Bishopp A., Atlan D., Beroud C., Cornell M., Dalgleish R., Devereau A., Patrinos G.P., Swertz M.A., Taschner P.E., Thorisson G.A., Vihinen M., Brookes A.J. and Muilu J. (2012). "VarioML framework for comprehensive variation data representation and exchange." BMC Bioinformatics **13**: 254.
- Campbell C.D. and Eichler E.E. (2013). "Properties and rates of germline mutations in humans." Trends in Genetics **29**(10): 575-584.
- Cavalli-Sforza L.L. and Feldman M.W. (2003). "The application of molecular genetic approaches to the study of human evolution." Nature Genetics **33**: 266-275.
- Collins A., Lonjou C. and Morton N.E. (1999). "Genetic epidemiology of single-nucleotide polymorphisms." Proceedings of the National Academy of Sciences of the United States of America **96**(26): 15173-15177.
- Crick F. (1970). "Central dogma of molecular biology." Nature **227**(5258): 561-563.
- \*Dahm R. (2005). "Friedrich Miescher and the discovery of DNA." Developmental biology **278**(2): 274-288.
- Darwin C. (1859). "On the origins of species by means of natural selection." London: Murray.
- \*Dodge J.A., Chigladze T., Donadieu J., Grossman Z., Ramos F., Serlicorni A., Siderius L., Stefanidis C.J., Tasic V. and Valiulis A. (2010). "The importance of rare diseases: from the gene to society." Archives of Disease in Childhood: archdischild193664.
- Fernandes J., Saudubray J.-M., Van Den Berghe G. and Walter J.H. (2006). Inborn Metabolic Diseases: Diagnosis and Treatment, Springer Science & Business Media.
- Flemming W. (1879). "Ueber das Verhalten des Kerns bei der Zelltheilung, und über die Bedeutung mehrkerniger Zellen." Virchows Archiv **77**(1): 1-29.
- Fu Q., Li H., Moorjani P., Jay F., Slepchenko S.M., Bondarev A.A., Johnson P.L., Aximu-Petri A., Prufer K., de Filippo C., Meyer M., Zwyns N., Salazar-Garcia D.C., Kuzmin Y.V., Keates S.G., Kosintsev P.A., Razhev D.I., Richards M.P., Peristov N.V., Lachmann M., Douka K., Higham T.F., Slatkin M., Hublin J.J., Reich D., Kelso J., Viola T.B. and Paabo S. (2014). "Genome sequence of a 45,000-year-old modern human from western Siberia." Nature **514**(7523): 445-449.
- Garcia-Diaz B., Barros Mario H., Sanna-Cherchi S., Emmanuele V., Akman Hasan O., Ferreira-Barros Claudia C., Horvath R., Tadesse S., El Gharaby N., DiMauro S., De Vivo Darryl C., Shokr A., Hirano M. and Quinzii Catarina M. (2012). "Infantile Encephalomyopathy and Defective Mitochondrial Translation Are Due to a Homozygous RMND1 Mutation." The American Journal of Human Genetics **91**(4): 729-736.
- Gibbs R.A., Belmont J.W., Hardenbol P., Willis T.D., Yu F., Yang H., Ch'ang L.-Y., Huang W., Liu B. and Shen Y. (2003). "The international HapMap project." Nature **426**(6968): 789-796.
- Griffiths A.J., Miller J.H., Suzuki D.T., Lewontin R.C. and Gelbart W.M. (2000). Somatic versus germinal mutation.
- Haldane J.B. (1947). "The mutation rate of the gene for haemophilia, and its segregation ratios in males and females." Ann Eugen **13**(4): 262-271.
- Hamosh A., Scott A.F., Amberger J.S., Bocchini C.A. and McKusick V.A. (2005). "Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders." Nucleic Acids Research **33**(Database issue): D514-D517.
- Harris K. (2015). "Evidence for recent, population-specific evolution of the human mutation rate." Proceedings of the National Academy of Sciences **112**(11): 3439-3444.
- Haussler D., O'Brien S.J., Ryder O.A., Barker F.K., Clamp M., Crawford A.J., Hanner R., Hanotte O., Johnson W.E. and McGuire J.A. (2009). "Genome 10K: a proposal to obtain whole-genome sequence for 10 000 vertebrate species." Journal of Heredity **100**(6): 659-674.
- Hawks J., Hunley K., Lee S.-H. and Wolpoff M. (2000). "Population bottlenecks and Pleistocene human evolution." Molecular Biology and Evolution **17**(1): 2-22.
- Higgins D.G. and Sharp P.M. (1988). "CLUSTAL: a package for performing multiple sequence alignment on a microcomputer." Gene **73**(1): 237-244.
- Hubbard T., Barker D., Birney E., Cameron G., Chen Y., Clark L., Cox T., Cuff J., Curwen V. and Down T. (2002). "The Ensembl genome database project." Nucleic Acids Research **30**(1): 38-41.

- Hunt J. and Ingram V. (1958). "Allelomorphism and the chemical differences of the human haemoglobins A, S and C." Nature.
- Choi M., Scholl U.I., Ji W., Liu T., Tikhonova I.R., Zumbo P., Nayir A., Bakkaloğlu A., Özen S., Sanjad S., Nelson-Williams C., Farhi A., Mane S. and Lifton R.P. (2009). "Genetic diagnosis by whole exome capture and massively parallel DNA sequencing." Proceedings of the National Academy of Sciences of the United States of America **106**(45): 19096-19101.
- Janer A., Antonicka H., Lalonde E., Nishimura T., Sasarman F., Brown Garry K., Brown Ruth M., Majewski J. and Shoubridge Eric A. (2012). "An RMND1 Mutation Causes Encephalopathy Associated with Multiple Oxidative Phosphorylation Complex Deficiencies and a Mitochondrial Translation Defect." The American Journal of Human Genetics **91**(4): 737-743.
- Janer A., van Karnebeek C.D., Sasarman F., Antonicka H., Al Ghamdi M., Shyr C., Dunbar M., Stockler-Ispiroglu S., Ross C.J., Vallance H., Dionne J., Wasserman W.W. and Shoubridge E.A. (2015). "RMND1 deficiency associated with neonatal lactic acidosis, infantile onset renal failure, deafness, and multiorgan involvement." European Journal of Human Genetics : EJHG.
- Jeffreys A.J., Wilson V. and Thein S.L. (1985). "Hypervariable 'minisatellite' regions in human DNA." Nature **314**(6006): 67-73.
- Karger B.L. and Guttman A. (2009). "DNA Sequencing by Capillary Electrophoresis." Electrophoresis **30**(Suppl 1): S196-S202.
- Keightley P.D. (2012). "Rates and fitness consequences of new mutations in humans." Genetics **190**(2): 295-304.
- Kent W.J., Sugnet C.W., Furey T.S., Roskin K.M., Pringle T.H., Zahler A.M. and Haussler D. (2002). "The human genome browser at UCSC." Genome Research **12**(6): 996-1006.
- Kmoch S., Stránecký V., Emes R.D. and Mitchison H.M. (2013). "Bioinformatic perspectives in the neuronal ceroid lipofuscinoses." Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease **1832**(11): 1831-1841.
- Kong A., Frigge M.L., Masson G., Besenbacher S., Sulem P., Magnusson G., Gudjonsson S.A., Sigurdsson A., Jonasdottir A., Jonasdottir A., Wong W., Sigurdsson G., Walters G.B., Steinberg S., Helgason H., Thorleifsson G., Gudbjartsson D.F., Helgason A., Magnusson O.T., Thorsteinsdottir U. and Stefansson K. (2012). "Rate of de novo mutations, father's age, and disease risk." Nature **488**(7412): 471-475.
- Kumar S. and Subramanian S. (2002). "Mutation rates in mammalian genomes." Proceedings of the National Academy of Sciences **99**(2): 803-808.
- Kwok S.C., Ledley F.D., DiLella A.G., Robson K.J. and Woo S.L. (1985). "Nucleotide sequence of a full-length complementary DNA clone and amino acid sequence of human phenylalanine hydroxylase." Biochemistry **24**(3): 556-561.
- Lande R. (1976). "Natural selection and random genetic drift in phenotypic evolution." Evolution: 314-334.
- Lercher M.J. and Hurst L.D. (2002). "Human SNP variability and mutation rate are higher in regions of high recombination." Trends in Genetics **18**(7): 337-340.
- Levy S., Sutton G., Ng P.C., Feuk L., Halpern A.L., Walenz B.P., Axelrod N., Huang J., Kirkness E.F. and Denisov G. (2007). "The diploid genome sequence of an individual human." PLoS Biology **5**(10): e254.
- Ley T.J., Mardis E.R., Ding L., Fulton B., McLellan M.D., Chen K., Dooling D., Dunford-Shore B.H., McGrath S., Hickenbotham M., Cook L., Abbott R., Larson D.E., Koboldt D.C., Pohl C., Smith S., Hawkins A., Abbott S., Locke D., Hillier L.W., Miner T., Fulton L., Magrini V., Wylie T., Glasscock J., Conyers J., Sander N., Shi X., Osborne J.R., Minx P., Gordon D., Chinwalla A., Zhao Y., Ries R.E., Payton J.E., Westervelt P., Tomasson M.H., Watson M., Baty J., Ivanovich J., Heath S., Shannon W.D., Nagarajan R., Walter M.J., Link D.C., Graubert T.A., DiPersio J.F. and Wilson R.K. (2008). "DNA sequencing of a cytogenetically normal acute myeloid leukemia genome." Nature **456**(7218): 66-72.
- Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G. and Durbin R. (2009). "The Sequence Alignment/Map format and SAMtools." Bioinformatics **25**(16): 2078-2079.
- Mailman M.D., Feolo M., Jin Y., Kimura M., Tryka K., Bagoutdinov R., Hao L., Kiang A., Paschall J., Phan L., Popova N., Pretel S., Ziyabari L., Lee M., Shao Y., Wang Z.Y., Sirotkin K., Ward



- M., Kholodov M., Zbicz K., Beck J., Kimelman M., Shevelev S., Preuss D., Yaschenko E., Graeff A., Ostell J. and Sherry S.T. (2007). "The NCBI dbGaP database of genotypes and phenotypes." *Nature Genetics* **39**(10): 1181-1186.
- Mathieson I. and McVean G. (2014). "Demography and the age of rare variants." *PLoS Genetics* **10**(8): e1004528.
- Maxam A.M. and Gilbert W. (1977). "A new method for sequencing DNA." *Proceedings of the National Academy of Sciences* **74**(2): 560-564.
- McClellan J. and King M.-C. (2010). "Genetic Heterogeneity in Human Disease." *Cell* **141**(2): 210-217.
- McVean G.T. and Hurst L.D. (1997). "Evidence for a selectively favourable reduction in the mutation rate of the X chromosome." *Nature* **386**(6623): 388-392.
- Mendel G., Bennett J. and Fisher S.R.A. (1965). *Experiments in plant hybridization*, Oliver & Boyd.
- Morgan T.H. (1911). "Chromosomes and associative inheritance." *Science*: 636-638.
- Morgan T.H., Bridges C. and Sturtevant A. (1925). "The genetics of *Drosophila melanogaster*." *Bibliophia genet* **2**: 1-262.
- Mullis K.B., Ferré F. and Gibbs R.A. (1994). *The polymerase chain reaction*, Birkhauser Boston Inc.
- Nachman M.W. and Crowell S.L. (2000). "Estimate of the mutation rate per nucleotide in humans." *Genetics* **156**(1): 297-304.
- Ng S.B., Buckingham K.J., Lee C., Bigham A.W., Tabor H.K., Dent K.M., Huff C.D., Shannon P.T., Jabs E.W., Nickerson D.A., Shendure J. and Bamshad M.J. (2010). "Exome sequencing identifies the cause of a Mendelian disorder." *Nature Genetics* **42**(1): 30-35.
- Ng S.B., Turner E.H., Robertson P.D., Flygare S.D., Bigham A.W., Lee C., Shaffer T., Wong M., Bhattacharjee A., Eichler E.E., Bamshad M., Nickerson D.A. and Shendure J. (2009). "Targeted Capture and Massively Parallel Sequencing of Twelve Human Exomes." *Nature* **461**(7261): 272-276.
- Nirenberg M., Caskey T., Marshall R., Brimacombe R., Kellogg D., Doctor B., Hatfield D., Levin J., Rottman F. and Pestka S. (1966). *The RNA code and protein synthesis*. Cold Spring Harbor Symposia on Quantitative Biology, Cold Spring Harbor Laboratory Press.
- Pauling L., Itano H.A., Singer S. and Wells I.C. (1949). "Sickle Cell Anemia." *Science* **110**.
- Pearson W.R. (1990). "[5] Rapid and sensitive sequence comparison with FASTP and FASTA." *Methods in Enzymology* **183**: 63-98.
- Pruitt K.D., Tatusova T. and Maglott D.R. (2005). "NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins." *Nucleic Acids Research* **33**(Database issue): D501-D504.
- Robinson J.T., Thorvaldsdóttir H., Winckler W., Guttman M., Lander E.S., Getz G. and Mesirov J.P. (2011). "Integrative genomics viewer." *Nature Biotechnology* **29**(1): 24-26.
- Ronaghi M., Karamohamed S., Pettersson B., Uhlen M. and Nyren P. (1996). "Real-time DNA sequencing using detection of pyrophosphate release." *Anal Biochem* **242**(1): 84-89.
- Rothberg J.M. and Leamon J.H. (2008). "The development and impact of 454 sequencing." *Nature Biotechnology* **26**(10): 1117-1124.
- Sachidanandam R., Weissman D., Schmidt S.C., Kakol J.M., Stein L.D., Marth G., Sherry S., Mullikin J.C., Mortimore B.J. and Willey D.L. (2001). "A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms." *Nature* **409**(6822): 928-933.
- Sanger F., Air G.M., Barrell B.G., Brown N.L., Coulson A.R., Fiddes C.A., Hutchison C.A., Slocombe P.M. and Smith M. (1977). "Nucleotide sequence of bacteriophage phi X174 DNA." *Nature* **265**(5596): 687-695.
- Sanger F., Nicklen S. and Coulson A.R. (1977). "DNA sequencing with chain-terminating inhibitors." *Proceedings of the National Academy of Sciences* **74**(12): 5463-5467.
- Shendure J., Porreca G.J., Reppas N.B., Lin X., McCutcheon J.P., Rosenbaum A.M., Wang M.D., Zhang K., Mitra R.D. and Church G.M. (2005). "Accurate multiplex polony sequencing of an evolved bacterial genome." *Science* **309**(5741): 1728-1732.
- Schena M., Shalon D., Davis R.W. and Brown P.O. (1995). "Quantitative monitoring of gene expression patterns with a complementary DNA microarray." *Science* **270**(5235): 467-470.
- \*Schieppati A., Henter J.-I., Daina E. and Aperia A. (2008). "Why rare diseases are an important medical and social issue." *The Lancet* **371**(9629): 2039-2041.

- \*Schieppati A., Henter J.I., Daina E. and Aperia A. (2008). "Why rare diseases are an important medical and social issue." *Lancet* **371**(9629): 2039-2041.
- Sutton W.S. (1902). "ON THE MORPHOLOGY OF THE CHROMOSOME GROUP IN BRACHYSTOLA MAGNA." *The Biological Bulletin* **4**(1): 24-39.
- Taruscio D., Gentile A., De Santis M., Ferrelli R., Posada d.I.P.M., Hens M., Huizer J., Fregonese L., Stefanov R. and Bottarelli V. (2012). "EUROPLAN: a project to support the development of national plans on rare diseases in Europe." *Public Health Genomics* **16**(6): 278-287.
- Taruscio D., Vittozzi L. and Stefanov R. (2010). National plans and strategies on rare diseases in Europe. *Rare Diseases Epidemiology*, Springer: 475-491.
- Taylor J., Tyekucheva S., Zody M., Chiaromonte F. and Makova K.D. (2006). "Strong and weak male mutation bias at different sites in the primate genomes: insights from the human-chimpanzee comparison." *Molecular Biology and Evolution* **23**(3): 565-573.
- Taylor R.W., Pyle A., Griffin H., Blakely E.L., Duff J., He L., Smertenko T., Alston C.L., Neeve V.C., Best A., Yarham J.W., Kirschner J., Schara U., Talim B., Topaloglu H., Baric I., Holinski-Feder E., Abicht A., Czermin B., Kleinle S., Morris A.A., Vassallo G., Gorman G.S., Ramesh V., Turnbull D.M., Santibanez-Koref M., McFarland R., Horvath R. and Chinnery P.F. (2014). "Use of whole-exome sequencing to determine the genetic basis of multiple mitochondrial respiratory chain complex deficiencies." *Jama* **312**(1): 68-77.
- Templeton A.R. (1980). "The theory of speciation via the founder principle." *Genetics* **94**(4): 1011-1038.
- The Genomes Project C. (2010). "A map of human genome variation from population scale sequencing." *Nature* **467**(7319): 1061-1073.
- Tomlinson I.P., Novelli M. and Bodmer W. (1996). "The mutation rate and cancer." *Proceedings of the National Academy of Sciences* **93**(25): 14800-14803.
- Wang K., Li M. and Hakonarson H. (2010). "ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data." *Nucleic Acids Research* **38**(16): e164-e164.
- Watson J.D. and Crick F.H. (1953). "Molecular structure of nucleic acids." *Nature* **171**(4356): 737-738.
- Wheeler D.A., Srinivasan M., Egholm M., Shen Y., Chen L., McGuire A., He W., Chen Y.-J., Makhijani V., Roth G.T., Gomes X., Tartaro K., Niazi F., Turcotte C.L., Irzyk G.P., Lupski J.R., Chinault C., Song X.-z., Liu Y., Yuan Y., Nazareth L., Qin X., Muzny D.M., Margulies M., Weinstock G.M., Gibbs R.A. and Rothberg J.M. (2008). "The complete genome of an individual by massively parallel DNA sequencing." *Nature* **452**(7189): 872-876.
- Wolfe K.H., Sharp P.M. and Li W.H. (1989). "Mutation rates differ among regions of the mammalian genome." *Nature* **337**(6204): 283-285.