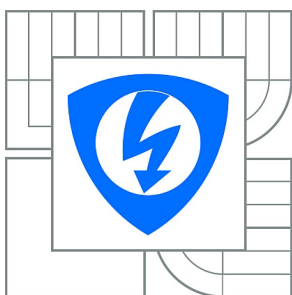


**VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ**

BRNO UNIVERSITY OF TECHNOLOGY



**FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH  
TECHNOLOGIÍ**

**ÚSTAV BIOMEDICÍNSKÉHO INŽENÝRSTVÍ**

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION  
DEPARTMENT OF BIOMEDICAL ENGINEERING

# **OBRAZOVÁ ANALÝZA MITOTICKÝCH CHROMOSOMŮ**

DIGITAL IMAGE ANALYSIS OF MITOTIC CHROMOSOMES

**DIPLOMOVÁ PRÁCE**

MASTER'S THESIS

**AUTOR PRÁCE**

AUTHOR

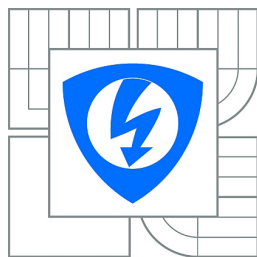
**Bc. MARIANA HÁVOVÁ**

**VEDOUCÍ PRÁCE**

SUPERVISOR

**Ing. HELENA ŠKUTKOVÁ**

BRNO 2014



VYSOKÉ UČENÍ  
TECHNICKÉ V BRNĚ

Fakulta elektrotechniky  
a komunikačních technologií

Ústav biomedicínského inženýrství

# Diplomová práce

magisterský navazující studijní obor  
Biomedicínské a ekologické inženýrství

**Studentka:** Bc. Mariana Hávová

**ID:** 149080

**Ročník:** 2

**Akademický rok:** 2013/2014

## NÁZEV TÉMATU:

### Obrazová analýza mitotických chromosomů

#### POKYNY PRO VYPRACOVÁNÍ:

1) Proveďte literární rešerši cytogenetických metod pro identifikaci chromosomů z mikroskopických snímků a zhodnoťte jejich výhody a nevýhody pro analýzu počtu a struktury chromosomů. 2) Seznamte se s technikami zpracování obrazových snímků cytogenetické analýzy chromosomů pro určení karyotypu organismů a vyhodnocení změn v počtu a struktuře chromosomů způsobujících genetická onemocnění. 3) Navrhněte a realizujte algoritmus pro vyhodnocení karyotypu člověka ze snímků mitotických chromosomů. 4) Algoritmus rozšířte o vyhodnocení strukturálních změn chromosomů oproti typickému lidskému karyotypu s využitím chromosomových ideogramů. 5) Vytvořte programové rozhraní v Matlabu pro vyhodnocení lidského karyotypu, chromosomových ideogramů a genetických změn. 6) Program otestujte na volně dostupných cytogenetických snímcích lidských chromosomů a proveďte diskuzi výsledků.

#### DOPORUČENÁ LITERATURA:

[1] GAJENDRAN, V., RODRI, X a J. J. GUEZ. Chromosome counting via digital image analysis. In Image Processing, 2004. ICIP '04. 2004 International Conference on. 2004, vol. 5, p. 2929-2932 Vol. 2925.

[2] PIPER, JIM a ERIK GRANUM. On fully automatic feature measurement for banded chromosome classification. Cytometry, 1989, 10(3), 242-255.

**Termín zadání:** 10.2.2014

**Termín odevzdání:** 23.5.2014

**Vedoucí práce:** Ing. Helena Škutková

**Konzultanti diplomové práce:**

**prof. Ing. Ivo Provazník, Ph.D.**

*Předseda oborové rady*

**UPOZORNĚNÍ:**

Autor diplomové práce nesmí při vytváření diplomové práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

## **Abstrakt**

Změny v počtu a ve struktuře chromozomů jsou příčinou řady vážných onemocnění. K odhalení chromozomálních změn slouží cytogenetická vyšetření, která nejčastěji vedou k sestavení karyotypu. Pro účely cytogenetických analýz se chromozomy vizualizují pomocí vhodných metod a nejčastěji se následně sestavují do karyotypu. Protože ruční stanovení karyotypu je časově i finančně náročné, vyvíjí se přístupy k automatickému karyotypování pomocí počítačového softwaru.

Automatické karyotypovací systémy klasifikují chromozomy do tříd na základě identifikačních znaků, specifických pro každý chromozom. Automatickou klasifikaci však nejvíce limituje přítomnost překrývající se a silně ohnutých chromozomů, přítomných v téměř každé mitóze. Přesnost a spolehlivost karyotypovacích systémů stále závisí na zásahu uživatele. Cílem vývoje nových přístupů k automatickému karyotypování je tedy zejména překonání výše zmíněných problémů a dále vývoj takových klasifikačních metod, které umožní klasifikaci chromozomů do párů bez lidské kontroly.

## **Klíčová slova**

Karyotyp, automatické karyotypování, chromozomální znaky, vizualizace chromozomů, chromozomální aberace, shluková analýza, distanční matice, dendrogram

## **Abstract**

Changes in chromosome number and structure may cause serious diseases. Cytogenetic tests leading to set of karyotype are done for detecting these abnormalities. Chromosomes are visualised with proper methods and karyotype is made up most often. Manual karyotyping is time-consuming and expensive task. Because of this, researchers have been developing automated karyotyping systems.

Karyotyping systems classify chromosomes into classes based on their characteristic features. Overlapping and bent chromosomes are limitations for automatic classification since they occur at almost every mitosis. Accuracy and reliability of karyotyping systems still depend on the human intervention. Overcoming of these problems and development of fully automated system is the aim of modern approaches.

## **Keywords**

Karyotype, automatic karyotyping, chromosome features, chromosome visualization, chromosome aberration, cluster analysis, distance matrix, dendrogram

HÁVOVÁ, M. *Obrazová analýza mitotických chromosomů*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2014. 60 s., 4 s. příloh. Vedoucí diplomové práce Ing. Helena Škutková.

# Prohlášení

Prohlašuji, že svoji diplomovou práci na téma *Obrazová analýza mitotických chromosomů* jsem vypracovala samostatně pod vedením vedoucího diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené diplomové práce dále prohlašuji, že v souvislosti s vytvořením této diplomové práce jsem neporušila autorská práva třetích osob, zejména jsem nezasáhla nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědoma následků porušení ustanovení § 11 a následujících zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

V Brně dne 23. 5. 2014

.....

(podpis autora)

## **Poděkování**

Na tomto místě bych ráda poděkovala Ing. Heleně Škutkové za odborné vedení mé práce, za pomoc, motivující přístup a čas, který mi věnovala. Dále bych ráda poděkovala Ing. Petru Walkovi a Ing. Janu Strmiskovi za cenné rady a jejich ochotu pomáhat.

# Obsah

1	Úvod.....	9
1	Cytogenetika .....	10
1.1	Lidský karyotyp .....	10
1.2	Chromozomální aberace .....	11
1.3	Chromozomy .....	13
1.4	Stavba metafázních chromozomů.....	16
1.5	Morfologické dělení.....	16
2	Cytogenetické metody pro vizualizaci a identifikaci chromozomů.....	18
2.1	Pruhovací barvicí techniky .....	19
2.2	Molekulárně cytogenetické metody.....	20
3	Zpracování obrazových snímků mitotických chromozomů.....	24
3.1	Předzpracování obrazu.....	24
3.2	Klasifikační znaky .....	24
3.3	Detekce znaků.....	25
4	Realizace programu pro klasifikaci chromozomů .....	30
4.1	Segmentace obrazu .....	31
4.2	Separace překrývajících se chromozomů.....	31
4.3	Separace dotýkajících se chromozomů.....	35
4.4	Vyhodnocení lidského karyotypu .....	36
4.5	Získ specifických znaků chromozomů .....	36
4.6	Klasifikace .....	40
5	Diskuze výsledků .....	46
	Závěr.....	54
	Seznam použitých zdrojů.....	55
	Seznam zkratk.....	58
	Seznam obrázků.....	59



Seznam tabulek.....	60
Seznam příloh.....	60

# 1 Úvod

Sestavení karyotypu patří mezi základní cytogenetická vyšetření. Karyotyp přehledně poskytuje informaci o celkovém počtu chromozomů, pohlaví jedince a o případných chromozomálních změnách, ať už početních či strukturních. Pro diagnostiku genetických onemocnění, podmíněných zejména numerickými a většími strukturálními aberacemi, je sestavení karyotypu klíčové.

Manuální sestavení karyotypu je časově i finančně náročné a vyžaduje zkušenosti cytogenetického odborníka. Tato skutečnost motivuje vývoj počítačových programů pro automatické karyotypování. Ačkoli bývají tyto programy pro klasifikaci chromozomů označovány jako automatické, zatím se jedná spíše o polo-automatické nástroje, které usnadňují práci cytogenetickým laborantům.

Hlavní překážkou na cestě k automatickému karyotypování je tvarová rozmanitost jednotlivých chromozomů, zachycených na mikroskopických snímcích. Ačkoli sami o sobě mají chromozomy obdélníkový tvar, v preparátech určených k analýze jsou často různě stočeny nebo pokřiveny, částečně nebo úplně se překrývají nebo dotýkají. Některé chromozomy mohou být ohnuté do takové míry, že není možné na nich odhadnout klasifikační znaky. Obecně je pak analýza mnohdy ztížena kvalitou samotných snímků, které mohou být špatně nasvětleny nebo jsou v nich přítomny nežádoucí artefakty, vzniklé v důsledku barvení chromozomů.

Teoretická část práce se zabývá cytogenetickými onemocněními a jejich příčinami, vysvětluje pojem chromozomů a jejich chování během buněčného cyklu. Cílem této práce je představit techniky vizualizace chromozomů a techniky zpracování mikroskopických snímků chromozomů. Teoretická část práce se zabývá jak klasickými, tak molekulárně biologickými metodami vizualizace chromozomů.

Cílem této práce je navrhnout algoritmus, který ze vstupního obrazu G-pruhovaných chromozomů vyhodnotí karyotyp člověka. G-pruhované chromozomy byli zvoleny z toho důvodu, že G-pruhování je v cytogenetických laboratořích nejčastější metoda vizualizace chromozomů. Je rychlá, ekonomická a lze s ní detekovat nejčastější příčiny genetických onemocnění.

Samotné vyhodnocení karyotypu je založeno na identifikačních znacích, které jsou pro chromozomy charakteristické. Některé tyto znaky, jako například délka, obsah či poloha centromery jsou pro některé chromozomy společné. Znak, jakým je sled pruhů na chromozomech vytvořený G-pruhováním, je pro každý chromozom unikátní.

Teoretická část práce se zabývá cytogenetikou, samotnými chromozomy a jejich chováním během buněčného cyklu. Věnuje se genetickým onemocněním a možnostem jak co nejrychleji a s největší přesností získat informace o jejich příčině – tedy sestavení karyotypu.

# 1 Cytogenetika

Cytogenetika se zabývá studiem buněk a buněčných struktur, zejména v souvislosti s přenosem genetické informace. Cytogenetické studie se týkají především struktury a funkce chromozomů a jejich chování během buněčného dělení [1].

Cytogenetika se rozvíjí od konce 19. století, nejvíce od roku 1956, kdy Tjio a Levan určili správný počet lidských chromozomů [2]. Z této doby (1960) pochází Denverská klasifikace chromozomů do sedmi skupin. V období 50. let 20. století bylo tedy možné identifikovat pouze změny v počtu chromozomů.

Na počátku 70. let byla objevena nová technika barvení chromozomů, která umožnila jejich podélnou diferenciaci [3]. Na základě první pruhovací techniky barvení byly s postupem času vyvinuty další, které umožňují spolehlivě odhalit strukturní přestavby chromozomů větší než 5 – 10 Mb [4]. Strukturní přestavby menšího rozsahu (200 – 300 kB) jsou spolehlivě identifikovány až metodami molekulární cytogenetiky. Jejich základ byl položen na konci 80. let, kdy byla zavedena metoda hybridizace *in situ* [5].

Strukturní a početní změny chromozomů jsou předmětem zájmu klinické cytogenetiky [1]. Důsledkem vrozených genetických změn vznikají v populaci vážná genetická onemocnění. U většiny plodů, postižených chromozomovými aberacemi dochází ke spontánním abortům. Aberace sluchitelné se životem jsou spojeny s těžkými vývojovými vadami, mentální retardací či s neplodností, některé vedou k onkologickým onemocněním.

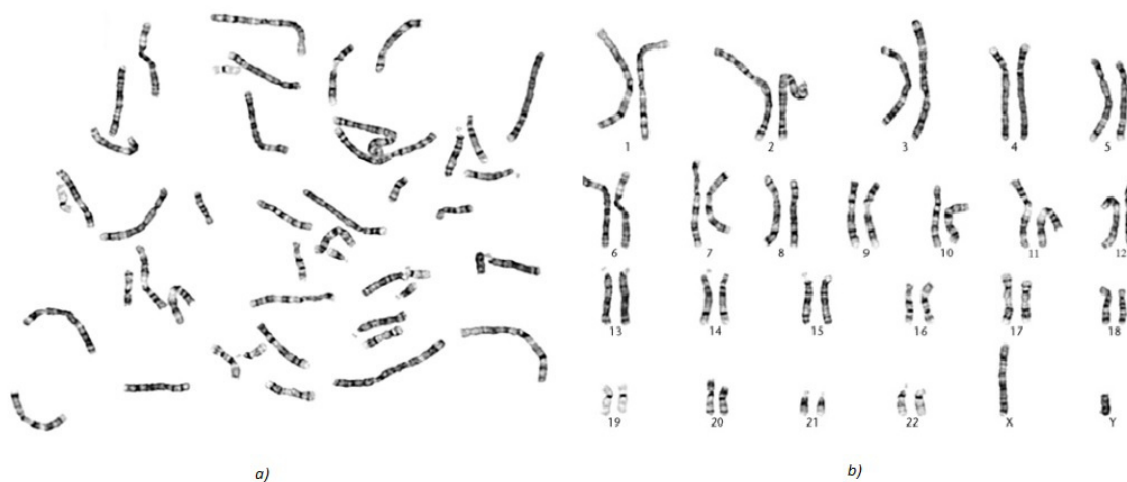
Rutinní součástí klasických cytogenetických vyšetření je sestavení karyotypu. Provádí se pro posouzení chromozomové výbavy jedinců s případnými vrozenými chromozomálními aberacemi, v prenatální diagnostice a při asistované reprodukci [6]. Své uplatnění nachází také v onkocytogenetice.

## 1.1 Lidský karyotyp

Velikost lidského diploidního genomu je  $6,4 \cdot 10^9$  párů bazí [1]. Toto množství genetické informace je rozděleno do 46 chromozomů – 22 homologních párů autozomů a 1 pár gonozomů (pohlavních chromozomů). Autozomy jsou očíslovány od 1 do 22, pohlavní chromozomy jsou označeny velkými písmeny X a Y. Gonozomy determinují pohlaví. U ženy se vyskytují v konstituci XX, u muže XY.

Karyotyp je standardní způsob obrazové prezentace lidských chromozomů pro diagnostické účely [7]. Chromozomy jsou v karyotypu vhodně orientovány a sestaveny do homologních párů. Vzhled karyotypu se řídí standardními klasifikačními pravidly ISCN (An International System for Human Cytogenetic Nomenclature), viz Obrázek 1 [8]. Na

základě karyotypu jsou prováděna cytogenetická vyšetření za účelem odhalení chromozomálních abnormalit [5].



**Obrázek 1: Chromozomy barvené pomocí G-pruhování**

a) snímek metafázních chromozomů muže: 46, XY a b) příslušný karyotyp. Převzato z [3].

## 1.2 Chromozomální aberace

Chromozomální aberace představují jakékoli odchylky od standardního počtu (numerické aberace) a struktury chromozomů (strukturní aberace) [9]. Často se jedná o rozsáhlé změny, které nejsou slučitelné se životem. Aberace, které se vyskytují u narozených jedinců mají většinou závažné klinické projevy. Existují však i takové aberace, které se nemusí fenotypově vůbec projevit nebo jsou její nositelné pouze přenašeči a aberace se může projevit až u příští generace.

Chromozomální změny jsou příčinou poloviny spontánních abortů během prvního trimestru [10]. Nejčastější změnou v karyotypu je trizomie (27%), dále polyploidie (10%), monozomie pohlavních chromozomů (9%) a strukturní přestavby (2%).

### Numerické aberace

Mezi numerické aberace patří polyploidie a aneuploidie [9]. Polyploidie, zmnožení celé sady chromozomů, nejsou u člověka slučitelné se životem. U narozených jedinců je popsána pouze aneuploidie, ztráta či nadbytek jednoho, vzácněji více chromozomů [5]. Aneuploidie postihují jak autozomy tak gonozomy.

Nejznámější a nejčastěji se vyskytující aneuploidii u člověka je trizomie chromozomu 21, známá jako Downův syndrom [1]. Lidé s Downovým syndromem trpí různými tělesnými malformacemi, mají charakteristický šikmý vzhled očí a v dlani tzv. opičí rýhu. Vyskytuje se u nich lehká mentální retardace.

Kromě Downova syndromu jsou v klinické praxi popsány trizomie chromozomu 13 (Patauův syndrom) a 18 (Edwardsův syndrom). Obě tyto trizomie však mají tak závažný klinický projev, že postižení jedinci umírají v brzké době po narození.

Numerické aberace, které postihující gonozomy a jsou slčitelné se životem jsou trizomie X a Klinefelterův syndrom s karyotypovou konfigurací 47, XXY. Trizomie X se vyskytuje u žen s obvykle normálním fenotypem. V některých případech je u postižených trizomií X snižená plodnost a lehká mentální retardace. Klinefelterův syndrom postihuje muže, kteří jsou neploidní. Fenotypovým projevem Klinefelterova syndromu je vysoké štíhlé tělo, malá varlata, zvětšená prsa o řídké ochlupení [1].

Monozomie existuje u člověka pouze jediná, a to monozomie chromozomu X, známá jako Turnerův syndrom. Vyskytuje se u žen, které jsou sterilní, s opožděným sexuálním vývojem, mají malou postavu a charakteristickou kožní řasu na krku.

### **Strukturní aberace**

Strukturní abnormality jsou méně časté než numerické a téměř nemají klinický význam [1]. U některých abnormalit však byla zjištěna souvislost s patogenezi nádorových onemocnění. Přenašeči strukturních abnormalit jsou většinou zdraví jedinci, avšak hrozí u nich tvorba nebalancovaných gamet.

Změny ve struktuře chromozomů vznikají jako následek chromozomových zlomů. S ohledem na to, zda dochází ke ztrátě genetického materiálu či nikoliv, rozlišujeme přestavby balancované a nebalancované. U nebalancovaných přestaveb část chromozomu chybí nebo naopak přebývá.

Mezi nebalancované přestavby řadíme delece a duplikace [5]. Delece představuje ztrátu části chromozomu. Nositel delece je monozomický pro tu část genetické informace, která byla deletována. Naproti tomu při duplikacích dochází ke zmnožení části chromozomu. Protože při duplikaci buňka nepřichází o genetický materiál, je její klinický projev méně závažný než je tomu u delecí.

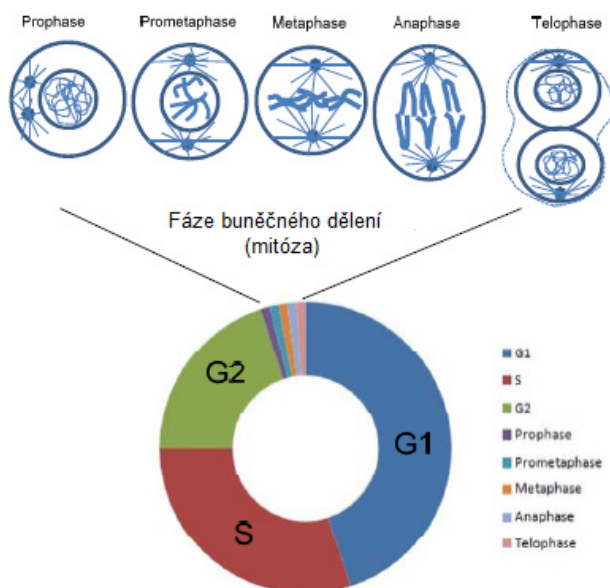
Balancovanými přestavbami jsou inverze a translokace. Během inverze dochází k přerušení chromozomu dvěma zlomy, uvolnění segmentu a jeho opětovné začlenění do chromozomu v inverzní podobě. U translokací dochází k přemístění uvolněného chromozomového segmentu na jiný chromozom.

Známy jsou reciproké translokace, představující výměnu segmentů mezi nehomologními chromozomy, kdy nedochází ke ztrátě genetického materiálu. Tyto aberace jsou běžné a neškodné. Další známou translokací je Robertsonská translokace, při které fúzují dva akrocentrické chromozomy a ztrácí svá krátká raménka. Robertsonská translokace nemá škodlivé následky [5].

### 1.3 Chromozomy

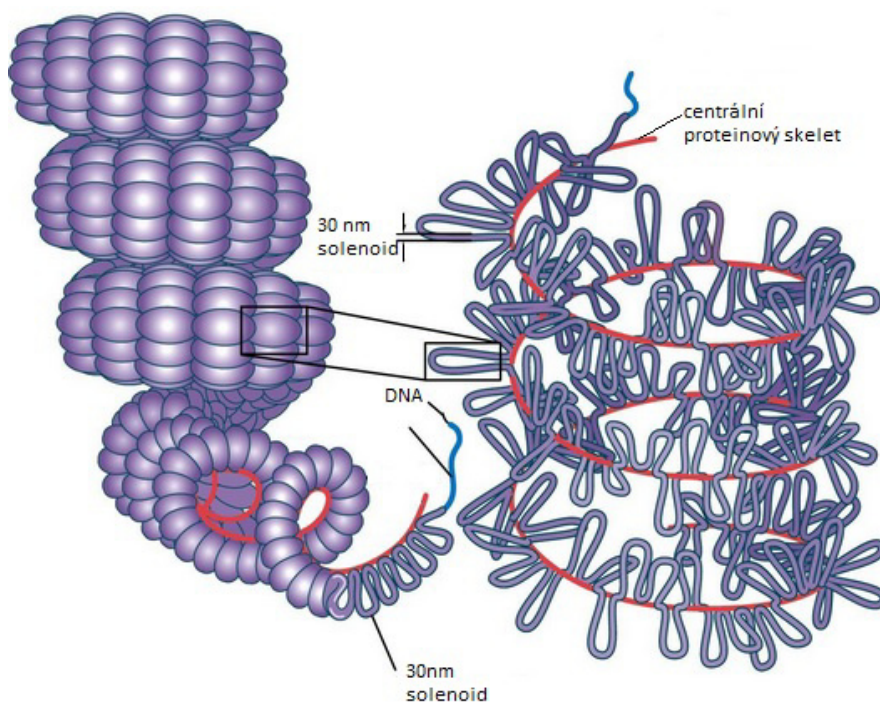
Chromozomy jsou jaderné útvary, tvořené chromatinem [1]. Na výstavbě chromatinu se společně s molekulou DNA podílejí bazické nukleoproteiny – histony a kyselé proteiny nehistonové povahy. Histony lze rozdělit do pěti základních tříd: H1, H2A, H2B, H3 a H4. Tyto proteiny zodpovídají za správné sbalení nukleohistonového (chromatinového) vlákna při kondenzaci chromozomu, viz **Chyba! Nenalezen zdroj odkazů..** Oktamer histonů H2A, H2B, H3 a H4 vytváří společně s ovinutou dvoušroubovicí DNA útvar nazývaný nukleozom. Jednotlivé nukleozomy jsou spojeny mezerníkovou sekvencí DNA navázanou na histon H1, jenž celý systém stabilizuje.

Struktura chromatinu se odvíjí od fáze buněčného cyklu, který je zobrazen na Obrázek 2. Během interfáze (fáze mezi buněčnými děleními) se chromatin vyskytuje buď ve své primární nebo sekundární struktuře. Primární strukturu představuje samotné nukleohistonové vlákno široké 10 nm, sekundární struktura je pak reprezentována cylindrickým solenoidem. Solenoid vzniká svinutím primárního vlákna a je 30 nm široký.



Obrázek 2: Buněčný cyklus a fáze buněčného dělení

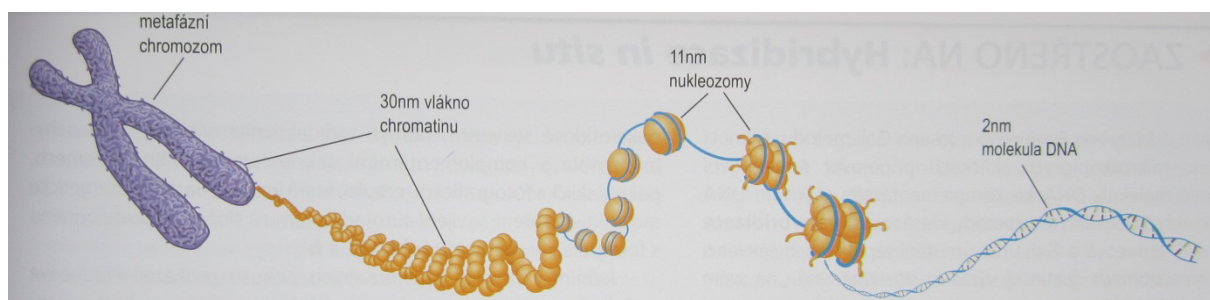
Další kondenzace probíhá až na počátku mitotického dělení a vede k mikroskopicky viditelným chromozomům. Buněčné dělení se skládá ze čtyř hlavních fází: profáze, metafáze, anafáze a telofáze. V profázi se solenoid stáčí do smyček, radiálně odbočujících z centrálního proteinového skeletu nehistonové povahy. Další kondenzací radiálních smyček vznikají vlákna široká okolo 300 nm. Metafáze představuje poslední fázi kondenzace chromozomu, kdy se vlákna skládají v 600 nm širokou chromatidu. Proces kondenzace názorně ukazuje Obrázek 3.



**Obrázek 3: Proces kondenzace chromozomu.**

Viz text. Převzato a upraveno z [50].

V metafázi jsou chromozomy kondenzovány na setiny až tisíce své původní délky. Dosahují velikosti v řádu desítek mikrometrů a jsou tedy vhodné k pozorování pod světelným mikroskopem. Kondenzované metafázní chromozomy jsou také nejvhodnější pro barvení jadernými barvivy.

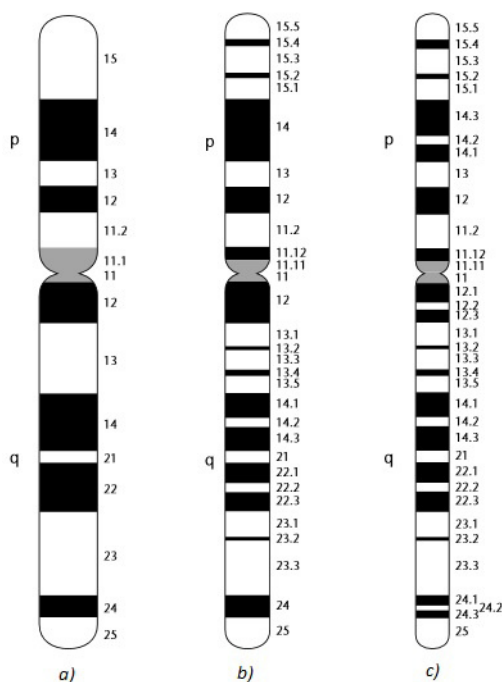


**Obrázek 4: Různé uspořádání DNA v chromozomech**

Na obrázku lze pozorovat navinutí dvoušroubovice DNA na histonový oktamer (11nm nukleozom) a svinutí primárního vlákna v 30nm široký solenoid až po výsledný metafázní chromozom. Převzato z [1].

Počet identifikovatelných pruhů (tzv. band resolution) je závislý na stupni mitotického dělení. Platí, že čím je stadium mitózy ranější, tím více pruhů lze na chromozomu detekovat [11]. Počet viditelných pruhů na haploidní sadě chromozomů jedné buňky, zachycených v pozdní metafázi (nejčastěji), je přibližně 350. V polovině profáze je pruhů 1250 – 2000 [12].

Na Obrázek 5 lze vidět ideogramy pro různé stupně kondenzace chromozomu. Ideogramy představují standardní referenci. Jedná se o schématickou reprezentaci chromozomů, která vyjadřuje jejich relativní velikost a sled pruhů. Na ideogramu lze také rozlišit polohu centromery a krátké (*p*-) a dlouhé (*q*-) raménko. Ideogramy jsou definovány mezinárodní klasifikací ISCN [8].



**Obrázek 5: Ideogramy G-pruhovaného lidského chromozomu č. 11.**

Rozlišení a) 350 pruhů, b) 550 pruhů a c) 850 pruhů. Převzato a upraveno z [11]

Vhodnost určitého stupně kondenzace chromozomu pro detekci abnormalit je dán doporučeními Asociace pro klinickou cytogenetiku. Tabulka udává souvislost detekce určitých aberací s počtem pruhů na chromozomech

**Tabulka 1: Minimální rozlišení na G-pruhovaných chromozomech pro detekci určité abnormality**

Typ abnormality	počet detekovatelných pruhů
aneuploidie	< 300
velká strukturní přestavba	300
menší strukturní přestavba	400
mikrodelece	700

Z předchozího textu vyplývá, že znalost míry kondenzace analyzovaného chromozomu je zásadní pro správnou identifikaci chromozomu podle standardní reference. Na základě informace o přítomném počtu pruhů jsou také detekovány případné chromozomální aberace. Nejmenší počet pruhů haploidní sady se pohybuje kolem 300, je



tedy zřejmé, že určení počtů pruhů jejich počítáním by bylo časově náročné. V cytogenetických laboratořích bývá počet pruhů odhadován na základě určitých znaků chromozomů, viditelných při určitém stupni kondenzace chromatinu [11]. Výběr některých znaků, umožňujících určení počtu pruhů na chromozomech je uveden v Tabulka 2.

**Tabulka 2: Identifikační znaky pro hodnocení stupně kondenzace chromozomu**

Převzato a upraveno z [11].

„Band resolution” snímku	Viditelné znaky
300	2 tmavé pruhy na 8p 3 tmavé pruhy na 10q viditelné 20p12
500	rozlišení 7q33 a 7q35 3 tmavé pruhy na 11p viditelné 14q32.2 a tmavé pruhy na 18q
700	rozlišení 2p25.2 rozlišení 2q37.2 rozlišení 10q21.1 a 10q21.3

#### 1.4 Stavba metafázních chromozomů

Na metafázních chromozomech lze pozorovat dvě chromatidy. Dvouchromatidový chromozom vzniká po replikaci DNA před jaderným dělením. Jeho charakteristický tvar lze vidět na **Chyba! Nenalezen zdroj odkazů.** Obě chromatidy obsahují stejnou genetickou informaci a jsou tvořeny jedinou molekulou DNA.

Obě chromatidy se stýkají v místě primární konstrikce neboli centromery, kterou tvoří chromatin s vysokým obsahem repetitivních sekvencí [1]. Centromera udržuje obě chromatidy ve spojení a umožňuje napojení na mikrotubuly dělicího vřetenka (přes kinetochor), čímž zodpovídá za správný rozchod chromozomů během buněčného dělení.

Konce chromozomů – telomery, jsou podobně jako centromera, tvořeny repetitivními sekvencemi. Na rozdíl od centromer jsou však repetitivní sekvence telomer u všech chromozomů shodné (TTAGGG) a liší se pouze počtem opakování.

#### 1.5 Morfologické dělení

Oblast primární konstrikce (centromery) rozděluje chromozom na krátké (p-) a dlouhé (q-) raménko [1]. Jednotlivá ramena chromozomů jsou charakteristická sledem světlých a tmavých pruhů, které číslujeme arabskými číslicemi směrem od centromery ke

konci chromozomu. Např. 20p12 označuje pruh na krátkém raménku chromozomu 20, který se nachází v blízkosti centromery. Při určitém stupni kondenzace jsou ještě některé základní pruhy rozděleny na další subpruhy, např. 5p13.1.

Podle poměru kratšího ramene k celkové délce chromozomu (centromerický index CI) rozlišujeme tři, někdy čtyři hlavní skupiny chromozomů – metacentrické, (submetacentrické), akrocentrické a telocentrické. Morfologické dělení chromozomů na základě centromerického indexu uvádí Tabulka 3. Chromozomy 1, 3, 16, 19 a 20 jsou metacentrické, 13, 14, 15, 21 a 22 jsou akrocentrické [13]. Zbývající chromozomy patří do skupiny submetacentrických chromozomů. Telocentrické chromozomy se v lidských buňkách nevyskytují [14].

**Tabulka 3: Rozlišení chromozomů podle centromerického indexu**

Typ chromozomu	Centromerický index
Metacentrický	0,39 - 0,50
Submetacentrický	0,13 – 0,38
Akrocentrický	do 0,13
Telocentrický	0

## 2 Cytogenetické metody pro vizualizaci a identifikaci chromozomů

Chromozomy určené pro analýzu karyotypu se získávají v cytogenetických laboratořích z buněk zastavených v metafázi či prometafázi buněčného dělení pomocí kolchicinu, alkaloidu který napadá mikrotubuly dělicího vřeténka a znemožňuje tak další postup dělení [12].

Pro cytogenetická vyšetření se využívají buňky periferní krve, nejčastěji lymfocyty, případně fibroblasty [15]. Prenatální diagnostika využívá buněk z plodové vody, onkocytogenetika buněk odebíraných přímo z nádoru či kostní dřeně. Po odběru se buňky kultivují na vhodném médiu. Řada buněčných typů při kultivaci spontánně proliferuje, některé typy (např. lymfocyty periferní krve) je k dělení potřeba stimulovat pomocí mitogenů (látek, které navozují mitotické dělení) [16].

Po sběru jsou buňky vystaveny působení hypotonického solného roztoku (nejčastěji KCl). Jádru nabírá voda a zvětšuje se. Uvolněné chromozomy jsou fixovány na podložním skle a obarveny jadernými barvivy. Nejlepších výsledků je dosahováno při inkubaci 10 – 30 min [16]. Není-li působení roztoku dostatečné, chromozomy jsou stažené a téměř je nelze vizualizovat. Naopak příliš dlouhé působení roztoku vede k velkému rozptýlu chromozomů, případně k protržení jádra a jejich ztrátě.

Působení hypotonického roztoku je zastaveno pomocí Carnoyova fixačního roztoku: 60% etanol, 30% chloroform a 10% ledová kyselina octová. Fixační roztok také udržuje jádro ve stabilním stavu a umožňuje barvení chromozomů. Chromozomy je možné na podložní sklo fixovat buď přímo – roztlaková metoda nebo ze suspenze po centrifugaci – nakapáním. Po odpaření fixačního roztoku jsou chromozomy připraveny k barvení.

Před objevem pruhovacích technik v 70. letech 20. století bylo využíváno tzv. konvenčních technik barvení. Konvenčně barvené chromozomy jsou barevně homogenní po celé délce [3]. Toto barvení umožňuje rozdělit chromozomy do sedmi skupin (A – G) s ohledem na jejich délku a polohu centromery. Až s rozvojem pruhování, které umožňuje vizualizovat podélnou diferenciaci chromozomů lze specificky identifikovat jednotlivé chromozomy.

Pruhovací barvicí techniky umožňují odhalit pouze chromozomové přestavby větší než 4 - 6 milionů párů bazí [17]. Za účelem identifikace změn v řádu tisíců párů bazí byla v roce 1969 poprvé zavedena do laboratorní praxe metoda *in situ* hybridizace (ISH) [3]. Zároveň s tímto objevem byly položeny základy pro rozvoj molekulárně cytogenetických přístupů pro vizualizaci chromozomů. Dnes jsou tyto metody běžně využívány v klinické praxi. Základní moderní metodou pro vyšetření strukturních a početních změn chromozomů je fluorescenční hybridizace *in situ* (FISH).

Chromozomy fixované na podložním skle a obarvené G-pruhovací technikou lze pozorovat pomocí světelného mikroskopu s imerzním objektivem (100 x zvětšení) [1]. Fluorescenčně značené chromozomy lze pozorovat ve fluorescenčním mikroskopu. Obraz chromozomů je nejprve hodnocen zkušeným cytogenetickým laborantem. Mitózy vhodné v analýze mohou být poměrně vzácné, zároveň některé chromozomy mohou chybět v důsledku procesu zpracování. Mikroskopy jsou vybaveny digitálním fotoaparátem pro mikrofotografii, který umožňuje převod obrazu do počítače pro jeho další zpracování

## **2.1 Pruhovací barvicí techniky**

Pruhovací barvicí techniky umožňují vizualizovat chromozom pomocí sledu světlých a tmavých pruhů [15]. Pruh je definován jako oblast chromozomu, která je svojí intenzitou odlišná od svého bezprostředního okolí. Každý chromozom disponuje unikátním vzorcem pruhů, tedy je sled pruhů na chromozomu znak vhodný pro identifikaci. Nejčastěji se k barvení využívá G-pruhování a R-pruhování [12].

### **Q-pruhování (quinacrine)**

Pro Q-pruhování jsou využívána fluorescenční barviva [3]. Poprvé byly Q-pruhy pozorovány za použití barviva chinakrinu. Na chromozomech lze s pomocí fluorescenčního mikroskopu pozorovat specifické světlé a tmavé pruhy. Tyto pruhy odpovídají specifitě jednotlivých fluorochromů ke konkrétním párům bazí v DNA.

AT-specifické fluorochromy (např. chinakrin, DAPI, Hoechst 33258) vytváří na chromozomech průhledné pruhy (Q-pruhy), které svojí pozicí odpovídají G-pruhům. GC-specifické fluorochromy (např. chromomycin) zobrazují pruhy neprůhledné (zářící) a tyto jsou pozičně podobné R-pruhům.

### **G-pruhování (Giemsa)**

G-pruhy se vytváří na chromozomech, které jsou nejprve vystaveny působení proteolytického enzymu (nejčastěji trypsinu) a poté jsou obarveny Giemsovým barvivem. Vzniklý sled tmavých a světlých pruhů je pro každý chromozom jedinečný. Ačkoli existuje celá řada barvicích technik, vhodných pro analýzu karyotypu, v cytogenetických laboratořích se nejčastěji využívá právě G-pruhování [11]. Důvodem je finanční i časová nenáročnost při současném zachování požadované informace. Obrázek 1 zobrazuje chromozomy barvené pomocí G-pruhování.

### **R-pruhování (reverse)**

Inkubací preparátů v horkých roztocích pufrů s následným obarvením pomocí Giemsova barviva (nejčastěji) vznikají na chromozomech R-pruhy [3]. R-pruhované chromozomy mají obrácený jas pruhů oproti G-pruhování a Q-pruhování. R-pruhy se

vytváří v oblastech genově bohatých oblastech chromatinu. Této skutečnosti je možno využít pro detekci malých strukturních přestaveb, které vedou k fenotypovým projevům [16].

### **C-pruhování (centromere)**

C-pruhy vznikají na chromozomech po jejich inkubaci v roztoku hydroxidu barnatého, opět s následným barvením Giemsovým barvivem [3]. C-pruhy se objevují v oblastech, kde je vysoký obsah repetitivních sekvencí. Takovými oblastmi jsou zejména centromery, telomery, interkalární úseky a oblasti přilehlé k NOR. C-pruhování je využíváno jako doplňkové vyšetření pro verifikaci aberací postihujících centromery, případně další heterochromatinové úseky [1].

### **Pruhování s vysokým rozlišením (high-resolution banding)**

Tento typ pruhování umožňuje vizualizovat chromozomy v profázi či ranné metafázi buněčné dělení [16]. Chromozomy nejsou plně kondenzované a jejich „band resolution“ je maximální. High-resolution pruhování lze dosáhnout několika způsoby. Buněčné cykly v kultuře jsou synchronizovány a buňky jsou vystaveny působení kolcemidu. Kolcemid je derivát kolchicinu, vřetenkové jedu. Pro high-resolution pruhování se používá z toho důvodu, že jeho působení není tak silné jako působení kolchicinu. Použitá dávka je navíc na dolní prahové hranici.

Podobného výsledku jako při působení kolcemidu lze dosáhnout jsou-li buňky kultivovány společně s látkami, které mají schopnost se inkorporovat do DNA a bránit kondenzaci chromozomů.

## **2.2 Molekulárně cytogenetické metody**

### **Fluorescenční hybridizace *in situ***

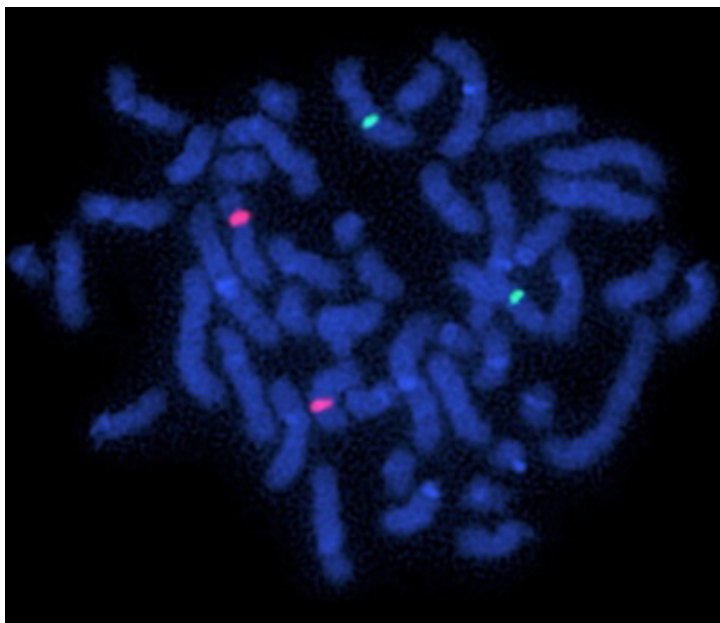
Pomocí FISH lze lokalizovat a vizualizovat žádané sekvence DNA, a to jak na metafázních, tak na interfázních chromozomech [5]. Možnost detekce konkrétních úseků na interfázní DNA představuje velkou výhodu této metody. V klinické praxi tím odpadá nutnost kultivace buněk, a tak se podstatně zkracuje doba nutná k získání výsledků cytogenetického vyšetření.

Základem přístupu FISH je proces hybridizace, tedy komplementární spojení dvou vláken nukleových kyselin. Jako matrice slouží jeden řetězec vyšetřované DNA, ke kterému se váže uměle připravená sekvence (sonda). DNA sondy mohou být komplementární k určitým genům, chromozomovým úsekům nebo i k celým chromozomům.

Aby se oba řetězce mohli spojit, je nejdříve nutné původní dvoušroubovici denaturovat, nejčastěji za působení vysoké teploty (100°C) v přítomnosti destabilizujících

látek. Denaturace probíhá přímo na podložním skle, tzv. „na místě“, tedy *in situ*, odtud název metody.

V době svého objevu byly hybridizační DNA sondy za účelem vizualizace značeny radioaktivní látkou (např.  $^3\text{H}$ ,  $^{14}\text{C}$ ) [5]. Dnes se používá výhradně fluorescenčního značení, a to buď přímo samotnými fluorescenčními barvivy nebo protilátkami s vázanými fluorescenčními barvivy. Chromozomy hybridizované s DNA sondou značenou fluorescenční látkou lze pozorovat ve fluorescenčním mikroskopu.



**Obrázek 6:** Mikroskopický snímek metafázních chromozomů hybridizujících se sondami specifickými pro centromerickou oblast chromozomů 7 (zeleně) a 8 (červeně)

Problémem „klasické“ FISH je malý počet rozeznatelných sekvencí během jedné hybridizace (aplikace 3 - 4 značených sond). Další nevýhodou je nutnost předem určit detekované sekvence. FISH tedy není možné účelně využít k detekci neznámých aberací, ale pouze k ověření jejich přítomnosti.

Za účelem odstranění těchto nevýhod byli vyvinuty moderní metody, které umožňují simultánní vizualizaci všech chromozomů v rozdílných barvách [49]. V klinické praxi se nejčastěji využívá M-FISH, SKY a CGH. Dvě z těchto metod – M-FISH a CGH jsou vhodné k vytvoření karyotypu a také se k tomuto účelu využívají.

### **M-FISH**

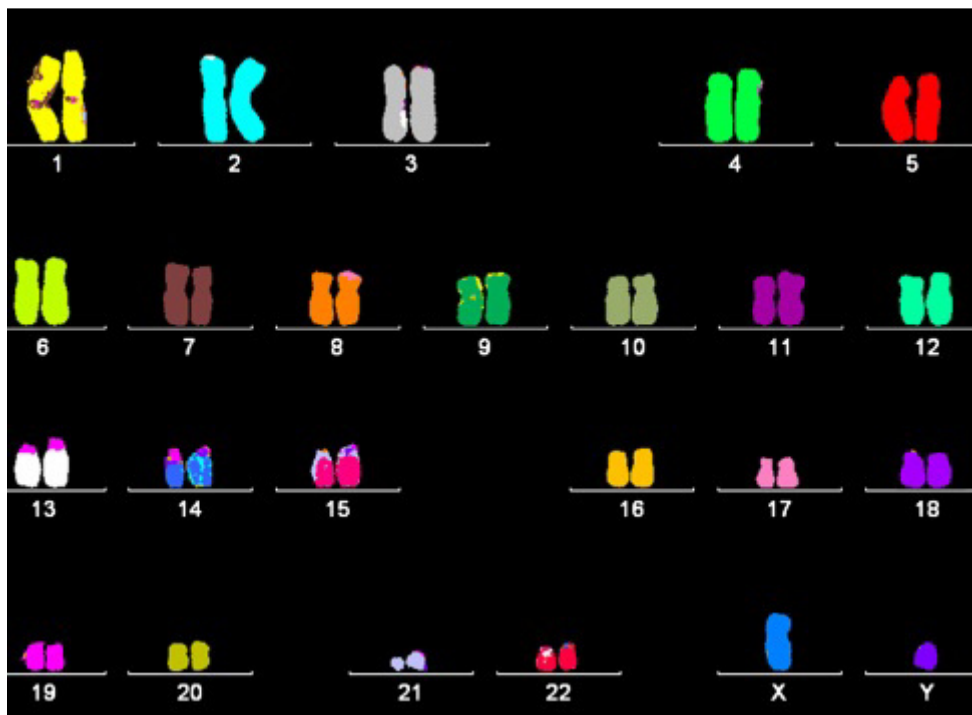
Multiplex (multicolor) fluorescence *in situ* hybridization byla představena v roce 1996 [18]. Využívá 24 rozdílně značených sond a umožňuje tak rozlišení všech 24 chromozomů. Sondy jsou značeny jak jednotlivými fluorochromy, tak jejich různou vzájemnou kombinací.

Počet kombinací  $k$  je dán vzorcem:

$$k = 2^n - 1, \quad (1)$$

kde  $n$  je počet použitých fluorochromů. Je zřejmé, že při použití již pěti spektrálně odlišných fluorochromů lze detekovat všech 24 chromozomů.

S využitím úzkopásmových filtrů jsou nasnímány samostatné obrazy pro každý z pěti fluorochromů. Následně jsou získané snímky zkombinovány a jednotlivým chromozomům je přiřazena specifická barva. Z výše uvedeného vyplývá, že M-FISH neumožňuje rozlišit strukturní aberace na jednom chromozomu, např. inverze, delece nebo duplikace neboť tyto jsou vyobrazeny stejnou barvou [19].



Obrázek 7: Karyotyp sestavený na základě metody M-FISH.

Převzato z [20].

### SKY

Princip značení chromozomů je u spektrálního karyotypování stejný jako u M-FISH, tyto dvě metody se od sebe liší pouze procesem snímání. Světlo emitované z fluorochromů prochází přes interferometr a je zobrazeno pomocí CCD kamery [21]. Vzniklé interferogramy prochází Fourierovou transformací a ze získaných příslušných spekter je možné jednotlivým chromozomům přiřadit falešné barvy.

SKY je společně s M-FISH vhodnou metodou pro detekci drobných interchromozomálních přestaveb, které jsou pod rozlišovací schopností pruhovacích technik [17].

## CGH

Komparativní genomová hybridizace (Comparative Genomic Hybridization) je metoda, která slouží k detekci a mapování jednotlivých genových sekvencí nebo celých chromozomů [22]. Její princip spočívá v měření rozdílů v počtu kopií nebo dávky konkrétního chromozomového úseku mezi dvěma různými vzorky DNA.

Jako sonda slouží DNA izolovaná ze vzorku vyšetřované tkáně. Tato DNA se značí fluorochromem zelené barvy. Vyšetřovaná DNA hybridizuje s kontrolní, zdravou DNA, která se značí červeným fluorochromem. Při stejném zastoupení bazí v obou vzorcích detekujeme signál červené a zelené sondy v poměru 1:1. Pokud v jednom ze vzorků genetický materiál přebývá nebo naopak chybí, detekovaný poměr se odsouvá od poměru 1:1. Je-li poměr menší než jedna, pak jsou na sondě určité úseky deletovány, při poměru větším než jedna jsou úseky amplifikovány.

CGH umožňuje citlivou identifikaci delecí a amplifikací, o kterých není předem nutné vědět kde se nacházejí [23]. Největší uplatnění CGH je v onkocytogenetice pro studium chromozomových aberací u nádorových buněk.



# 3 Zpracování obrazových snímků mitotických chromozomů

Rostoucí počet rutinních cytogenetických analýz a zvyšující se nároky na rychlost zpracování výsledků, motivují vývoj metod pro poloautomatické či automatické sestavení karyotypu. Různé přístupy ke karyotypování pomocí počítače se vyvíjejí od 60. let.

Pro sestavení karyotypu je nezbytná identifikace všech 46 chromozomů. Automatické sestavení karyotypu pomocí počítačového softwaru se děje v několika základních krocích: předzpracování obrazu, segmentace (detekce chromozomů), získání specifických znaků a klasifikace [6].

## 3.1 Předzpracování obrazu

Účelem předzpracování digitalizovaného mikroskopického snímku je získat obraz v takové formě, která nejvíce podporuje nalezení specifických znaků, které jsou vhodné k identifikaci jednotlivých chromozomů [24]. Na základě těchto znaků je následně provedena klasifikace.

Základem je převod vstupního RGB obrazu na šedotónový a pro některé další operace také na binární obraz. V závislosti na kvalitě snímku je možné využít filtraci obrazu k odstranění šumu. Aplikace např. mediánového filtru je vhodná pro odstranění šumu typu „sůl a pepř“ a také k vyhlazení hran chromozomů, což usnadňuje další práci s obrazem a vede k přesnější analýze znaků [24]. K separaci jednotlivých chromozomů od pozadí se nejčastěji využívá prahování. Často je využíváno adaptivního prahování, hysterezní prahování nebo prahování metodou Otsu [25].

Prahovací metody sami o sobě neumožňují separaci dotýkajících se či překrývajících se chromozomů [26]. Tento problém je jednou z hlavních překážek v plně automatickém karyotypování. Z tohoto důvodu lze systémy pro automatické karyotypování považovat spíše za nástroje pro usnadnění než za plně automatizované systémy, nevyžadující zásah lidského operátora.

## 3.2 Klasifikační znaky

Základem automatického či poloautomatického karyotypování je nalezení identifikačních znaků chromozomů, na základě kterých mohou být chromozomy spolehlivě zařazeny do jedné z 24 skupin [15]. Nejčastěji se k tomuto účelu využívá délka chromozomu, centromerický index (CI) a charakteristický sled pruhů.

Analyzované znaky vyžadují různou míru apriorní informace, nutné k jejich výpočtu [27]. První kategorie obsahuje znaky, které lze odečíst přímo ze samotného segmentovaného snímku (např. plocha), v druhé kategorii jsou znaky, které pro výpočet vyžadují znalost podélné osy (např. délka), třetí kategorie obsahuje znaky založené na znalosti jak podélné osy tak příslušných profilů. Poslední, čtvrtá kategorie zahrnuje znaky, vyžadující znalost polohy centromery a orientace chromozomu (např. index centromery).

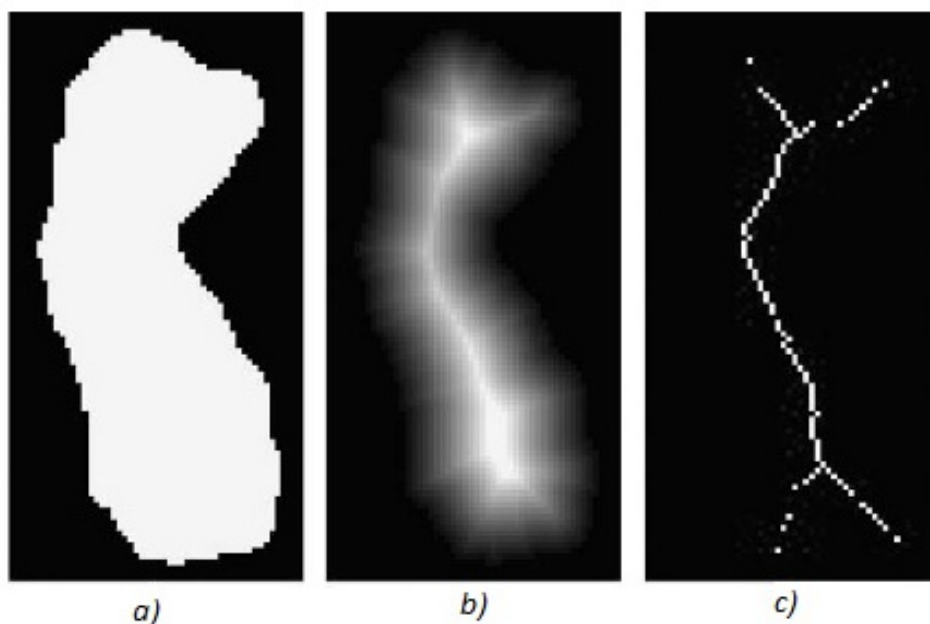
Analyzované znaky lze rozdělit do dvou hlavních skupin [28]. První skupinu tvoří tzv. geometrické znaky, druhá skupina znaků vychází z pruhovacího vzorce. Mezi nejvýznamnější geometrické znaky patří délka chromozomu a poloha centromery [29].

### **3.3 Detekce znaků**

Mitotické chromozomy zachycené na mikroskopických snímcích vykazují značnou tvarovou rozmanitost. Téměř všechny chromozomy vykazují určitý stupeň ohnutí, který často komplikuje získání charakteristických znaků [30]. Z důvodu nerigidní povahy chromozomů nelze klasifikační znaky počítat podél rovných linií (horizontálních či vertikálních). Obejít ohnutí chromozomu a získat tak znaky s dostatečnou přesností umožňuje detekce podélné osy chromozomu.

Podélná osa symetrie popisuje tvar objektu na základě topologického skeletonu [31].

Způsobů detekce podélné osy byla navržena celá řada. Nejčastěji je pro získání topologického skeletonu využívána tzv. transformace podélné osy (medial axis transform – MAT), známá také jako skeletonizace [32]. Pojmy MAT a skeletonizace jsou v literatuře často považovány za synonyma. Ve skutečnosti skeletonizace představuje operaci na binárním obraze, kdy dochází k postupnému ztenčování objektu až do zachování jeden pixel široké středové osy. MAT vychází z “distance transform” funkce [33]. V obraze, který prošel operací “distance transform” je MAT reprezentována pixely s nejvyšší hodnotou intenzity a tedy maximální vzdáleností od hranice objektu, Obrázek 8. Výsledkem obou operací je topologická kostra objektu – skeleton. Tuto kostru je možno považovat za vhodnou aproximaci podélné osy chromozomu.



Obrázek 8: a) binární obraz chromozomu, b) obraz chromozomu po použití „distance transform“ funkce, c) topologická kostra chromozomu, získaná procesem ztenčování na obraze b). Převzato a upraveno z [32].

Znalost podélné osy umožňuje učinit další krok v analýze – získat tzv. „profily“ [27]. Pod pojmem "profil" rozumíme výsledek převodu 2D obrazu chromozomu na 1D graf, viz Obrázek 9. Profil je tedy jednodimenzionální funkce jedné proměnné a reprezentuje vhodné vlastnosti chromozomu, nejčastěji šířku (profil tvaru) a sled pruhů (jasový profil).

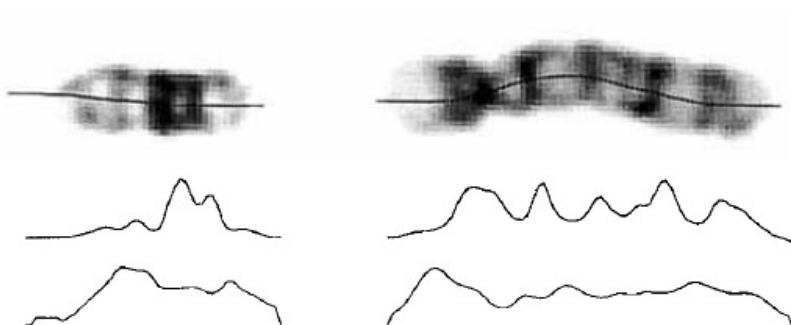
**Tvarový profil** reprezentuje šířku chromozomu. Jeho výpočet je realizován podél kolmice k podélné ose v každém jejím bodě. Je definován jako:

$$\sum_{i=1}^n (g_i \times d_i) / \sum_{i=1}^n g_i, \quad (2)$$

tedy podíl sumy součinu intenzity pixelu  $g_i$  a jeho vzdálenosti  $d_i$  od středové osy a sumy příslušných vzdáleností. Minimum tohoto profilu indikuje potenciální pozici centromery.

**Jasový profil** nese informaci o specifickém sledu pruhů jednotlivých chromozomů [27]. Hodnoty jasového profilu jsou, stejně jako v předchozím případě, vypočteny jako průměrná hodnota intenzit pixelů ve směru kolmém na podélnou osu v určité zvolené sekvenci bodů:

$$\sum_{i=1}^n g_i/n. \quad (3)$$



**Obrázek 9: Chromozomy barvené pomocí G-pruhování**

Jejich příslušný jasový profil (nahore) a tvarový profil (dole). Převzato z [27].

S ohledem na využití sledu pruhů na chromozomu při jeho klasifikaci jsou důležité tři faktory [32]:

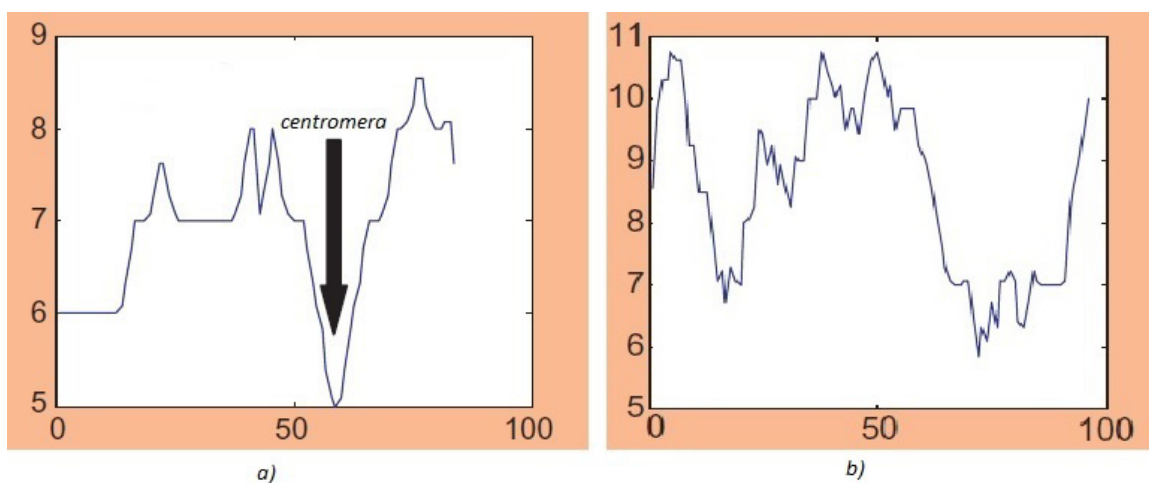
1. šířka pruhu
2. průměrná intenzita pixelů v rámci pruhu
3. relativní pozice vzhledem k centromere či ostatním pruhům.

**Délka chromozomu** je základním a zároveň významným morfologickým znakem, vhodným pro klasifikaci [34]. Pozvolna klesá od chromozomu č. 1 k chromozomu č. 22 [26]. Jeho výhodou je snadný výpočet, který za předpokladu dobré segmentace představuje počet pixelů podélné osy. Častěji než samotná délka se v analýze využívá délky normalizované:

$$I_{ri} = \frac{I_i}{I_t}, \quad (4)$$

kde  $I_i$  je délka  $i$ -tého chromozomu a  $I_t$  je součet délek všech 46 chromozomů. Délka sama o sobě klasifikaci neumožňuje, v kombinaci s polohou centromery lze však chromozomy zařadit do sedmi skupin (A- G).

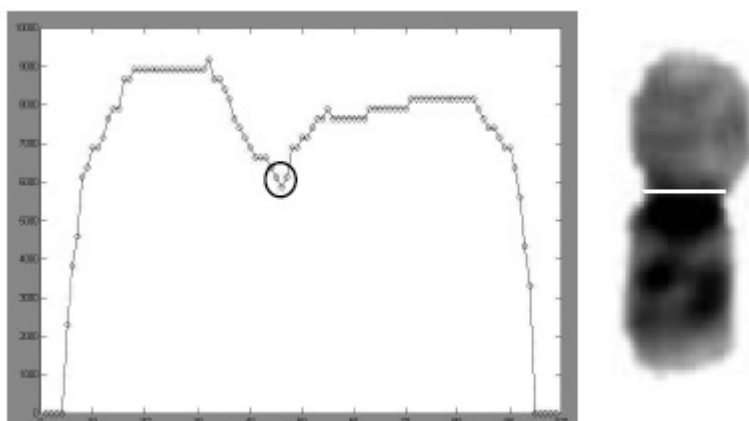
Dalším významným identifikačním znakem je **poloha centromery**. Kromě určení správné orientace chromozomu, usnadňuje znalost polohy centromery samotnou klasifikaci [35]. Z morfologického hlediska se jedná o nejužší část chromozomu v podélném směru [13]. Pro automatickou lokalizaci centromery byla navržena řada postupů. Stále využívaný přístup, navržený již v roce 1988, je detekce globálního minima ve tvarovém profilu chromozomu [27]. Tato metoda však dává spolehlivé výsledky pouze v případě, že chromozomy nejsou silně ohnuty. Pokud jsou chromozomy ohnuté, není často možné v jasovém profilu detekovat žádné významné minimum, viz Obrázek 10.



**Obrázek 10: Tvarový profil chromozomu, detekce centromery**

a) jasový profil rovného chromozomu s výrazným minimem. b) jasový profil zakřiveného chromozomu bez zřejmého minima. Převzato a upraveno z [14].

Pro detekci dobře patrné centromery na chromozomech jejichž ohnutí od svislé osy není větší než 90% lze aplikovat jednoduchý algoritmus, který pro detekci centromery vytváří projekce ve směru řádků a sloupců binárního obrazu, viz Obrázek 11 [33].



**Obrázek 11: Horizontální projekce s minimem naznačujícím pozici centromery.**

Nalezená centromera je v obraze chromozomu označena bílou linií. Převzato z [13].

V tomto případě je projekcí myšlena suma hodnot intenzit pixelů v jednotlivých řádcích (horizontální projekce  $V_h$ ) či sloupcích (vertikální projekce  $V_v$ ). Protože hodnoty intenzit pixelů v binárním obraze jsou pouze 1 a 0, hodnoty  $V_h$  a  $V_v$  jsou dány počtem pixelů v příslušném řádku resp. sloupci. [13].

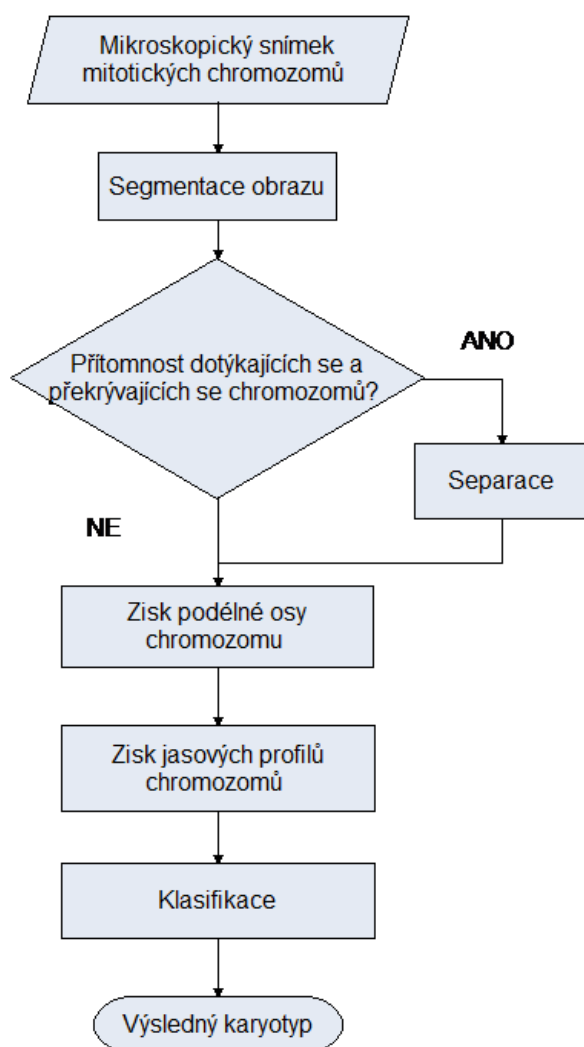
$$V_h(x) = \sum_{y=1}^Y f(x, y), \quad (5)$$

$$V_v(y) = \sum_{x=1}^X f(x, y). \quad (6)$$

## 4 Realizace programu pro klasifikaci chromozomů

Návrh vlastního algoritmu je založen na metodách zpracování obrazových snímků mitotických chromozomů, které jsou popsány v předchozích kapitolách. Algoritmus je realizován v prostředí výpočetního programu MATLAB R2012b. Výhodou MATLABu je dispozice knihovny funkcí pro zpracování obrazu, tzv. Image Processing Toolbox.

Vlastní algoritmus byl testován na mikroskopických snímcích roztlaku metafázních chromozomů. Protože na chromozomech v analyzovaných snímcích není patrná pozice centromery, jsou pro klasifikaci zvoleny následující znaky: délka, obsah a jasový profil chromozomu. Souhrnný průběh algoritmu demonstruje následující vývojový diagram.



Obrázek 12: Vývojový diagram navrženého algoritmu

## 4.1 Segmentace obrazu

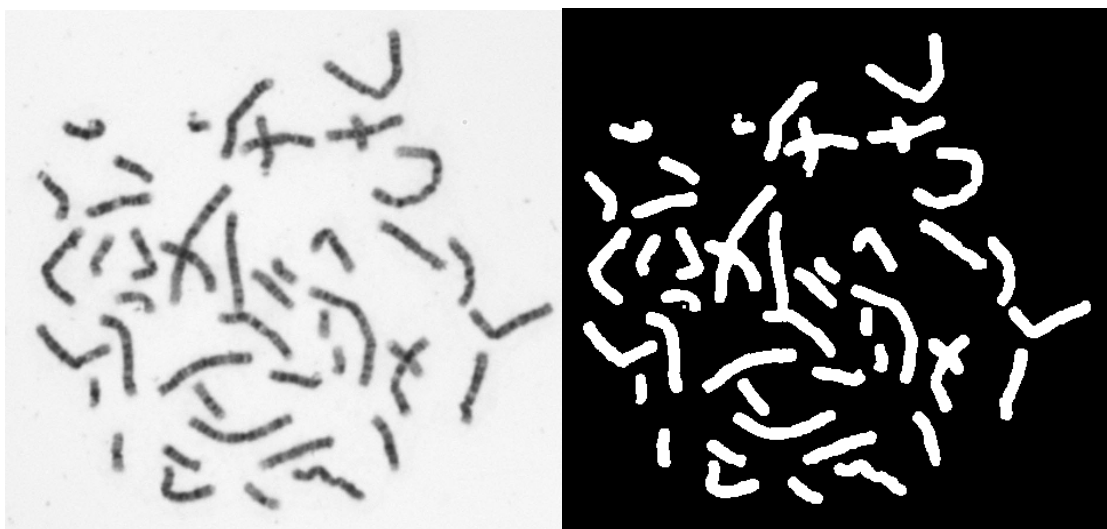
Vstupem algoritmu je šedotónový obraz, v němž je před aplikací dalších metod nezbytné rozlišit samotné chromozomy a oddělit je od pozadí. V této práci jsou snímky segmentovány pomocí hysterezního prahování. Práh je vypočten automaticky pomocí Otsuho prahovací metody [36].

Metoda Otsu vychází z analýzy histogramu, přesněji z relativní četnosti jednotlivých intenzit. Optimální práh je hledán iterativně pomocí rozptylu. V každé iteraci je počítán rozptyl uvnitř a mezi dvěma třídami, vzniklými rozdělením histogramu prahem  $t$ . Optimální práh je takové  $t$ , pro které platí, že je při něm hodnota rozptylu mezi třídami maximální (což odpovídá minimální hodnotě rozptylu uvnitř tříd).

Získaný práh je využit pro hysterezní prahování. Hysterezní prahování využívá dvou prahů  $T_{Low}$  a  $T_{High}$ , které jsou 10% pod a nad Otsuho prahem [24]. Pro převod šedotónového obrazu  $A$  na binární obraz  $B$  platí:

- pokud  $A(i, j) < T_{Low}$ , pak  $B(i, j) = 1$ ,
- pokud  $A(i, j) > T_{High}$ , pak  $B(i, j) = 0$ .

V případě, že se hodnota  $A(i, j)$  nachází mezi prahy  $T_{Low}$  a  $T_{High}$ , pak je příslušné hodnotě  $B(i, j)$  přiřazena hodnota 0 pouze tehdy, když se ve čtyřokolí  $A(i, j)$  nachází hodnota 0. Obrázek 13 ilustruje výsledek hysterezního prahování vstupního snímku.



Obrázek 13: Šedotónová a binární verze téhož obrazu mitotických chromozomů

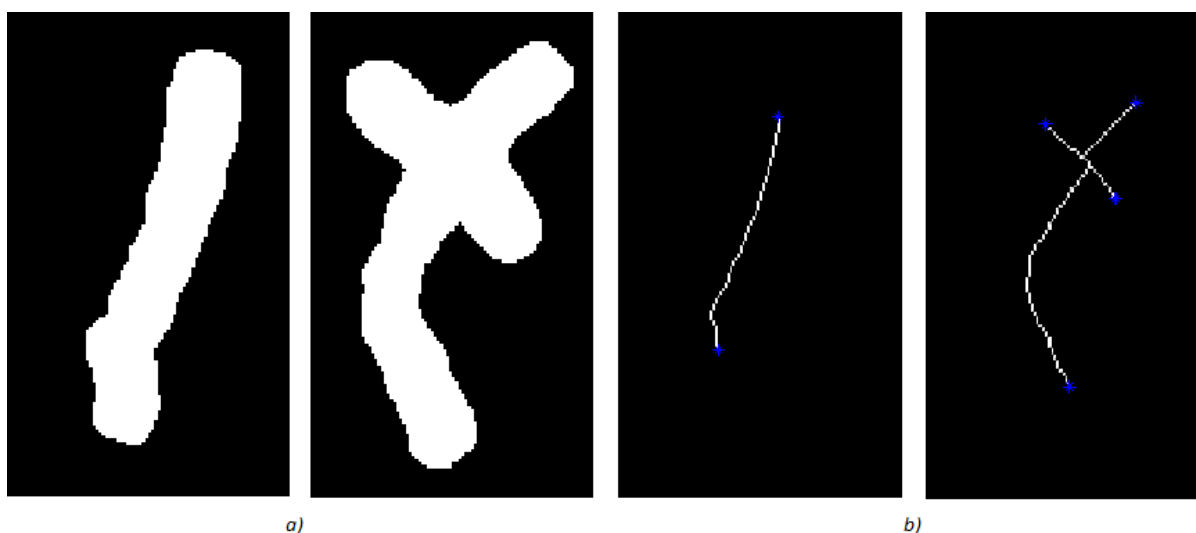
## 4.2 Separace překrývajících se chromozomů

Současné počítačové systémy pro automatickou tvorbu karyotypu jsou interaktivní, a to zejména z důvodu přítomnosti dotýkajících a překrývajících se chromozomů [37].



Prozatimní vědecká činnost byla věnována minimalizaci lidského zásahu, avšak kompletně automatický systém je zatím otázkou budoucnosti [38].

Detekce překrývajících se klastrů byla ve vlastním algoritmu navržena dvěma způsoby. První přístup využívá skeletonizace binárního obrazu a nalezení koncových bodů získaných skeletonů, viz Obrázek 14. V případě, že počet tzv. end-pointů je větší než dva, lze předpokládat, že daný objekt může být klastrem překrývajících se či dotýkajících se chromozomů.



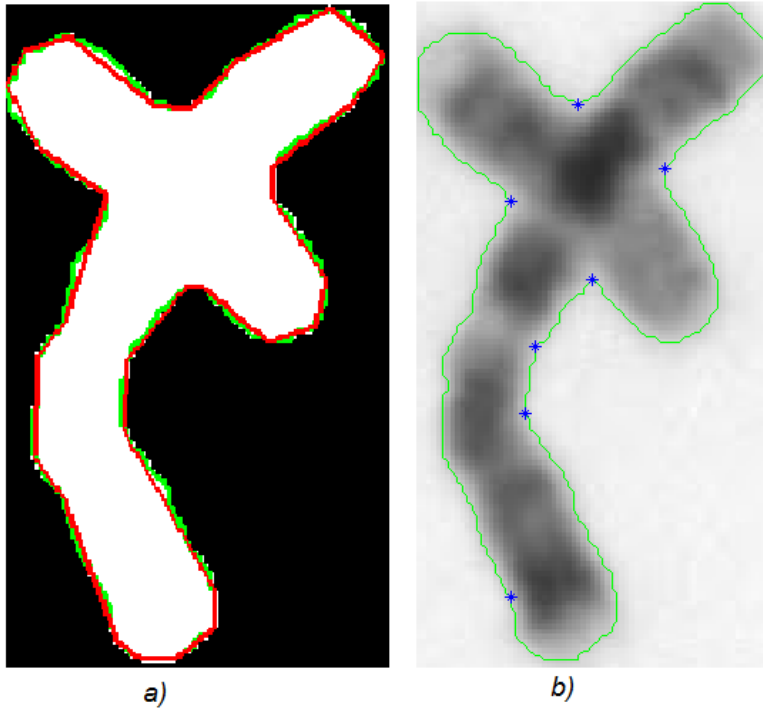
**Obrázek 14: Jednotlivé chromozomy, binární a skeletonizované verze**

a) binární verze chromozomů, b) skeletonizace, včetně naznačených koncových bodů

Přístup k této automatické detekci klastrů však komplikuje přítomnost tzv. zbytečných větví (*spurious branches*) [39]. Přesto, že se v algoritmu vyskytuje krok pro odstranění těchto větví, není vyloučeno, že některé klastry budou chybně detekovány. Aby byla při detekci zajištěna dostatečná přesnost, disponuje algoritmus druhým přístupem pro detekci klastrů – manuálním výběrem pomocí tlačítka myši.

Po označení jsou vybrané klastry uloženy do buňkového pole, které umožňuje snadný přístup k datům a jejich další analýzu. Protože algoritmus téměř výhradně pracuje s jednotlivými chromozomy, je těchto tzv. buňkových polí (*cell arrays*) často využíváno pro ukládání dat různých typů.

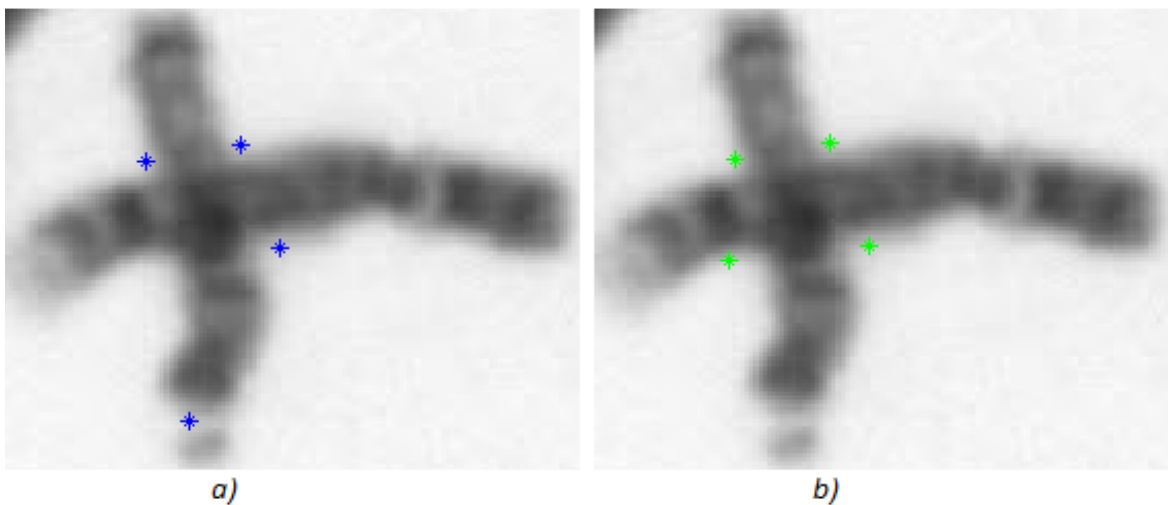
Samotná separace využívá aproximace hranice chromozomu polygonem [40]. K získání nových vrcholů polygonu je použit Douglas-Peuckerův algoritmus [41]. Následně jsou detekovány ty vrcholy polygonu, jejichž úhel je konkávní. Tyto vrcholy jsou považovány za potenciální body, mezi kterými mohou vést dělicí linie. Aproximace polygonem a detekce konkávních vrcholů je ilustrována na Obrázek 15.



**Obrázek 15: Aproximace hranice chromozomu polygonem**

a) hranice chromozomu (zeleně) a její aproximace polygonem (červeně), b) vyznačení konkávních úhlů na polygonu

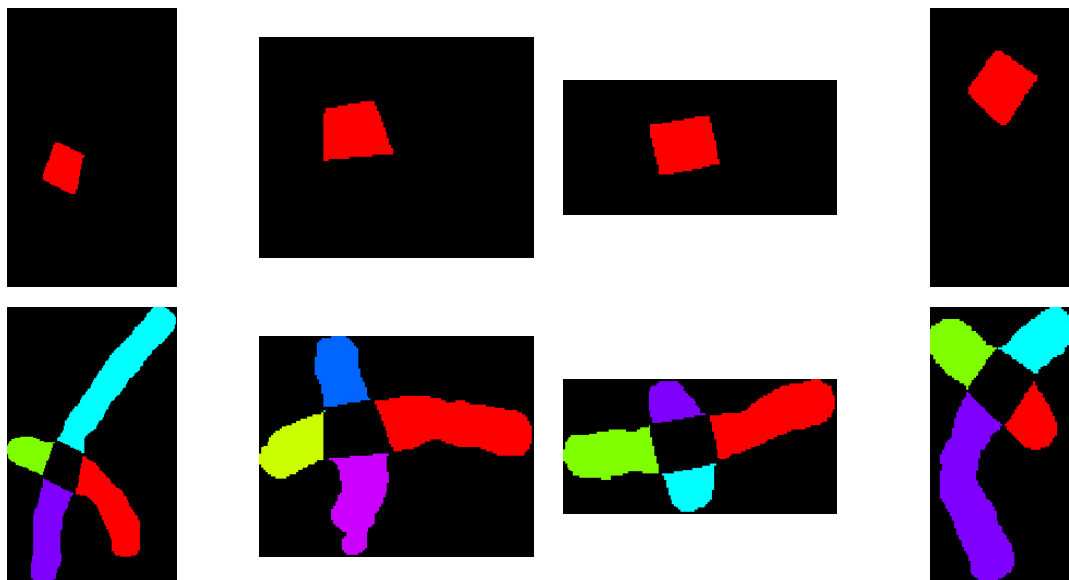
Z Obrázek 15 lze též odhadnout, že body, které algoritmus hledá, mají největší hodnotu tupého úhlu. Tyto body jsou vrcholem čtyřúhelníkové oblasti, společné pro oba klastry. Přes tvarovou rozmanitost chromozomů je automatická detekce těchto bodů většinou účinná. V případech kdy není (Obrázek 16a), je opět vyžadován zásah lidského operátora pro případnou korekci (Obrázek 16b).



**Obrázek 16: Označení vrcholů čtyřúhelníkové oblasti společné pro oba klastry**

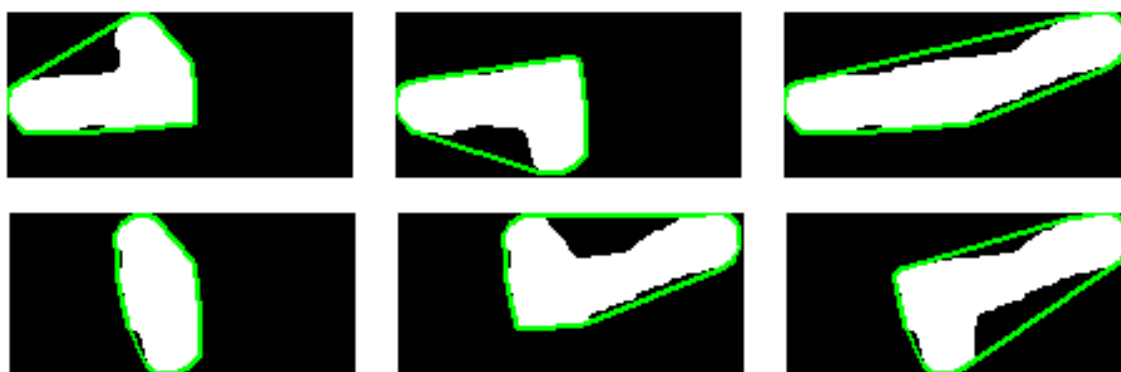
a) chybné označení vrcholů čtyřúhelníka, b) správné označení vrcholů čtyřúhelníka

Jsou-li vrcholy uznány za správné, je možné jimi definovat čtyřúhelníkovou oblast, společnou pro oba klastry. Je-li tato oblast v binární verzi snímku nastavena na hodnotu pozadí, ramena jsou separována a je možno s nimi nakládat jako se samostatnými objekty. Chromozomální klastř je rozdělen na 5 částí – středovou část a čtyři ramena, Obrázek 17.



Obrázek 17: Čtyři označená ramena chromozomu a překrývající se část

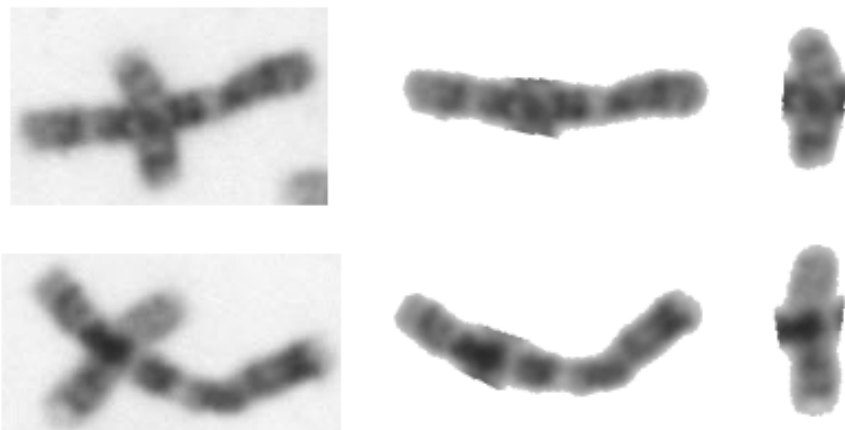
V každém klastřu existuje 6 kombinací ramen, které mohou vytvořit výsledný chromozom, jak ukazuje Obrázek 18. Pro každou z šesti kombinací byly vypočteny dva znaky, na základě kterých je možné z šestice vybrat vhodnou kombinaci. Těmito znaky jsou poměr obsahu chromozomu a obsahu konvexní obálky a dále velikost negativních úhlů.



Obrázek 18: Kombinace ramen a středové společné části pro jeden chromozomální klastř

Zeleně vyznačení konvexní obálky

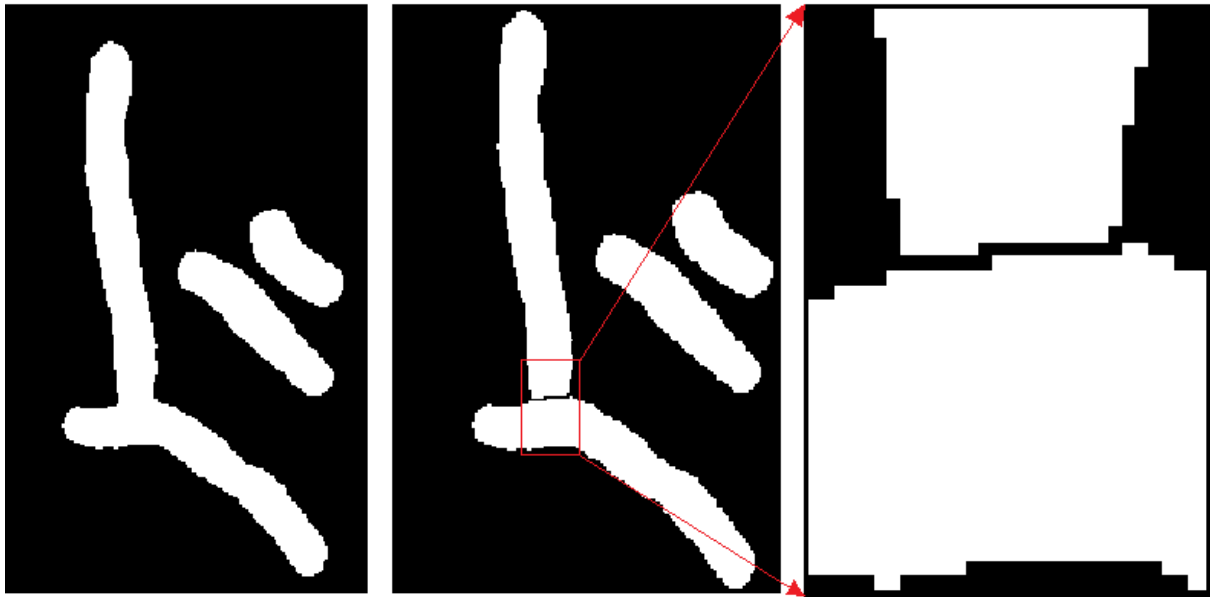
Poměr obsahu chromozomu a obsahu konvexní obálky nedává vždy dobré výsledky. Pomocí přítomnosti negativních úhlů byly detekovány správné kombinace ve všech případech za předpokladu, že byla hranice chromozomu dostatečně vyhlazena. Následující obrázek ukazuje výsledek separace čtyř překrývajících se chromozomů.



Obrázek 19: Výsledek separace překrývajících se chromozomů

### 4.3 *Separace dotýkajících se chromozomů*

Problém dotýkajících se chromozomů je v této práci řešen manuálním výběrem bodů, mezi kterými je vytvořena separační linie. Body jsou vybírány kliknutím tlačítka myši, jejich souřadnice jsou uloženy do matice, přičemž nezáleží na pořadí výběru. Následně jsou vybírány pouze souřadnice patřící bodům, mezi kterými se dotýkající se chromozomy mají rozdělit. Při hledání souřadnic pixelů dělicí linie je zohledněna  $x$ -ová a  $y$ -ová vzdálenost koncových bodů. Pro větší ze vzdáleností se pixely doplňují od jednoho bodu ke druhému s krokem 1, pro menší vzdálenost je volen krok menší než 1. Jeho velikost je dána poměrem menší ku větší vzdálenosti. V posledním kroku jsou nově získané souřadnice nastaveny na hodnotu pozadí. Tím jsou chromozomy programem považovány za dva samostatné objekty, jak lze vidět na Obrázek 20.



**Obrázek 20: Manuální separace dotýkajících se chromozomů**

Levá část obrázku ilustruje dotýkající se chromozomy před rozdělením. Body, mezi kterými je vytvořena linie z černých pixelů, jsou vybrány tlačítkem myši. Pravá část zobrazuje zvětšené místo separace.

Další zpracování se týká jednotlivých chromozomů, proto jsou jako jednotlivé objekty z celkového obrazu vyřiznuty a uloženy jak v binární tak v šedotónové verzi. Pomocí implementované funkce v Matlabu je zjištěna orientace hlavní osy chromozomu k ose  $x$ . Následně je každý chromozom rotován do vertikální polohy.

#### **4.4 Vyhodnocení lidského karyotypu**

Chromozomy je možné uspořádat do karyotypu pomocí identifikačních znaků. V této práci jsou využity znaky jako je délka, obsah a jasový profil chromozomu. Klasifikace je provedena na základě korelací jasových profilů. Délka a obsah slouží k pomocným výpočtům. Chromozomy na analyzovaných snímcích nemají patrnou centromeru a proto nemá význam využívat délku jako klasifikační znak.

Samotná klasifikace je provedena pomocí shlukové analýzy. Vstupem klasifikačního algoritmu je distanční matice RMSD a korelačních koeficientů jasových profilů. Na základě distanční matice je metodou hierarchického shlukování vytvořen dendrogram, ze kterého je možné odečíst správně klasifikované chromozomy.

#### **4.5 Získ specifických znaků chromozomů**

##### **Detekce podélné osy**

Základem pro získ specifických znaků je detekce podélné osy. V navrženém algoritmu je podélná osa získána metodou ztenčování binárního obrazu chromozomu

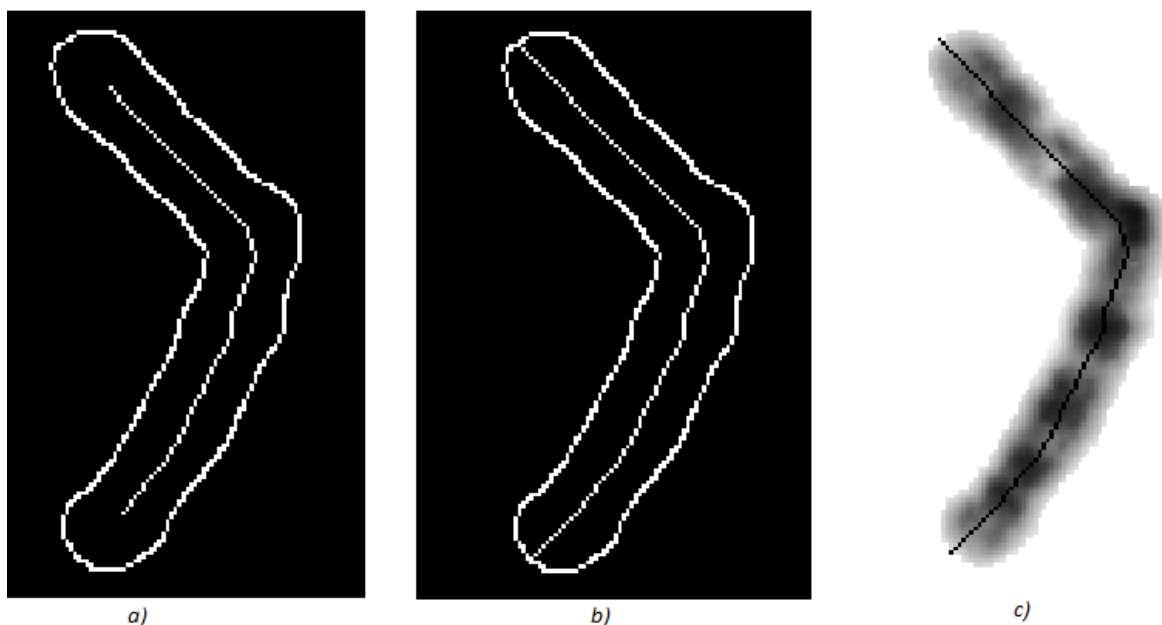
pomocí implementované funkce *bwmorph* s argumentem *thin*. V důsledku výpočetního algoritmu je délka této osy kratší než je samotná délka chromozomu, jak lze vidět na Obrázek 21 a), proto je zapotřebí ji prodloužit ke koncům chromozomu.

Celý skeleton je dopočítán pomocí části algoritmu, který vyhledá aktuální koncové body a vypočítá sklon přímky, aproximující poslední tři pixely skeletonu. Na základě znalosti sklonu přímky a bodu, kterým tato přímka prochází, lze nalézt další body přímky v určité euklidovské vzdálenosti od výchozího bodu. Jsou-li souřadnice výchozího bodu  $x_1$ ,  $y_1$ , sklon přímky  $m$  a vzdálenost od tohoto bodu  $d$ , pak souřadnice hledaného bodu  $x_A$ ,  $y_A$  jsou dány následujícími rovnicemi:

$$x_A = x_1 + \frac{d}{\sqrt{1 + m^2}} \quad (7)$$

$$y_A = y_1 + \frac{d \cdot m}{\sqrt{1 + m^2}} \quad (8)$$

Algoritmus iterativně počítá body  $x_A$ ,  $y_A$  s rostoucí vzdáleností  $d$  od koncového bodu skeletonu. Ve chvíli, kdy je dosaženo hranice chromozomu je cyklus ukončen. Všechny získané body jsou nastaveny na hodnotu 1. Ve výsledku je vytvořena podélná osa chromozomu od jednoho konce chromozomu ke druhému, jak ilustruje Obrázek 21 b) a c).



**Obrázek 21: Hranová reprezentace chromozomu s naznačenou podélnou osou**

a) hranová reprezentace chromozomu včetně topologické kostry (aproximace podélné osy), získané procesem „ztenčování“, b) tentýž chromozom s prodlouženou osou, c) zobrazení skeletonu přes šedotónovou verzi chromozomu.

## Obsah a délka chromozomů

Obsah chromozomu odpovídá počtu bílých pixelů v binárním obraze chromozomu. Délka chromozomu je zjištěna jako počet pixelů skeletonu po jeho prodloužení k okrajům chromozomu, obrázek 22.



**Obrázek 22: Šedotónová verze, binární verze a topologická kostra chromozomu**

Obsah chromozomu odpovídá počtu bílých pixelů v binární verzi chromozomu. Délku chromozomu reprezentuje počet bílých pixelů jeho topologické kostry. Obsah chromozomu vlevo: 4011 pixelů, délka: 176 pixelů. Obsah chromozomu vpravo: 1373 pixelů, délka: 70 pixelů

V následujících tabulkách jsou uvedeny sestupně seřazené hodnoty obsahů a délek pro jednu skupinu 46 chromozomů.

**Tabulka 4: Příklady obsahů jednotlivých chromozomů**

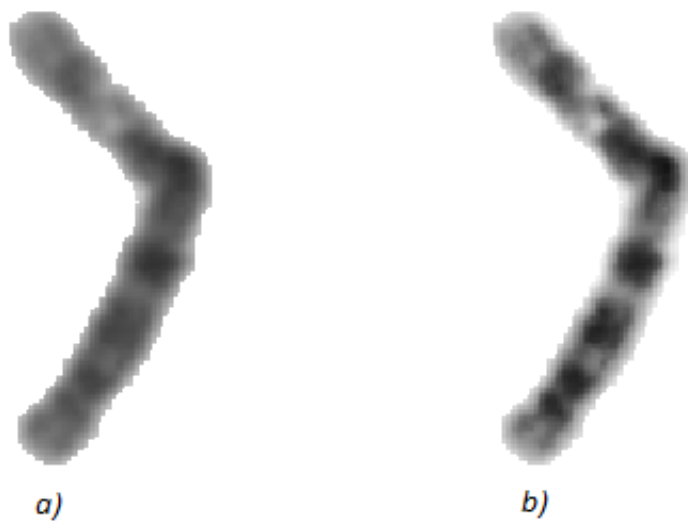
3400	2073	3668	2053	2555	1569	936	3398	1380	1617
1574	2373	3874	2800	1544	968	4011	1594	3078	3673
3532	1373	1991	2859	1984	1221	2853	4636	2511	966
4813	981	4794	2031	2922	2594	2751	3622	3119	4990
2271	3013	3041	1553	3269	1940				

**Tabulka 5: Příklady délek jednotlivých chromozomů**

226	208	201	188	176	173	170	162	157	153
152	145	140	136	136	134	132	130	129	127
123	120	120	118	110	109	105	100	95	93
93	79	78	77	76	74	74	73	70	69
65	56	52	50	49	49				

## Získ jasových profilů

Jasové profily všech chromozomů je možné získat sestrojením kolmic na podélnou osu [42]. Tato konstrukce je, až na extrémní případy, nezávislá na stočení chromozomu. Pro zvýšení přesnosti při vytváření jasového profilu jsou jednotlivé pruhy zvýrazněny pomocí adaptivní ekvalizace histogramu, Obrázek 23 [43].



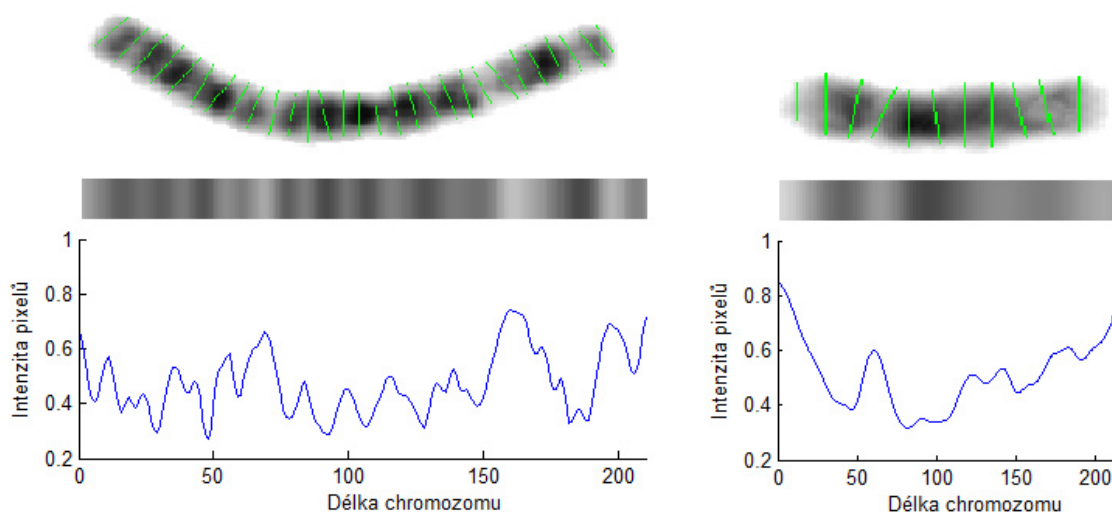
**Obrázek 23: Zvýšení kontrastu šedotónového obrazu chromozomu**

a) originální verze, b) verze se zvýšeným kontrastem

Jednotlivé kolmice jsou vytvořeny na základě lokálního sklonu podélné osy [11]. Algoritmus prochází pixely podélné osy od jednoho konce k druhému. Pro každý pixel je zjištěn lokální sklon z předchozího a následujícího pixelu – aproximační přímka. Konstrukce kolmic vychází ze skutečnosti, že jsou-li dvě přímky kolmé, pak sklon jedné z nich je negativní reciprokou hodnotou sklonu druhé. Znalost sklonu kolmice a bodu, kterým prochází (středový bod aproximační přímky), umožňuje iterativně získávat souřadnice kolmice na obě strany od příslušné aproximační přímky směrem k hranici chromozomu.

Pomocí sestavených kolmic lze získat jasové profily, které reprezentují sled pruhů na chromozomu. Hodnoty jasového profilu jsou získány jako medián intenzit pixelů podél všech přímek na podélnou osu. Každý jasový profil je před dalším použitím interpolován kubickým splinem. Výsledkem je vektor hodnot, který reprezentuje rozložení intenzity jasu podél celého chromozomu.





**Obrázek 24: Jasový profil chromozomu**

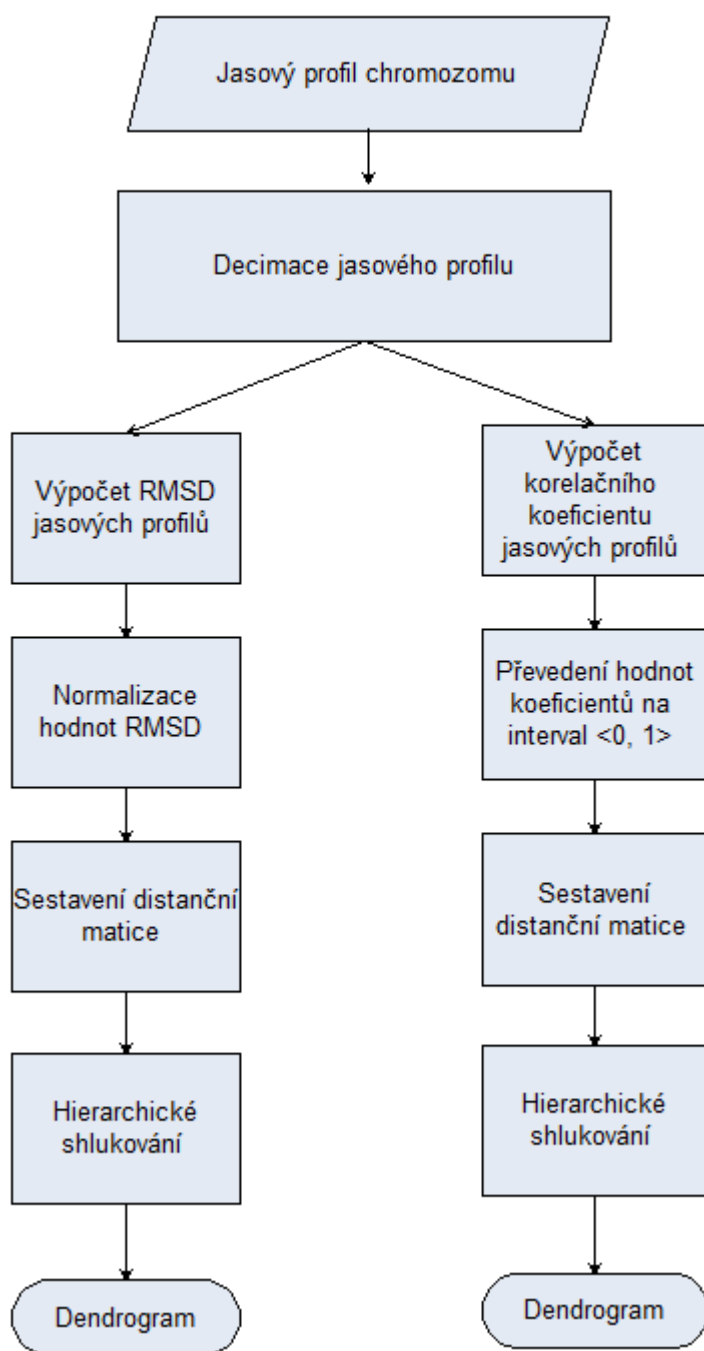
a) dolní část obrázku zobrazuje jasový profil chromozomu interpolovaný kubickým splinem a normalizovaný na délku od nuly do jedné. Ve středu je zobrazen medián intenzit jasu, vypočítaný z hodnot získaných podél kolmic na podélnou osu, které jsou naznačena na chromozomu v horní části obrázku

## 4.6 Klasifikace

Pro klasifikaci chromozomů byla zvolena shluková analýza. Shluková analýza (cluster analysis) třídí data do shluků (klastřů) tak, aby si v rámci jednoho shluku byla podobnější než data z různých shluků [44]. Klasifikace dat do shluků je založena na znacích, kterými jsou data popsána. Jako klasifikačních znaků bylo využito délky, obsahu a jasového profilu chromozomů, jejichž zisk je popsán v předchozí kapitole.

Pro vyjádření míry podobnosti (tzv. vzdálenosti) znaků se využívají tzv. metriky [45]. Metrik existuje celá řada a jejich volba ovlivňuje výsledné zařazení objektu do určitého shluku. V této práci je podobnost obsahu a délky chromozomů porovnána pomocí výpočtu RMSD, v případě jasového profilu je ještě využit korelační koeficient. Vypočtenými hodnotami vzdáleností je naplněna distanční matice, která vstupuje do shlukové analýzy. Grafickým vyjádřením výsledku shlukování je dendrogram.

Při klasifikaci chromozomů se předpokládá, že klasifikační znaky homologního páru jsou si podobnější než znaky chromozomů nehomologních. Tento předpoklad komplikuje skutečnost, že znaky v rámci homologního páru nejsou identické, naopak, mnohdy lze nalézt větší podobnost např. mezi nehomologními chromozomy blízké délky [46]. Následující vývojový diagram ilustruje průběh klasifikačního algoritmu.



Obrázek 25: Vývojový diagram klasifikačního algoritmu

### Výpočet RMSD a korelačního koeficientu

Míra podobnosti jednotlivých znaků chromozomů je hodnocena na základě RMSD (root mean square deviation) koeficientu. Čím je hodnota koeficientu bližší nule, tím je podobnost větší.

$$RMSD = \sqrt{\frac{\sum_{i=1}^n (x_{1,i} - x_{2,i})^2}{n}} \quad (9)$$

Obsah a délka chromozomu představují vektory hodnot, pro které lze bez dalších úprav RMSD spočítat. V případě výpočtu RMSD jasových profilů je nejprve potřeba sjednotit jejich různou délku, danou různou délkou chromozomů. K tomuto účelu je využito podvzorkování (decimace) profilů s celočíselným faktorem  $D$ . Faktor  $D$  je získán jako poměr aktuálního a nejkratšího jasového profilu. Spektrum profilu je filtrováno antialiasingovým filtrem, aby vynechání některých hodnot nevedlo k výrazné změně průběhu profilu. Následně je vybrán každý  $D$ -tý vzorek, čímž je sjednocena délka všech jasových profilů.

Protože na snímcích není patrná poloha centromery, není v této části algoritmu známa orientace chromozomu. Z toho důvodu je nutné vypočítat korelace pro dvě skupiny jasových profilů. Algoristmus iterativně vybírá vždy jeden jasový profil, spočítá korelace s ostatními, pak tento profil otočí a znovu spočítá všechny korelace. Ve výsledku vznikají dvě distanční matice, které jsou procházeny pozice po pozici a je určeno, která z hodnot je menší. Tato hodnota je pak překopírována do třetí, výsledné matice.

Podobnost jasových profilů je kromě pomocí RMSD hodnocena také pomocí Pearsonova korelačního koeficientu, rovnice 10. Korelační koeficient udává lineární závislost dvou statistických veličin a nabývá hodnot v intervalu  $(-1, 1)$ . Hodnoty  $\pm 1$  nabývá koeficient v případě, že mezi veličinami je absolutní závislost. Kladné hodnoty reprezentují pozitivní závislost (obě veličiny stoupají), záporné negativní (jedna veličina stoupá, druhá klesá). Pokud mezi veličinami neexistuje lineární závislost, nabývá koeficient hodnoty 0.

$$r = \frac{\sum_{i=1}^n (x_{1,i} - \bar{x}_1)(x_{2,i} - \bar{x}_2)}{\sqrt{(\sum_{i=1}^n (x_{1,i} - \bar{x}_1)^2 \sum_{i=1}^n (x_{2,i} - \bar{x}_2)^2)}} \quad (10)$$

### Distanční matice

Jak již bylo zmíněno, vypočtené hodnoty RMSD a korelačního koeficientu naplňují distanční matice. Část distanční matice pro RMSD a korelační koeficient ilustruje Tabulka

6 a Tabulka 7. Protože však distanční matice nemohou vytvářet záporné hodnoty, jsou záporné hodnoty korelačního koeficientu převedeny na kladné přičtením hodnoty 1 a dělením hodnotou 2. Je zohledněno, že míra korelace je při použití RMSD opačná než při použití korelačního koeficientu. Korelace je maximální při  $RMSD = 0$  a při  $r = 1$ . V dendrogramu platí, že čím více se vzdálenost blíží nule, tím je vztah dvou objektů silnější. Proto je míra podobnosti mezi dvěma objekty vyjádřena jako  $r_p = 1 - r$ .

Hodnoty délky a obsahu jsou po výpočtu distanční matice normalizovány na interval  $\langle 0, 1 \rangle$  pomocí min-max normalizace, jak ilustruje rovnice 9 [47]. Všechny znaky tak ovlivňují klasifikaci se stejnou vahou.

$$x'_i = \frac{x - \min(x_i)}{\max(x_i) - \min(x_i)} \quad (11)$$

**Tabulka 6: Část distanční matice pro RMSD jasových profilů**

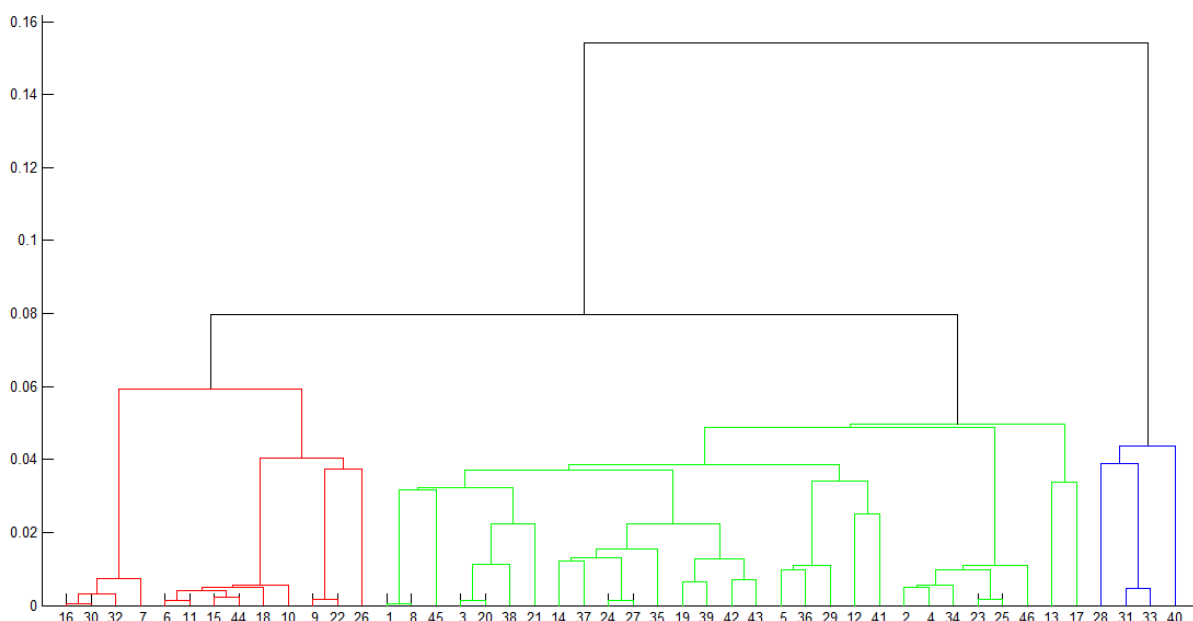
chr. č.	1	2	3	4	5	6	7	8	9	10
1	0	0,171	0,122	0,105	0,151	0,121	0,142	0,130	0,118	0,126
2	0,171	0	0,146	0,165	0,203	0,176	0,098	0,154	0,154	0,161
3	0,122	0,146	0	0,113	0,177	0,132	0,127	0,115	0,101	0,127
4	0,105	0,165	0,113	0	0,188	0,158	0,138	0,109	0,113	0,121
5	0,151	0,203	0,177	0,188	0	0,123	0,192	0,184	0,194	0,171
6	0,121	0,176	0,132	0,158	0,123	0	0,158	0,165	0,130	0,183
7	0,142	0,098	0,127	0,138	0,192	0,158	0	0,142	0,112	0,125
8	0,130	0,154	0,115	0,109	0,184	0,165	0,142	0	0,121	0,114
9	0,118	0,154	0,101	0,113	0,194	0,130	0,112	0,121	0	0,138
10	0,126	0,161	0,127	0,121	0,171	0,183	0,125	0,114	0,138	0

**Tabulka 7: Část distanční matice pro korelační koeficient jasových profilů**

chr. č.	1	2	3	4	5	6	7	8	9	10
1	0	0,654	0,429	0,341	0,331	0,342	0,601	0,414	0,384	0,380
2	0,654	0	0,441	0,565	0,506	0,606	0,203	0,397	0,511	0,403
3	0,429	0,441	0	0,336	0,449	0,403	0,448	0,269	0,266	0,310
4	0,341	0,565	0,336	0	0,556	0,560	0,491	0,288	0,297	0,371
5	0,331	0,506	0,449	0,556	0	0,139	0,528	0,465	0,533	0,407
6	0,342	0,606	0,403	0,560	0,139	0	0,666	0,494	0,425	0,599
7	0,601	0,203	0,448	0,491	0,528	0,666	0	0,395	0,381	0,205
8	0,414	0,397	0,269	0,288	0,465	0,494	0,395	0	0,264	0,279
9	0,384	0,511	0,266	0,297	0,533	0,425	0,381	0,264	0	0,315
10	0,380	0,403	0,310	0,371	0,407	0,599	0,205	0,279	0,315	0

## Dendrogram

Dendrogram je grafický výstup hierarchického shlukování. Při vertikální orientaci dendrogramu je na pomyslené ose y uvedena vzdálenost mezi shluky. Čím je tato vzdálenost menší, tím více jsou si jednotlivé shluky podobné. V této práci je dendrogram sestaven na základě distanční matice pomocí implementované funkce *linkage*. Pro sestavení dendrogramu byla zvolena metoda UPGMA [48]. Na Obrázek 26 lze vidět dendrogram pro RMSD obsahu chromozomů. Dendrogramy pro RMSD akorelační koeficient jasových profilů jsou součástí příloh.



Obrázek 26: Dendrogram (RMSD obsahu chromozomu)

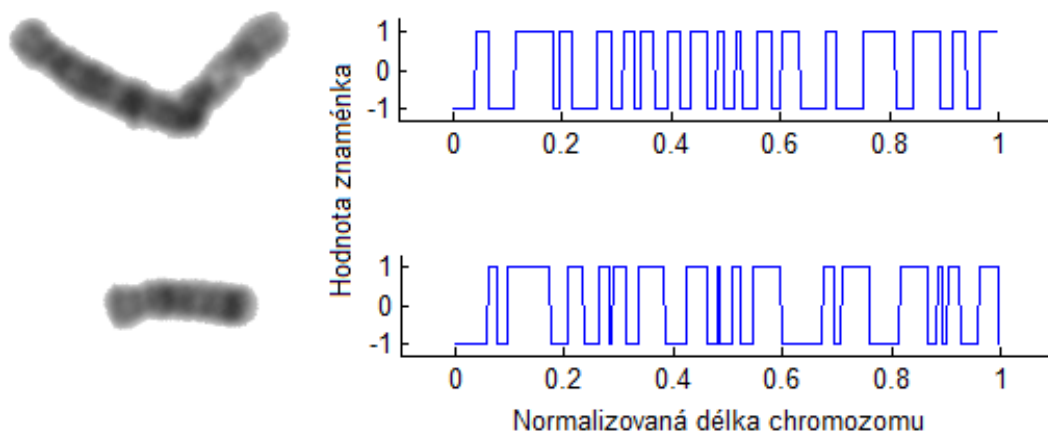
## Srovnání s ideogramy

Jak již bylo zmiňováno v předchozích kapitolách, počet pruhů na chromozomu závisí na stupni buněčného dělení. Protože není známo z jaké fáze dělení analyzované snímky pochází, byl navržen algoritmus, který odhaduje “band resolution” snímku [11].

Odhad počtu pruhů vychází z druhé derivace jasových profilů interpolovaných kubickým splinem. Je založen na předpokladu, že kde se v jasovém profilu mění průběh funkce z konvexní na konkávní a naopak, tam jsou hranice pruhů na chromozomu. Protože platí, že kde je funkce konvexní, tam je druhá derivace kladná a kde konkávní, tam je druhá derivace záporná, byl odhad počtu pruhů na základě hodnot druhé derivace stanoven následujícím způsobem:

$$\begin{aligned} Y &= 1 \quad \text{kde } d'' > 0 \\ Y &= 0 \quad \text{kde } d'' = 0 \\ Y &= -1 \quad \text{kde } d'' < 0 \end{aligned} \tag{12}$$

V jasovém profilu odpovídají konvexní části funkce tmavým pruhům a konkávní části pruhům světlým. Analogicky, v nově vytvořeném vektoru hodnot  $Y$  odpovídají tmavé pruhy hodnotě 1, světlé hodnotě -1. Vynesení hodnot vektoru  $Y$  do grafu je na obrázek 27.



**Obrázek 27: Změna znaménka druhé derivace jasového profilu**

Počet „vrcholů“ v grafu na obrázku 27 tedy reprezentuje počet pruhů na jednom chromozomu. Počet je zjištěn tak, že algoritmus hledá ve vektoru  $Y$  ty pozice, kde dochází ke změně znaménka. Pro chromozom v horní části obrázku 27 je tato hodnota 33 pro spodní chromozom 32. Součet všech hodnot dává počet pruhů v diploidní sadě chromozomů. Pro analyzovaný případ je tato hodnota 1068. Pro haploidní sadu pak 534, což odpovídá rozlišení chromozomů zachycených v metafázi buněčného dělení.

## 5 Diskuze výsledků

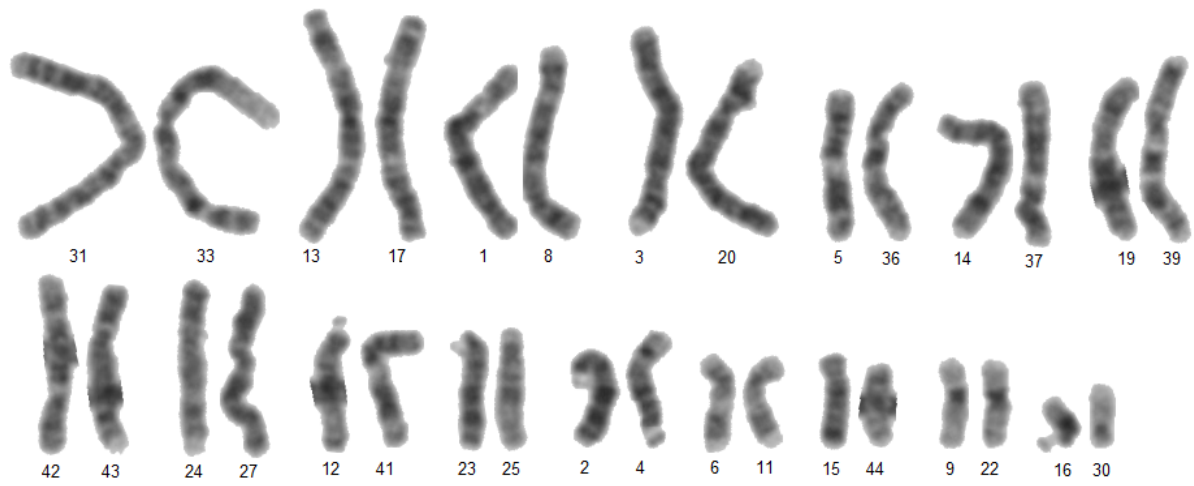
Tato práce je zaměřena na klasifikaci chromozomů pomocí shlukové analýzy. Navržený algoritmus byl testován na mikroskopických snímcích chromozomů bez apriorní informace. Tedy není např. předem známo z jaké fáze mitotického dělení chromozomy pocházejí a tedy není znám tzv. banding resolution (rozlišení pruhů - součet pruhů na všech chromozomech haploidní sady) Z tohoto důvodu je těžké jednoznačně přiřadit takovou standardní referenci (ideogram), která svým rozlišením odpovídá rozlišení analyzovaných chromozomů.

Chromozomy na analyzovaných snímcích nemají patrnou polohu centromery, proto není možné provést klasifikaci do sedmi tříd podle délky a polohy centromery. Též není možné podle centromery určit správnou orientaci chromozomů. Tyto skutečnosti ovlivňují návrh dalšího algoritmu, který zohledňuje absenci některé informace – např. zmíněné orientace chromozomu.

Do klasifikačního algoritmu vstupuje vektor obsahů všech chromozomů a dále cell array s uloženými jasovými profily. Míra korelace obsahů chromozomů je hodnocena pomocí jejich kvadratické odchylky (RMSD). Klasifikace chromozomů podle míry korelace jejich jasových profilů je provedena na základě výpočtu jejich kvadratické odchylky (RMSD) a Pearsonova korelačního koeficientu.

Obrázek 26 ukazuje dendrogram, který vychází z distanční matice obsahů. Dendrogramy jasových profilů jsou součástí příloh této práce. Další text diponuje obrázky, které vznikly odečtením z dendrogramu, aby bylo možno zhodnotit s jakou přesností probíhá klasifikace na základě té které metriky (RMSD/Korelační koeficient).

Pro sestavení distanční matice obsahu chromozomů byla vypočtena kvadratická odchylka jednotlivých obsahů. Ze 23 potenciálních párů, bylo zaklasifikováno 16, z toho 3 představují skutečně homologní pár. Tabulka 8 obsahuje indexy z dendrogramu pro správně zařazené chromozomy. Indexy v tabulce odkazují jak do denrogramu tak do obrázků chromozomových párů.



**Obrázek 28: Chromozomy zařazené do párů dle RMSD obsahu**

Správně zařazené chromozomy jsou vypsány v Tabulka 8 a v tabulce 9

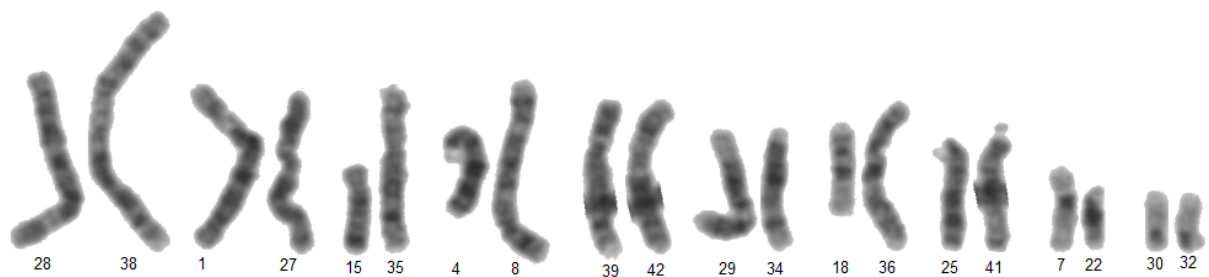


**Tabulka 8: Procentuální úspěšnost při hledání homologních párů**

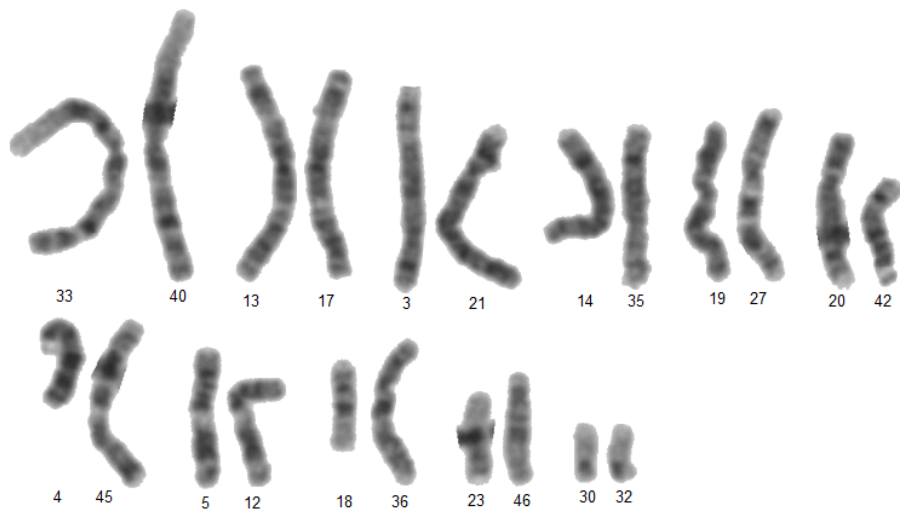
Pomocí RMSD obsahu, RMSD jasového profilu a pomocí korelačního koeficientu

kritérium	RMSD obsahu				RMSD jasového profilu				Korelační koeficient jasového profilu			
	všechny nalezené páry		správně nalezené páry		všechny nalezené páry		správně nalezené páry		všechny nalezené páry		správně nalezené páry	
index chromozomu	31	33	6	11	28	31	30	32	33	40	30	32
	13	17	9	22	1	27			13	17	5	12
	1	8	23	25	15	35			3	21	13	17
	3	20			4	8			14	35	33	40
	5	36			39	42			19	27	14	35
	14	37			29	34			20	42	19	27
	19	39			18	36			4	45		
	42	43			25	41			5	12		
	24	27			7	22			18	36		
	12	41			30	32			23	46		
	23	25							30	32		
	2	4										
	6	11										
	15	44										
	9	22										
	16	30										
počet nalezených párů	16		3		10		1		11		6	
úspěšnost	18,75%				10%				54,54%			





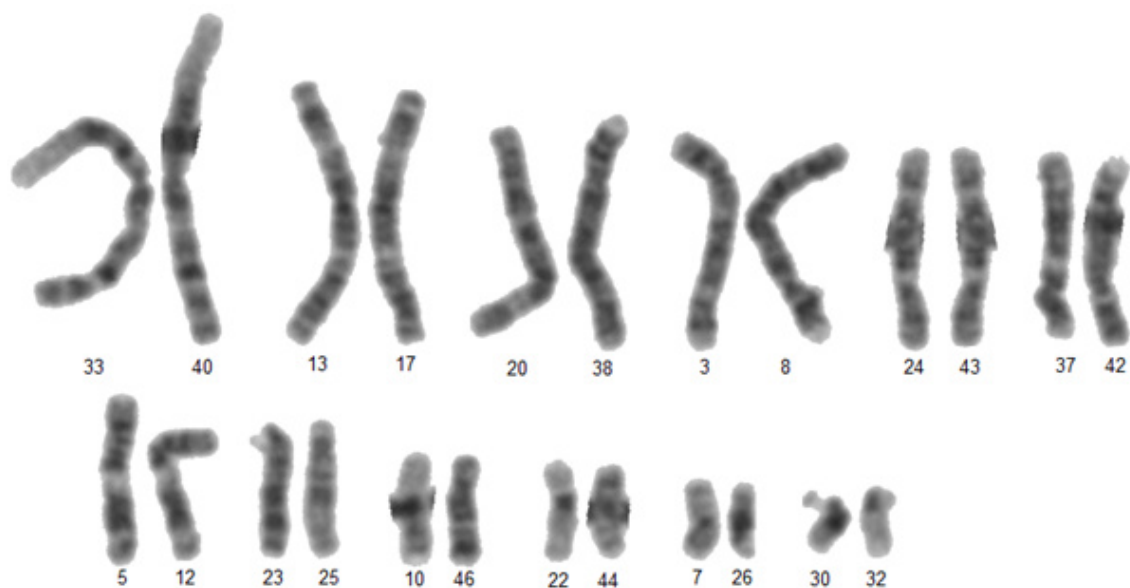
**Obrázek 29: Chromozomy klasifikované do párů pomocí RMSD jasových profilů**



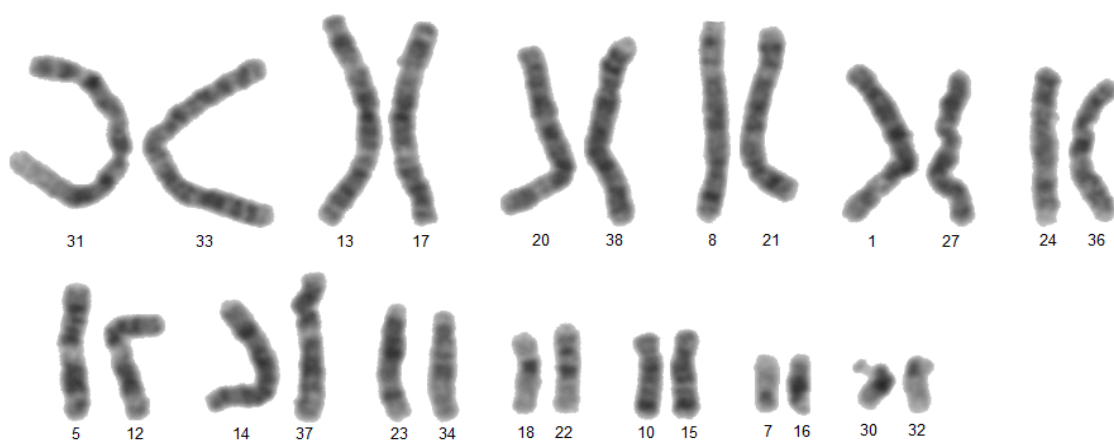
**Obrázek 30: Chromozomy klasifikované do párů pomocí Pearsonova korelačního koeficientu**

Obrázek 29 a Obrázek 30 ukazují, že je často nalezena korelace mezi jasovými profily chromozomů značně odlišné délky. Z tohoto důvodu byl navržen přístup, který omezuje rozsah délek (počet hodnot) jasových profilů v rámci kterého se korelace hledá. Je stanovena procentuální odchylka délky každého chromozomu, která představuje meze intervalu, do kterého musí jasový profil druhého potenciálního chromozomu svojí délkou spadat.

Test korelace je proveden pouze mezi těmi chromozomy, jejichž rozdíl délek není větší než nastavená tolerance. Meze intervalu vhodných délek jsou nastaveny na  $\pm 10\%$  (Obrázek 31) a  $\pm 20\%$  (Obrázek 32) délky jasového profilu (která koresponduje s délkou chromozomu). Chromozomy zařazené do párů pomocí korelace jejich jasových profilů na intervalu omezené délky ukazují obrázky 31 a 32.

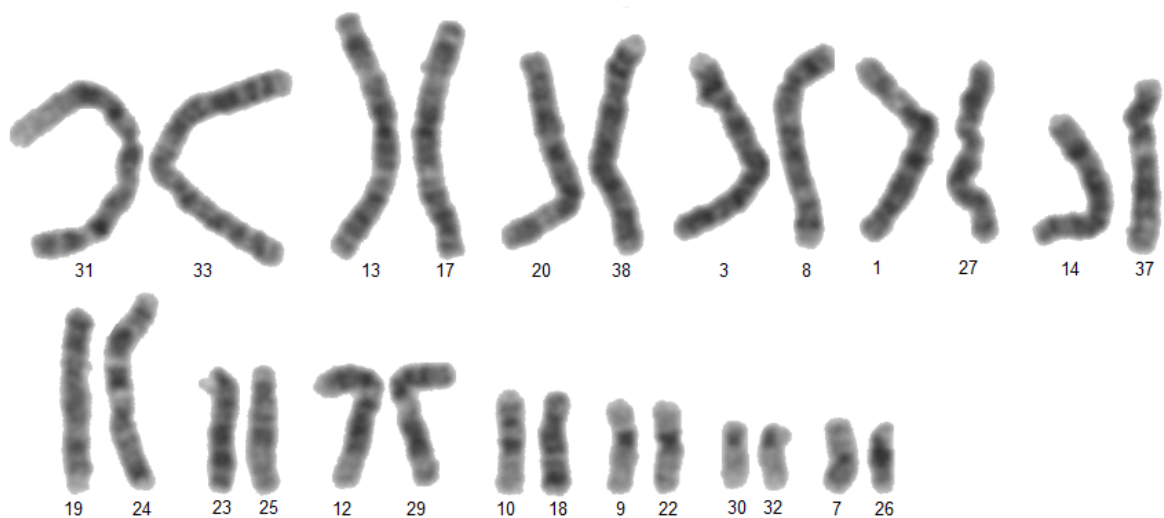


**Obrázek 31: Chromozomy klasifikované do párů pomocí Pearsonova korelačního koeficientu, omezení výběru na 10%.**



**Obrázek 32: Chromozomy klasifikované do párů pomocí Pearsonova korelačního koeficientu, tolerance 20%**

Separované chromozomy mají jasový profil částečně poškozen v důsledku ztráty informace v místě, kde došlo k překryvu. Bylo testováno, zda má tato změna zásadní vliv na přesnost klasifikace. V rámci 10% odchylky byly provedeny dvě klasifikace, jedna se všemi 46 chromozomy, z druhé byly rozseparované chromozomy vyloučeny. Výsledek klasifikace bez separovaných chromozomů je na obrázek 33. Správně zaklasifikované homologní páry ze všech typů klasifikací uvádí Tabulka 8.

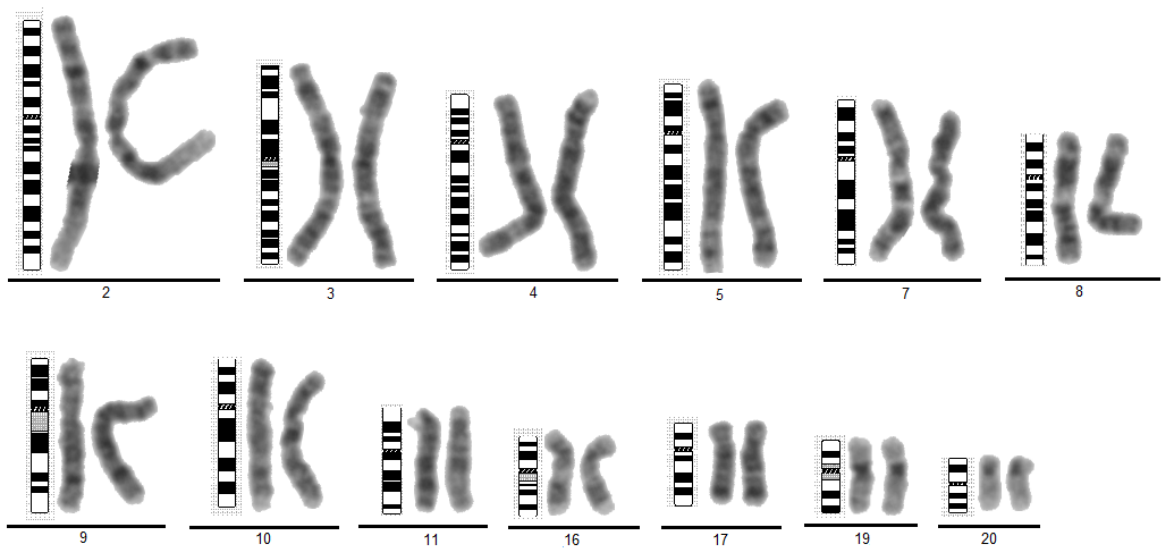


**Obrázek 33: Chromozomy klasifikované do párů pomocí Pearsonova korelačního koeficientu bez separovaných chromozomů, tolerance 10%**

V tabulce 10 jsou uvedeny indexy všech správně zaklasifikovaných párů pomocí kombinací všech použitých metod. Pomocí porovnání těchto párů se standardní referencí byla určena jejich ISCN čísla. Obrázek 34 zobrazuje všechny správně klasifikované homologní páry včetně jejich standardní reference – ideogramu [51]. Uvedená čísla odpovídají číslům chromozomů v karyotypu.

**Tabulka 10: Seznam správně zaklasifikovaných homologních párů**

Číslo chromozomu	Indexy chromozomu	
16	6	11
19	9	22
11	23	25
20	30	32
8	5	12
3	13	17
2	33	40
9	14	35
7	19	27
4	20	38
5	8	21
10	24	36
17	10	15



Obrázek 34: Správně seřazené homologní páry zařazené do příslušných ISCN kategorií

# Závěr

V této práci byly shrnuty metody vizualizace chromozomů sloužící k identifikaci chromozomů z mikroskopických snímků. V základu byly popsány obecné skutečnosti, které se sestavováním karyotypu souvisí, např. popis chromozomů a jejich chování během buněčného cyklu. V práci jsou zmíněna onemocnění, jejichž příčinou jsou změny v počtu či ve struktuře chromozomů a která jsou tedy důvodem k cytogenetickému vyšetření. Nejčastější cytogenetická vyšetření vedou právě k sestavení karyotypu.

Překážkou na cestě k automatickému karyotypování je přítomnost překrývajících se a dotýkajících se chromozomů. Analyzovaný snímek takové chromozomy obsahoval. Byl navržen poloautomatický přístup pro separaci překrývajících se a dotýkajících se chromozomů.

Karyotyp je sestavován na základě znaků chromozomů, které umožňují jejich rozlišení. Část této práce se věnuje identifikaci těchto znaků jak v teoretické, tak praktické části. Identifikačních znaků je celá řada a záleží na uvážení autorů klasifikačních algoritmů, které znaky budou detekovat a využívat k analýze. Tato práce se drží znaků, které jsou k identifikaci využívány nejčastěji, tedy jsou zmiňovány v převážné většině citovaných článků.

Při návrhu vlastního klasifikačního algoritmu bylo využito hlavně jasových profilů chromozomů. Pro zjištění korelací mezi profily byla použita metoda RMSD a Pearsonův korelační koeficient. Klasifikace chromozomů podle míry korelace jejich jasových profilů byla provedena pomocí shlukové analýzy. Různé modifikace základního přístupu za účelem zvýšení přesnosti klasifikace jsou uvedeny v diskuzi.

Identifikace chromozomů, jejich zařazení do homologních párů a sestavení do karyotypu bez počítačové podpory je časově náročné a finančně neefektivní. Na rozvoji automatického karyotypovacího systému se v souladu s rozvojem informačních technologií pracuje už tři desetiletí. Ani nejnovější přístupy však stále neumožňují plně automatickou klasifikaci chromozomů.

# Seznam použitých zdrojů

## Literární zdroje

1. Snustad, D.P. and M.J. Simmons, *Principles of genetics*. 3rd ed2003, New York: Wiley. xix, 840 p.
2. Kannan, T.P. and B.A. Zilfalil, *Cytogenetics: past, present and future*. Malays J Med Sci, 2009. **16**(2): p. 4-9.
3. Moore, C.M. and R.G. Best, *Chromosome Preparation and Banding*, in *eLS2001*, John Wiley & Sons, Ltd.
4. Riegel, M., *Human molecular cytogenetics: From cells to nucleotides*. Genet Mol Biol, 2014. **37**(1 Suppl): p. 194-209.
5. Nussbaum, R.L., et al., *Thompson & Thompson genetics in medicine*. 6th ed2001, Philadelphia: W.B. Saunders. xi, 444 p.
6. Lerner, B., *Toward a completely automatic neural-network-based human chromosome analysis*. IEEE Trans Syst Man Cybern B Cybern, 1998. **28**(4): p. 544-52.
7. Nanni, L., *A reliable method for designing an automatic karyotyping system*. Neurocomputing, 2006. **69**(13–15): p. 1739-1742.
8. Holtzman, N.A. and J.F. Merz, *ISCN 2005 - An International System for Human Cytogenetic Nomenclature (2005) Introduction*. Community Genetics, 2005. **8**(4): p. 201-202.
9. Strachan, T. and A.P. Read, *Human molecular genetics 2*. 2nd ed1999, Oxford: BIOS Scientific. xxiii, 576 p.
10. Nagaishi, M., et al., *Chromosome abnormalities identified in 347 spontaneous abortions collected in Japan*. Journal of Obstetrics and Gynaecology Research, 2004. **30**(3): p. 237-241.
11. Sethakulvichai, W., et al. *Estimation of band level resolutions of human chromosome images*. in *Computer Science and Software Engineering (JCSSE), 2012 International Joint Conference on*. 2012.
12. Bickmore, W.A., *Karyotype Analysis and Chromosome Banding*, in *eLS2001*, John Wiley & Sons, Ltd.
13. Moradi, M., S.K. Setarehdan, and S.R. Ghaffari. *Automatic locating the centromere on human chromosome pictures*. in *Computer-Based Medical Systems, 2003. Proceedings. 16th IEEE Symposium*. 2003.
14. Loganathan, E., M.R. Anuja, and N. Madian. *Analysis of human chromosome images for the identification of centromere position and length*. in *Point-of-Care Healthcare Technologies (PHT), 2013 IEEE*. 2013.
15. Moradi, M., S.K. Setarehdan, and S.R. Ghaffari. *Automatic landmark detection on chromosomes' images for feature extraction purposes*. in *Image and Signal Processing and Analysis, 2003. ISPA 2003. Proceedings of the 3rd International Symposium on*. 2003.
16. Howe, B., A. Umrigar, and F. Tsien, *Chromosome preparation from cultured cells*. J Vis Exp, 2014. **28**(83): p. 50203.
17. Lee, C., et al., *Limitations of chromosome classification by multicolor karyotyping*. Am J Hum Genet, 2001. **68**(4): p. 1043-7.



18. Speicher, M.R., S. Gwyn Ballard, and D.C. Ward, *Karyotyping human chromosomes by combinatorial multi-fluor FISH*. Nat Genet, 1996. **12**(4): p. 368-75.
19. Imataka, G. and O. Arisaka, *Chromosome analysis using spectral karyotyping (SKY)*. Cell Biochem Biophys, 2012. **62**(1): p. 13-7.
20. Catalina, P., et al., *Conventional and molecular cytogenetic diagnostic methods in stem cell research: A concise review*. Cell Biology International, 2007. **31**(9): p. 861-869.
21. Schrock, E., et al., *Multicolor spectral karyotyping of human chromosomes*. Science, 1996. **273**(5274): p. 494-497.
22. Weiss, M.M., et al., *Comparative genomic hybridisation*. Mol Pathol, 1999. **52**(5): p. 243-51.
23. Bayani, J. and J.A. Squire, *Comparative genomic hybridization*. Curr Protoc Cell Biol, 2005. **22**(22).
24. Gajendran, V. and J.J. Rodriguez, *Chromosome counting via digital image analysis*. Icip: 2004 International Conference on Image Processing, Vols 1- 5, 2004: p. 2929-2932.
25. Rungruangbaiyok, S. and P. Phukpattaranont, *Chromosome image classification using a two-step probabilistic neural network*. Songklanakarin Journal of Science & Technology, 2010. **32**(3): p. 255-262.
26. Wang, X.W., et al., *Development and evaluation of automated systems for detection and classification of banded chromosomes: current status and future perspectives*. Journal of Physics D-Applied Physics, 2005. **38**(15): p. 2536-2542.
27. Piper, J. and E. Granum, *On Fully-Automatic Feature Measurement for Banded Chromosome Classification*. Cytometry, 1989. **10**(3): p. 242-255.
28. Jong-Man, C., *Chromosome classification using backpropagation neural networks*. Engineering in Medicine and Biology Magazine, IEEE, 2000. **19**(1): p. 28-33.
29. Roshtkhari, M.J. and S.K. Setarehdan, *A novel algorithm for straightening highly curved images of human chromosome*. Pattern Recognition Letters, 2008. **29**(9): p. 1208-1217.
30. Mohammadi, M.R., *Accurate localization of chromosome centromere based on concave points*. Vol. 2. 2012. 88-94.
31. Arachchige, A.S., et al. *An image processing algorithm for accurate extraction of the centerline from human metaphase chromosomes*. in *Image Processing (ICIP), 2010 17th IEEE International Conference on*. 2010.
32. Moradi, M. and S.K. Setarehdan, *New features for automatic classification of human chromosomes: A feasibility study*. Pattern Recogn. Lett., 2006. **27**(1): p. 19-28.
33. Madian, N. and K.B. Jayanthi, *Analysis of human chromosome classification using centromere position*. Measurement, 2014. **47**(0): p. 287-295.
34. Cho, J.M., *Chromosome classification using backpropagation neural networks - A process that overcomes nonlinearity problems to correctly classify giemsa-stained human chromosomes*. Ieee Engineering in Medicine and Biology Magazine, 2000. **19**(1): p. 28-33.
35. Groen, F.C.A., et al., *Human chromosome classification based on local band descriptors*. Pattern Recognition Letters, 1989. **9**(3): p. 211-222.
36. *A Threshold Selection Method from Gray-Level Histograms*. Systems, Man and Cybernetics, IEEE Transactions on, 1979. **9**(1): p. 62-66.

37. Agam, G. and I. Dinstein, *Geometric separation of partially overlapping nonrigid objects applied to automatic chromosome classification*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1997. **19**(11): p. 1212-1222.
38. Poletti, E., E. Grisan, and A. Ruggeri. *Automatic classification of chromosomes in Q-band images*. in *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*. 2008.
39. Kao, J.-h., J.-h. Chuang, and T. Wang, *Chromosome classification based on the band profile similarity along approximate medial axis*. Pattern Recogn., 2008. **41**(1): p. 77-89.
40. Madian, N. and K.B. Jayanthi. *Overlapped chromosome segmentation and separation of touching chromosome for automated chromosome classification*. in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*. 2012.
41. Wu, S.-T., A.C.G.d. Silva, and M.R.G. Márquez, *The Douglas-peucker algorithm: sufficiency conditions for non-self-intersections*. Journal of the Brazilian Computer Society, 2004. **9**: p. 67-84.
42. Kao, J., J. Chuang, and T. Wang, *Automatic Chromosome Classification Using Medial Axis Approximation and Band Profile Similarity*, in *Computer Vision – ACCV 2006*, P.J. Narayanan, S. Nayar, and H.-Y. Shum, Editors. 2006, Springer Berlin Heidelberg. p. 274-283.
43. Zuiderveld, K., *Contrast limited adaptive histogram equalization*, in *Graphics gems IV*, S.H. Paul, Editor 1994, Academic Press Professional, Inc. p. 474-485.
44. Trebuña, P. and J. Halčinová, *Experimental Modelling of the Cluster Analysis Processes*. Procedia Engineering, 2012. **48**(0): p. 673-678.
45. Kaufman, L. and P.J. Rousseeuw, *Introduction*, in *Finding Groups in Data 2008*, John Wiley & Sons, Inc. p. 1-67.
46. Kao, J.-h., J.-h. Chuang, and T. Wang, *Chromosome classification based on the band profile similarity along approximate medial axis*. Pattern Recognition, 2008. **41**(1): p. 77-89.
47. Gonzalez, R.C. and R.E. Woods, *Digital Image Processing (3rd Edition)* 2006: Prentice-Hall, Inc.
48. Romesburg, H.C., *Cluster analysis for researchers* 1984, Belmont: Lifetime Learning Publications. xiii,334p.

## Internetové zdroje

49. NCI and NCBI's SKY/M-FISH and CGH Database (2001),  
<http://www.ncbi.nlm.nih.gov/sky/skyweb.cgi>
50. CARR, Steven. *Www.mun.ca* [online]. Dostupné z:  
[http://www.mun.ca/biology/scarr/Histone\\_Protein\\_Structure.html](http://www.mun.ca/biology/scarr/Histone_Protein_Structure.html)
51. Adler, David, University of Washington, Department of Pathology. Dostupné z:  
<http://www.pathology.washington.edu/research/cytopages/idiograms/human/>

# Seznam zkratek

<b>CGH</b>	<i>Comparative genomic hybridization – Komparativní genomová hybridizace</i>
<b>CI</b>	<i>Centromerický index</i>
<b>DNA</b>	<i>Deoxyribonucleic acid – kyselina deoxyribonukleová</i>
<b>FISH</b>	<i>Fluorescence in situ hybridization – fluorescenční in situ hybridizace</i>
<b>MAT</b>	<i>Medial axis transform – transformace podélné osy</i>
<b>M-FISH</b>	<i>Multicolor Fluorescence in situ hybridization – vícebarevná fluorescenční in situ hybridizace</i>
<b>Mb</b>	<i>Megabazí</i>
<b>NOR</b>	<i>Nucleolus organizer region – oblast organizátoru jádérka</i>
<b>RMSD</b>	<i>Root mean square deviation (kvadratická odchylka)</i>
<b>SKY</b>	<i>Spectral karyotyping – spektrální karyotypování</i>
<b>UPGMA</b>	<i>Unweighted pair-group method using arithmetic averages</i>

# Seznam obrázků

Obrázek 1: Chromozomy barvené pomocí G-pruhování.....	11
Obrázek 2: Buněčný cyklus a fáze buněčného dělení.....	13
Obrázek 3: Proces kondenzace chromozomu. ....	14
Obrázek 4: Různé uspořádání DNA v chromozomech .....	14
Obrázek 5: Ideogramy G-pruhovaného lidského chromozomu č. 11. ....	15
Obrázek 6: Mikroskopický snímek metafázních chromozomů hybridizujících se sondami specifickými pro centromerickou oblast chromozomů 7 (zeleně) a 8 (červeně) .....	21
Obrázek 7: Karyotyp sestavený na základě metody M-FISH. ....	22
Obrázek 8: a) binární obraz chromozomu, b) obraz chromozomu po použití „distance transform“ funkce, c) topologická kostra chromozomu, získaná procesem ztenčování na obraze b). Převzato a upraveno z [32]. ....	26
Obrázek 9: Chromozomy barvené pomocí G-pruhování.....	27
Obrázek 10: Tvarový profil chromozomu, detekce centromery .....	28
Obrázek 11: Horizontální projekce s minimem naznačujícím pozici centromery.....	28
Obrázek 12: Vývojový diagram navrženého algoritmu .....	30
Obrázek 13: Šedotónová a binární verze téhož obrazu mitotických chromozomů .....	31
Obrázek 14: Jednotlivé chromozomy, binární a skeletonizované verze .....	32
Obrázek 15: Aproximace hranice chromozomu polygonem .....	33
Obrázek 16: Označení vrcholů čtyřúhelníkové oblasti společné pro oba klastry .....	33
Obrázek 17: Čtyři označená ramena chromozomu a překrývající se část.....	34
Obrázek 18: Kombinace ramen a středové společné části pro jeden chromozomální klastr .....	34
Obrázek 19: Výsledek separace překrývajících se chromozomů .....	35
Obrázek 20: Manuální separace dotýkajících se chromozomů .....	36
Obrázek 21: Hranová reprezentace chromozomu s naznačenou podélnou osou.....	37
Obrázek 22: Šedotónová verze, binární verze a topologická kostra chromozomu.....	38
Obrázek 23: Zvýšení kontrastu šedotónového obrazu chromozomu .....	39
Obrázek 24: Jasový profil chromozomu.....	40
Obrázek 25: Vývojový diagram klasifikačního algoritmu.....	41
Obrázek 26: Dendrogram (RMSD obsahu chromozomu) .....	44
Obrázek 27: Změna znaménka druhé derivace jasového profilu.....	45
Obrázek 28: Chromozomy zařazené do párů dle RMSD obsahu .....	47
Obrázek 29: Chromozomy klasifikované do párů pomocí RMSD jasových profilů .....	50
Obrázek 30: Chromozomy klasifikované do párů pomocí Pearsonova korelačního koeficientu .....	50
Obrázek 31: Chromozomy klasifikované do párů pomocí Pearsonova korelačního koeficientu, omezení výběru na 10% .....	51
Obrázek 32: Chromozomy klasifikované do párů pomocí Pearsonova korelačního koeficientu, tolerance 20% .....	51
Obrázek 33: Chromozomy klasifikované do párů pomocí Pearsonova korelačního koeficientu bez separovaných chromozomů, tolerance 10% .....	52
Obrázek 34: Správně seřazené homologní páry zařazené do příslušných ISCN kategorií.....	53

# Seznam tabulek

Tabulka 1: Minimální rozlišení na G-pruhovaných chromozomech pro detekci určité abnormality ....	15
Tabulka 2: Identifikační znaky pro hodnocení stupně kondenzace chromozomu .....	16
Tabulka 3: Rozlišení chromozomů podle centromerického indexu .....	17
Tabulka 4: Příklady obsahů jednotlivých chromozomů .....	38
Tabulka 5: Příklady délek jednotlivých chromozomů .....	38
Tabulka 6: Část distanční matice pro RMSD jasových profilů.....	43
Tabulka 7: Část distanční matice pro korelační koeficient jasových profilů.....	43
Tabulka 8: Procentuální úspěšnost při hledání homologních párů .....	48
Tabulka 9: Procentuální úspěšnost při hledání homologních párů .....	49

# Seznam příloh

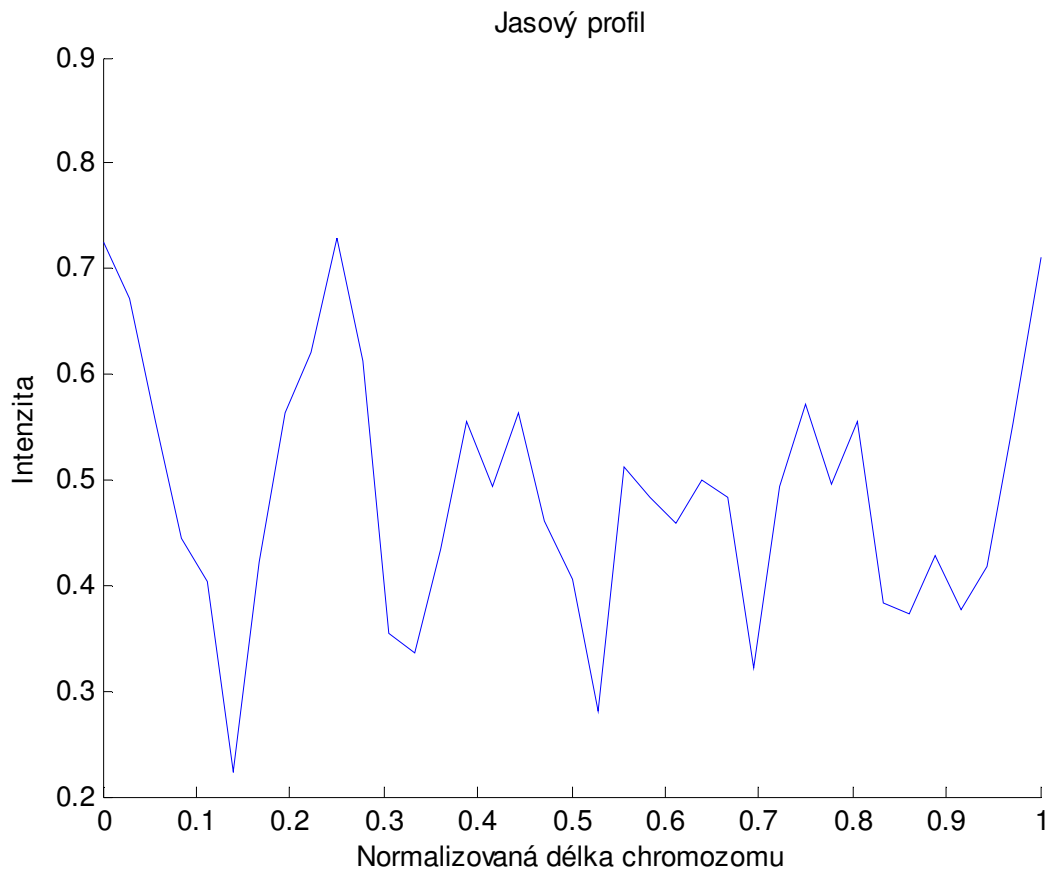
*Příloha A.*

Jasové profily

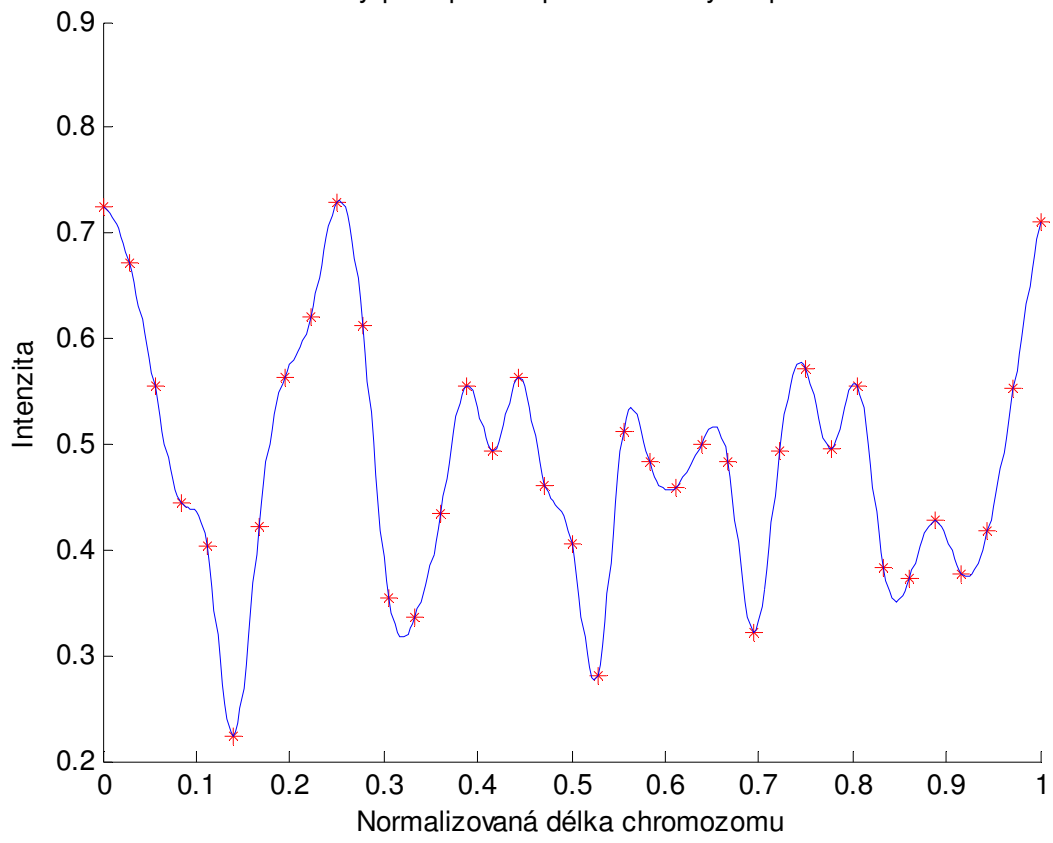
*Příloha B.*

Dendrogramy

# A. Jasový profil chromozomu

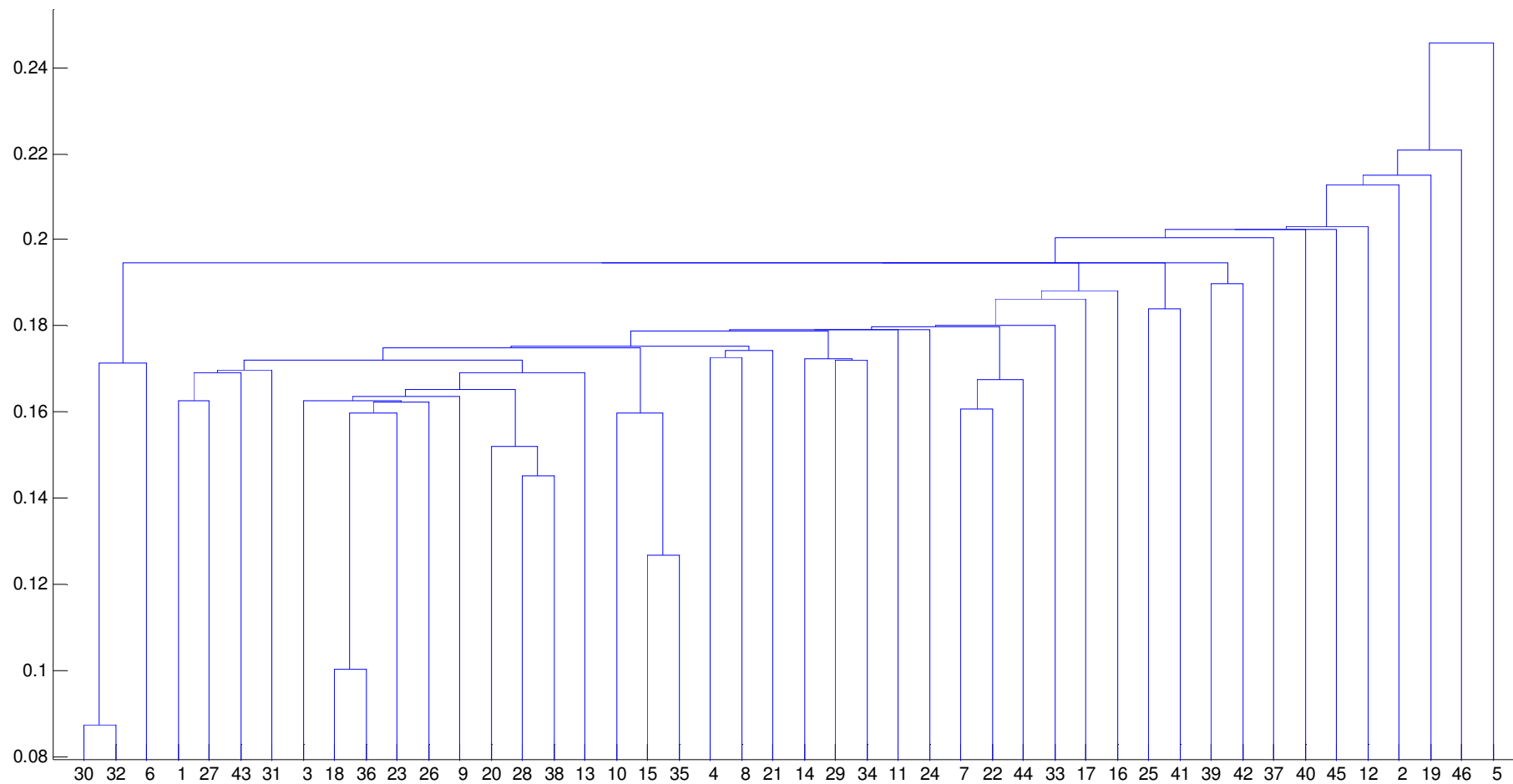


Jasový profil po interpolaci kubickým splinem



# B. Dendrogramy

Dendrogram pro RMSD jasových profilů





Dendrogram pro korelační koeficient jasových profilů

