

Česká zemědělská univerzita v Praze
Fakulta agrobiologie, potravinových a přírodních zdrojů
Katedra genetiky a šlechtění



**Celogenomová asociační studie rezistence mastitidy u
dojeného skotu**

Diplomová práce

Autor práce: Bc. Kristýna Forejt

Obor studia: Reprodukční biotechnologie (AMGB2)

Vedoucí práce: prof. Ing. Luboš Vostrý, Ph.D.

© 2021 ČZU v Praze

Čestné prohlášení

Prohlašuji, že svou diplomovou práci „Celogenomová asociační studie rezistence mastitidy u dojeného skotu“ jsem vypracovala samostatně pod vedením vedoucího diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou citovány v práci a uvedeny v seznamu literatury na konci práce. Jako autorka uvedené diplomové práce dále prohlašuji, že jsem v souvislosti s jejím vytvořením neporušila autorská práva třetích osob.

V Praze dne 25. 3. 2021

Poděkování

Ráda bych touto cestou poděkovala Svazu chovatelů holštýnského skotu za poskytnutí dat pro asociační analýzu a panu prof. Ing. Luboši Vostrému, Ph.D. za jeho vždy vstřícné vedení, věnovaný čas a cenné rady.

Celogenomová asociační studie rezistence mastitidy u dojeného skotu

Souhrn

Klinická mastitida je zánětlivé onemocnění mléčné žlázy, které má rozsáhlé negativní dopady na ekonomiku i welfare v chovech mléčného skotu. V rámci této práce byla nejprve sestavena rešerše českých i mezinárodních odborných publikací seznamující s problematikou mastitidy, její etiologií, epizootologií a závažnými ekonomickými dopady. Následně byly prezentovány také aktuální poznatky o možném genetickém založení odolnosti vůči tomuto onemocnění a bioinformatické metody vhodné k jeho studiu.

V experimentální části práce byla za použití DNA čipu Illumina BovineSNP50 DNA Analysis BeadChip získána genotypová data o 53 218 jednonukleotidových polymorfismech (SNP) od celkem 1 258 dojených krav holštýnského plemene a podrobena celogenomové asociační analýze s cílem identifikovat úseky genomu související s rezistencí vůči mastitidě. Chovatelé dotyčných zvířat poskytli pro účely této práce binární fenotypová data o výskytu mastitidy.

Po očištění dat na základě úspěšnosti genotypování, frekvence minoritní alely a populační stratifikace bylo u celkem 51 557 SNP a 1 042 jedinců přistoupeno k vlastní analýze za použití specializovaného softwaru využívajícího obecný lineární model. U dvou polymorfismů, *BTA-121769-no-rs* a *BTB-00265951*, byla pozorována statisticky významná asociace s výskytem mastitidy – oba se nacházejí na chromozomu BTA6. Detekované SNP byly zasazeny do kontextu genomové mapy a diskutovány s výsledky dosavadních publikací obdobného zaměření. Bylo zjištěno, že oba polymorfismy leží mimo transkribovanou oblast, ale v těsné blízkosti genů významných pro imunitní odpověď, což nadále zvyšuje pravděpodobnost jejich významu pro odolnost vůči mastitidě.

Klíčová slova: mastitida, asociační studie, GWAS, SNP, skot, *Bos taurus*

Genome-wide association study for clinical mastitis traits in dairy cattle

Summary

Clinical mastitis is an inflammatory disease of the mammary gland which largely impacts profitability and welfare in dairy farming. In the theoretical part of this thesis, a comprehensive overview of mastitis, its aetiology, epizootology and economic impacts has been compiled using international scientific resources. Furthermore, current findings regarding the potential link of certain genotypes to the resistance to mastitis and the appropriate bioinformatic methods to study it have been presented.

In the experimental part of this thesis, DNA microarray data has been obtained from 1 258 Holstein cows via Illumina BovineSNP50 DNA Analysis BeadChip, examining 53 218 single-nucleotide polymorphisms (SNPs). Owners of these cows have provided binary phenotype data capturing the incidence of mastitis in each individual.

After data pruning, controlling for genotype missingness, minor allele frequency and population stratification, the total of 51 557 SNPs from 1 042 animals has been analysed with the appropriate software using the general linear model (GLM). Two SNPs, *BTA-121769-no-rs* and *BTB-00265951*, have demonstrated statistically significant associations – both located on the chromosome BTA6. The detected SNPs have been further examined via genome mapping and discussed with relevant publications. They have been found to lie outside of transcribed regions but within the immediate vicinity of genes essential for the immune response, which further supports the case for their significance in the resistance to mastitis.

Keywords: mastitis, association study, GWAS, SNP, cattle, *Bos taurus*

Obsah

1	Úvod	1
2	Vědecká hypotéza a cíle práce.....	2
2.1	Vědecká hypotéza	2
2.2	Konkrétní cíle práce	2
3	Literární přehled	3
3.1	Úvod do problematiky zánětů mléčné žlázy	3
3.1.1	Stavba vemene	3
3.1.2	Fyziologie vemene.....	5
3.1.3	Fyziologické složení mléka	6
3.1.4	Klasifikace zánětů vemene	6
3.1.4.1	Subklinická mastitida.....	6
3.1.4.2	Klinická mastitida	7
3.2	Mastitida u skotu	7
3.2.1	Etiologie	7
3.2.1.1	Environmentální původci.....	7
3.2.1.2	Infekční původci	8
3.2.2	Diagnostika.....	8
3.2.2.1	Zevní vyšetření mléčné žlázy.....	8
3.2.2.2	Vyšetření sekretu mléčné žlázy	10
3.2.3	Prevence	10
3.2.4	Metody léčby	12
3.2.5	Epidemiologie.....	12
3.2.6	Ekonomický význam	12
3.3	Genom skotu.....	13
3.4	Metody genomového mapování	14
3.4.1	Izolace DNA	14

3.4.2	PCR – polymerázová řetězová reakce	14
3.4.3	Detekce variability	16
3.4.3.1	Vazebná analýza	16
3.4.3.2	Analýza počtu změn kopií DNA	16
3.4.3.3	Exomové sekvenování	16
3.4.3.4	Homozygoní mapování	17
3.4.3.5	Analýza pomocí DNA čipů (<i>DNA microarray</i>)	17
3.4.4	Zpracování dat z DNA čipů	19
3.4.4.1	Převod surových dat do pracovního formátu	19
3.4.4.2	Čištění dat	20
3.4.4.3	Možnosti zápisu fenotypu	22
3.4.4.4	Stanovení hladiny významnosti	22
3.4.4.5	Vlastní asociační analýza	23
3.4.4.6	Vyhodnocení výsledků	23
3.5	Úskalí studia genetického založení odolnosti vůči mastitidě	25
3.6	Dosavadní asociační studie odolnosti vůči mastitidě	26
3.7	Jiná odvětví asociačních studií u skotu	29
3.8	Genomická selekce plemenných zvířat	31
4	Metodika	32
4.1	Sběr dat	32
4.2	Čištění dat	33
4.3	Stanovení hladiny významnosti	37
4.4	Asociační analýza	38
4.5	Grafické znázornění výsledků	40
4.5.1	Manhattan graf	40
4.5.2	Kvantil-kvantil (Q-Q) graf	41

4.6	Anotace nejvýznamnějších SNP	41
5	Výsledky	42
5.1	Výsledky stratifikační analýzy.....	42
5.2	Souhrn výsledků asociační analýzy	44
5.3	Významné SNP.....	46
5.3.1	<i>BTB-00265951</i>	48
5.3.2	<i>BTA-121769-no-rs</i>	48
5.3.3	<i>BTB-01283269</i>	49
5.3.4	<i>BTA-36062-no-rs</i>	49
5.3.5	<i>ARS-BFGL-NGS-109683</i>	49
5.3.6	<i>ARS-BFGL-NGS-37189</i>	50
5.3.7	<i>ARS-BFGL-NGS-104096</i>	50
5.3.8	<i>Hapmap43710-BTA-86183</i>	51
5.3.9	<i>BTA-28028-no-rs</i>	51
5.3.10	<i>ARS-BFGL-BAC-23887</i>	51
5.3.11	<i>ARS-BFGL-NGS-115947</i>	52
5.3.12	<i>ARS-BFGL-NGS-1837</i>	52
5.3.13	<i>Hapmap50627-BTA-23969</i>	53
5.3.14	<i>UA-IFASA-7226</i>	53
5.4	Žádoucí alelické sestavy významných SNP	54
6	Diskuse.....	55
6.1	Metody a významné SNP v předchozích asociačních analýzách.....	55
6.2	Funkce genů v blízkosti významných SNP	57
7	Závěr	59
8	Seznam použité literatury.....	60

1 Úvod

Kravske mléko je hojně využívaným zdrojem živin pro lidi po celém světě – jen v České republice bylo v roce 2019 vyprodukováno přes tři miliardy litrů (ČSÚ 2020). Vzhledem ke své rozšířenosti a ekonomickému významu je chov mléčného dobytka důkladně monitorován za účelem dosažení co největší efektivity; mezi lety 2002 a 2019 vzrostla průměrná roční dojivost (výtěžek mléka na jednu dojnici) z 5717,9 na 8471,4 litrů, tedy o 48 % (ČSÚ 2020). Zintenzivnění výroby však provází také vysoká zátěž na organismus zvířat.

Záněty mléčné žlázy neboli mastitidy byly zaznamenány u mnoha druhů savců včetně člověka, u žádného však nemají tak dalekosáhlé ekonomické dopady jako u tura domácího (*Bos taurus*). Průměrná ztráta na jednu klinickou mastitidu připadá v České republice na více než devět tisíc korun (Kvapilík 2014), přičemž prevalence tohoto onemocnění v našich chovech mléčného skotu činí přibližně 19,8 % (Kašná *et al.* 2018).

Není proto překvapením, že je mastitida intenzivně zkoumána se záměrem nalézt co nejúčinnější metody prevence. Jednou z průlomových událostí v tomto snažení bylo zpřístupnění nových, komplexnějších metod genetické analýzy – v nedávných celogenomových asociačních studiích, na něž tato práce navazuje, byla pozorována spojitost zvýšené odolnosti vůči mastitidě s určitými genotypy.

2 Vědecká hypotéza a cíle práce

2.1 Vědecká hypotéza

- Asociační studie na základě SNP čipů umožní detekovat oblasti genomu skotu zodpovědné za výskyt mastitidy.

2.2 Konkrétní cíle práce

- Cílem práce je na základě asociační studie pomocí dostupných molekulárně genetických dat (SNP markerů) analyzovat úseky genomu zodpovědné za výskyt mastitidy u dojeného skotu v České republice.

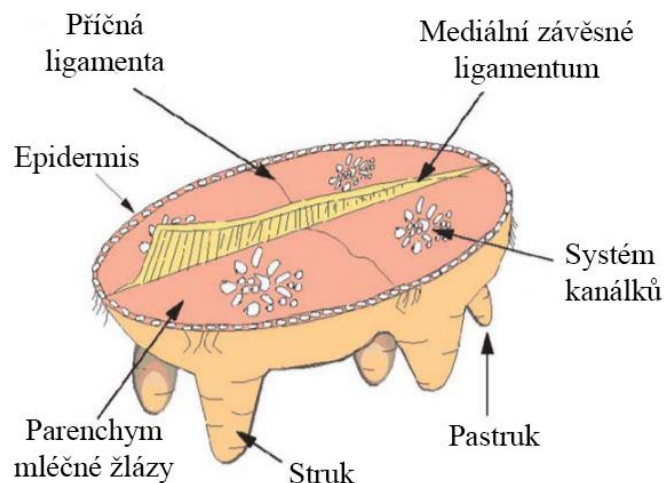
3 Literární přehled

3.1 Úvod do problematiky zánětů mléčné žlázy

Mastitida je souborný název pro záněty mléčné žlázy zaviněné specifickými patogeny (viz níže), v některých publikacích jsou však tímto pojmem označovány také nespecifické záněty projevující se změnami počtu buněčných elementů v mléce (VFÚ 2018; Kvapilík 2014). Vyskytuje se typicky v souvislosti s laktací a byla pozorována u široké škály druhů savců včetně člověka. Její průběh se rozděluje dle délky na perakutní, akutní, subakutní a chronický, dle projevů na klinický a subklinický (Hofírek & Haas 2003).

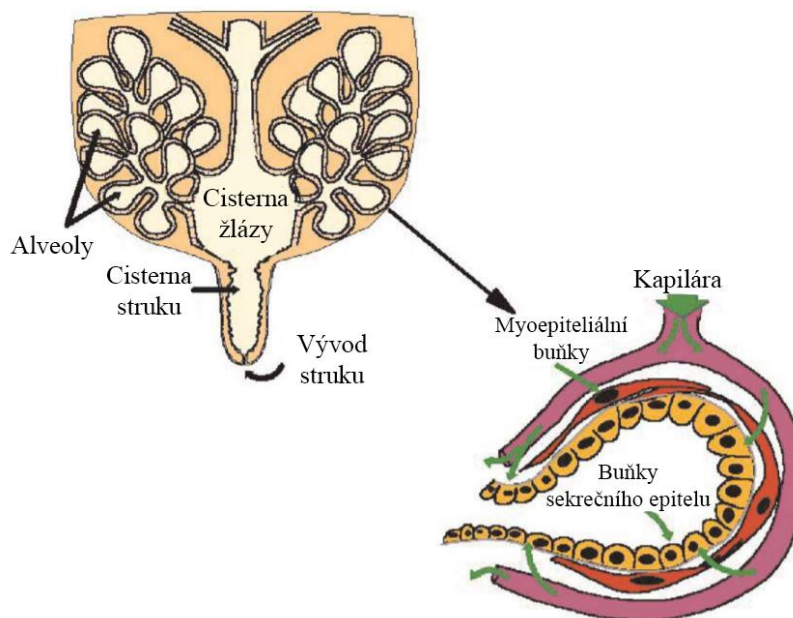
3.1.1 Stavba vemene

Mléčná žláza skotu má fyziologicky oválný tvar a je rozdělena na čtyři části, z nichž každá zahrnuje vlastní mléčnou žlázu, cévní zásobení, systém vývodných kanálků a struk. Podélně je vemeno rozděleno na dvě laterálně souměrné poloviny mediálním závěsným vazem, který jej upíná ke šlachám břišních svalů a kosti pánevní. Ze stran vemeno drží laterální závěsné vazy. Příčná ligamenta jsou méně výrazná a dělí vemeno asymetricky, kaudální část je větší než kraniální (znázorněno na Obrázku 1). Nejčastěji ze zadní strany vemene se mohou vyskytovat jeden nebo dva nefunkční struky, takzvané pastruky. Povrch vemene je pokryt jemnou srstí v závislosti na plemeni, struky jsou však vždy bez osrstění (Reece 2010).



Obrázek 1: Schéma příčného řezu vemenem (adaptováno z Nickerson & Ackers 2011).

Základní strukturou parenchymu mléčné žlázy je sekreční alveol. Mléčné alveoly představují váčky o průměru 50-250 μm ústící do vývodných kanálků. Na jejich povrchu se nachází kubický až cylindrický jednovrstevný epitel (znázorněno na Obrázku 2). Kromě vlastních alveolů je mléko ve vemeni skladováno také v cisternách, které přiléhají k vývodům struků (Reece 2010).



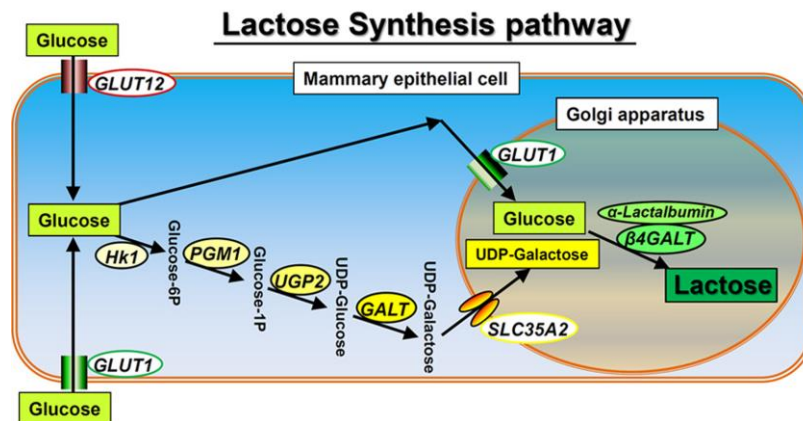
Obrázek 2: Schéma podélného řezu vemenem a alveolem (adaptováno z Nickerson & Ackers 2011).

3.1.2 Fyziologie vemene

Krev do vemene přivádí zevní stydká tepna, odváděna je zevní stydkou žílou skrze tříselný kanál. Některé žíly vemene ústí také do mléčné žíly (známé též jako kaudální povrchová nadbřišková žíla), která společně se zevní tříselnou žílou tvoří u báze vemene žilný prstenec. Intersticiální tekutina je odváděna lymfatickými cévami, hlavními mízními uzlinami pro oblast vemene jsou nadvemenné tříselné uzliny u kaudální části jeho základny (Reece 2010).

Buňky epitelu v sekrečních alveolech vemene vstřebávají z kapilární krve mléčné prekurzory a přetváří je na mléčný cukr (laktózu), tuk a bílkoviny (kasein, syrovátkovou bílkovinu). Pokud je lumen alveolu naplněn, tlak způsobí zploštění buněk sekrečního epitelu, které v důsledku přestávají syntetizovat nové mléko.

Prekurzorem laktózy je glukóza (viz Obrázek 3). V buňkách sekrečního epitelu dochází k postupné přeměně molekul glukózy na galaktózu za účasti enzymů HK1, PGM1, UGP2 a GALT. Takto připravená galaktóza je dopravena translokátorem SLC35A2 do Golgiho aparátu, kde spojením s druhou glukózou vzniká laktóza (Kobayashi *et al.* 2016).



Obrázek 3: Schéma syntézy laktózy (Kobayashi *et al.* 2016).

Mléčné bílkoviny jsou syntetizovány z aminokyselin z krevní plazmy, albuminu, globulinu, fibrinogenu a nebílkovinných dusíkatých látek. Tyto prekurzory jsou zpracovány na drsném endoplazmatickém retikulu, jmenovitě na kasein a syrovátkový protein. Mléčný tuk je

z naprosté většiny tvořen triacylglyceroly. Ty se do mléka dostávají buď rozkladem krevních lipoproteinů (VLDL) nebo syntézou *de novo* z kyseliny octové a β -hydroxymáselné (Agropress 2017).

3.1.3 Fyziologické složení mléka

Mléko zdravé krávy je bílé barvy, homogenní konzistence, bez zápachu, pH 6,2-6,7. Podíl jednotlivých složek je v průměru (Drbohlav & Vodičková 2002):

- 87 % vody
- 3,3 % bílkovin, z toho 75-85 % kasein, 15-22 % syrovátkové bílkoviny (β -laktoglobulin, α -laktalbumin, albumin, globulin), 1,9-3,3 % imunoglobuliny, 2-6 % proteázo-peptonová frakce.
- 4,4-4,7 % laktózy
- 2,5-6 % tuku
- 0,7 % minerálních látek (zejména vápník, draslík, fosfor, rozpustné i koloidně vázané)
- lipofilní i hydrofilní vitaminy

Nutno podotknout, že se složení mléka liší podle fáze laktace, přičemž kolostrum má tyto hodnoty výrazně posunuty ve prospěch tuku a bílkovin. Je také kyselější (pH 6-6,6) (Kopřiva 2011).

3.1.4 Klasifikace zánětů vemene

3.1.4.1 Subklinická mastitida

Subklinická mastitida je stav, při němž jsou do mléka z místa zánětu uvolňovány patogeny a zvýšené množství buněčných elementů, ale na mléčné žláze nejsou pozorovány typické projevy zánětu. Subklinická mastitida může předcházet klinické (Hofírek & Haas 2003).

3.1.4.2 Klinická mastitida

Klinická mastitida je stav, při němž jsou do mléka uvolňovány patogeny, zvýšené množství buněčných elementů, a zároveň jsou na mléčné žláze pozorovány typické projevy zánětu podle stupně závažnosti (Hofírek & Haas 2003):

- Klinická mastitida mírného stupně – nevýrazné příznaky zánětu, prokazatelné strukturní změny v mléce (vločky).
- Klinická mastitida středního stupně – středně závažné příznaky zánětu (otok, zarudnutí, zvýšená teplota, povolení závěsného aparátu), vodnaté mléko, systémové změny (snížený příjem potravy, tachykardie).
- Klinická mastitida vysokého stupně – závažné příznaky zánětu (výrazný otok, zarudnutí až vznik gangrenózních ložisek, vysoká bolestivost), výrazné změny sekretu (hnisavý až hemoragický), závažné systémové změny (horečka, anorexie, slabost).

3.2 Mastitida u skotu

3.2.1 Etiologie

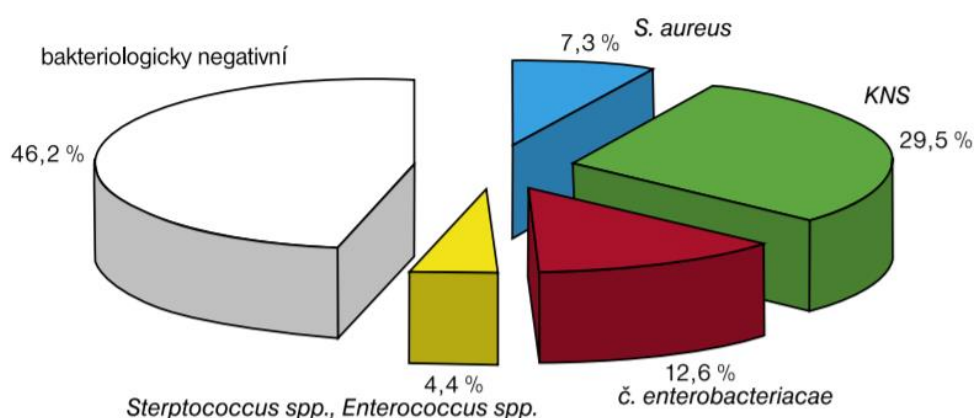
Původci mastitidy se u jednotlivých savců liší – u skotu se nejčastěji jedná o bakterie (viz Obrázek 4), ale zaznamenány jsou i mastitidy způsobené viry, kvasinkami, plísněmi a řasami (*Candida*, *Aspergillus*, *Galactomyces* a *Prototheca*) (Smola 2003).

3.2.1.1 Environmentální původci

Do organismu se dostávají skrze trávicí soustavu a kůži. Jejich výskyt je spojován zejména s nízkou úrovní hygieny prostředí. Patří mezi ně gramnegativní fekální bakterie, konkrétně *Escherichia coli*, *Enterobacter* spp., *Klebsiella* spp. a *Serratia* spp. Z grampozitivních se do této kategorie řadí *Streptococcus uberis* a některé stafylokoky (Smola 2003).

3.2.1.2 Infekční původci

Infekční původci mastitid jsou vázáni na prostředí mléčné žlázy a přenášejí se prostřednictvím kontaminovaných tělesných tekutin. Do této skupiny se řadí některé viry, z bakterií *Stafylococcus aureus* a další stafylokoky, *Streptococcus agalactiae*, *Streptococcus dysgalactiae* a *Mycoplasma bovis*. Tito původci nejčastěji zaviňují subklinické infekce (Hofírek & Haas 2003).



Obrázek 4: Frekvenční zastoupení bakteriálních původců zjištěných při mastitidách u skotu (Hofírek & Haas 2003). KNS značí koaguláza-negativních stafylokoky.

3.2.2 Diagnostika

3.2.2.1 Zevní vyšetření mléčné žlázy

Hlavními metodami zevní diagnostiky mastitidy u skotu je adspekce (pozorování) a palpace (pohmat). Při adspekci je posuzována pevnost závěsného aparátu vemene, tvar a velikost vemene, barva kůže, tvar struků a jejich schopnost zadržet mléko. Při palpaci je posuzována teplota, bolestivost a pevnost vemene, kontroluje se také zduřelost mízních uzlin (VFU 2018).

Závěsný aparát je fyziologicky pevně upnutý k břišní stěně, čímž se zaslouhuje o typický polovejčitý tvar vemene. S věkem a narůstajícím počtem laktací dochází k jeho povolení. To způsobuje komplikace při strojním dojení a vytváří predispozici pro zánětlivá onemocnění – názorné srovnání na Obrázku 5.



Obrázek 5: Srovnání pevného (nalevo) a uvolněného (napravo) závěsu vemene (VFU 2018).

Zdravé vemeno by mělo být bilaterálně souměrné a znatelně rozdělené na čtyři části příslušící k jednotlivým strukům. Z bočního pohledu by měly být zadní (kaudální) čtvrtě větší než přední (kraniální). Patologickým nálezem je narušení symetrie zmenšením (atrofií) nebo naopak otokem jedné ze čtvrtí (VFU 2018).

Barva kůže je fyziologicky růžová nebo odpovídající pigmentaci jedince, příznakem zánětu je zarudlost, případně až gangrenózní černo-fialové zbarvení. Na kůži vemene by se neměly vyskytovat rány a oděrky, které představují vstupní bránu pro některé původce mastitidy.

Při vyšetření struků se posuzuje jejich tvar, velikost, počet a pevnost jejich svěrače. Měly by být kuželovité, jeden až dva centimetry dlouhé, v počtu čtyř. Je možný výskyt třetího páru nefunkčních struků (nejčastěji na zadní části vemene) – funkční pastruky představují riziko úniku mléka a místo vstupu infekce. Predispozicí pro záněty je výhřez strukového kanálku, který vzniká příliš intenzivním dojením (VFU 2018).

Při palpačním vyšetření se srovnává teplota jednotlivých čtvrtí – vyšší teplota jedné čtvrti je příznakem zánětu. Zanícená čtvrť nebo čtvrtě vykazují bolestivost, která se zjišťuje podle obranné reakce zvířete na tlak. Na zánět ukazuje také zduření supramamárních mízních uzlin, které lze nahmatat z kaudální strany nad vememem (VFU 2018).

3.2.2.2 Vyšetření sekretu mléčné žlázy

Prvním krokem při vyšetření sekretu je posouzení struktury mléka v misce s černým dnem. Za fyziologický nález se považuje bílé homogenní mléko nebo kolostrum bez zápachu. Za patologický nález se považuje netypický zápach, vločky v mléce, vodnatý sekret, žlutý sekret či růžově až červeně zbarvený sekret (s krví).

Pokud se v mléce nevyskytnou viditelné změny, ale vemenem jeví známky zánětu, přistupuje se ke stájovému NK testu. V případě pozitivního nálezu následuje kultivace a volba vhodné léčby.

- NK test neboli kalifornský test je stájová metoda detekce zvýšeného počtu somatických buněk v mléce. Do mléka se fyziologicky uvolňují epiteliální buňky z kůže či mléčné žlázy a bílé krvinky, které do něj pronikají z krve (makrofágy, neutrofilové a lymfocyty). Počet somatických buněk na jeden litr mléka závisí na fázi a pořadí laktace (zvýšený je například u starších dojnic nebo v kolostru), příliš vysoký je však typickým znakem zánětu mléčné žlázy. Test funguje na principu reakce činidla s buněčnými membránami – při vysokém počtu buněk je výsledný roztok hustší a podle pH mění barvu (při zánětech je pH vyšší). Strojové měření počtu somatických buněk pomocí optických metod může kalifornský test nahradit a detekci zpřesnit (Hofírek & Haas 2003).
- Stájová kultivace se využívá pro stanovení vhodné léčby – do Petriho misky rozdělené na tři části se sterilní kličkou naočkuje vzorek mléka a inkubuje se při teplotě 37 °C po 20 hodin. Jedna třetina média v misce je neselektivní, druhá je selektivní na gramnegativní bakterie, třetí na grampozitivní. Podle vzešlých kultur je možné zvolit antibiotika cílená na konkrétní skupinu původců (Hofírek & Haas 2003).

3.2.3 Prevence

Vzhledem k různorodosti možných původců mastitidy jsou základní doporučení pro její prevenci poměrně obecná (Weisberg 2013):

- I. Udržování čistého prostředí bez odvrátitelných stresorů.

- II. Pečlivé očištění struků před připojením dojící jednotky. K čištění struků by se neměla používat běžná voda z vodovodu, která je pro řadu původců mastitidy vhodným prostředím.
- III. Vyšetření struků před dojením na oděrky či jiná poranění. Poraněné struky by neměly být intenzivně dojeny.
- IV. Namáčení struků do speciálního germicidního roztoku před dojením a gelu mezi dojeními. Roztoky zanechávají struk suchý a dezinfikovaný pro připojení dojící jednotky, gely pro použití mezi dojeními vytvoří na povrchu struků film, který zabraňuje vstupu patogenů (viz Obrázek 6). K aplikaci dezinfekčních prostředků se používají aplikační kalíšky, které by měly být mezi aplikacemi u různých dojnic důkladně očištěny.
- V. Pravidelná kontrola konzistence mléka ze všech čtvrtí vemene pomocí NK testu.
- VI. Vhodné používání dojících jednotek – dojení by měl provádět dostatečně kvalifikovaný personál. Nesprávná aplikace může způsobit poranění struku a tím vytvořit vstupní bránu pro patogeny.
- VII. Šlechtitelské programy zaměřené na zvýšení odolnosti vůči mastitidám.



Obrázek 6: Struky namočený v gelu Barrier LuxDIP 50 B.

3.2.4 Metody léčby

Léčba má dva hlavní cíle: snížení počtu buněčných elementů v mléce na takovou hodnotu, aby se kráva mohla co nejdříve vrátit do produkce, a eliminaci původce, je-li to dosažitelné.

Základem léčby zůstávají antibiotika. Ta mohou být aplikována přímo do mléčné žlázy (intramamárně), nebo mohou být podávána systémově (podkožně, do svalu; k zánětu se dostávají skrze krev). Intramamární antibiotika jsou vhodná zejména u krav s mírným zánětem v jedné čtvrti vemene (Laven 2016).

U krav se silným zánětem nebo zánětem postihujícím více čtvrtí se přistupuje k systémové léčbě. Antibiotická látka je zvolena podle výsledků kultivačního testu. U závažných zánětů se navíc přistupuje k aplikaci nesteroidních antiflogistik (léků s protibolestivým, protihorečnatým a protizánětlivým účinkem). Staré krávy s pokleslými vemeny a krávy s vysokou recidivitou onemocnění je často ekonomicky nevýhodné léčit, bývají tudíž vyřazeny z chovu (Laven 2016).

3.2.5 Epidemiologie

Prevalence mastitidy se liší lokálně podle oblasti i hygienických podmínek konkrétního chovatelského zařízení, od průměrné prevalence 74,7 % v Etiopii (Abebe *et al.* 2016) přes 44,7 % v Indii (Bansal & Gupta 2009), 25 % v Norsku (Østerås *et al.* 2006) po 19,8 % v České republice (Kašná *et al.* 2018).

3.2.6 Ekonomický význam

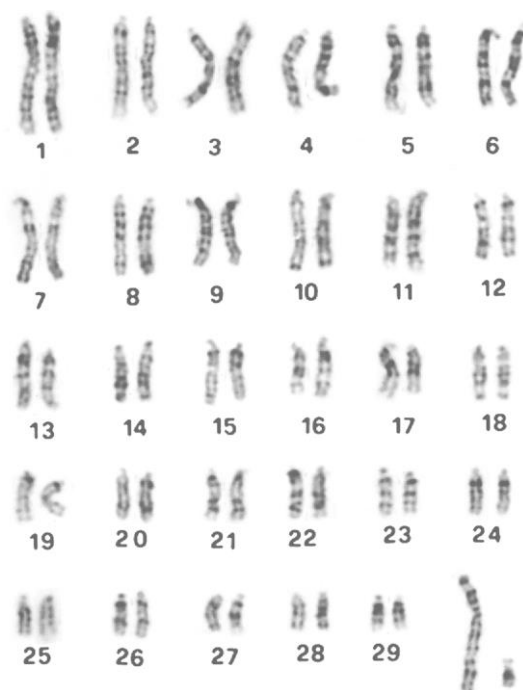
Užitkovost je základním ukazatelem hodnoty jakéhokoliv produkčního chovu, v případě chovu zaměřeného na produkci mléka hovoříme o mléčné užitkovosti. Mléčná užitkovost je kombinací dojnosti (schopnosti tvořit mléko) a dojitelnosti (schopnosti uvolňovat mléko z mléčné žlázy), souhrnně se tyto dvě vlastnosti vyjadřují jako dojivost (množství získaného mléka za dané časové období) (Nosková 2016).

Mastitida negativně ovlivňuje obě složky dojivosti. Vzhledem k vysoké prevalenci a nutnosti dočasného či trvalého vyřazení nemocného kusu z produkce přináší mastitida

mlékárenskému průmyslu významné ztráty – v České republice přes 9 000 Kč (rozpětí 4 000-18 000 Kč) na jednu mastitidu (Kvapilík 2014), 265 EUR (přes 7 000 Kč) průměrně v Evropské unii (Nielsen 2009), studie z USA uvádí 444 USD (přes 11 000 Kč) (Rollin *et al.* 2015). Na ztrátách se podílejí z 53 % nižší tržby za prodané mléko, z 20 % vyřazování nemocných krav, z 14 % náklady na léky a léčení krav, ze 7 % práce na ošetřování nemocných krav a 6 % srážky z nákupní ceny mléka (Kvapilík 2014).

3.3 Genom skotu

Standardní genom tura domácího je dlouhý okolo 2 857 605 192 bází (bp) a je rozčleněn do 30 párů chromozomů, z toho 29 párů autozomů a jednoho páru gonozomů (XX u krav, XY u býků). Jednotlivé autozomy se v literatuře označují BTA[číslo příslušného chromozomu] (zkratka pro *Bos taurus autosome*), například BTA6, BTA29,... Normální karyotyp skotu je zachycen na Obrázku 7.



Obrázek 7: Karyotyp skotu (Iannuzzi 1996)

3.4 Metody genomového mapování

V posledních letech je intenzivně zkoumána dědičná složka náchylnosti, respektive odolnosti vůči mastitidě. Jedná se zřejmě o polygenně podmíněný znak s nízkou dědivostí (h^2), odhadem Banos *et al.* (2017) 0,09-0,18. Při úspěšné identifikaci dostatečného množství souvisejících genotypů je však možné uplatnit poznatky genetické analýzy v procesu šlechtění.

Pomocí celogenomových asociačních studií (v odborné literatuře často pod zkratkou GWA nebo GWAS jako *genome-wide association study*) byla vytipována řada kandidátních jednonukleotidových polymorfismů (SNP). Pro asociační studii jsou klíčové následující kroky:

3.4.1 Izolace DNA

Izolace předchází všem dalším metodám genomového inženýrství, které pro své fungování vyžadují čistou nukleovou kyselinu.

Izolace DNA se provádí kombinací enzymatických činidel (proteinázy, RNázy), pufrů (lyzační pufr, promývací pufr, ředící pufr) a fyzikálního působení (centrifugace přes filtr vázající DNA, míchání třepačkou, zahřívání). Koncentraci DNA ve výsledném izolátu je následně nezbytné zkontrolovat pomocí UV/VIS (ultrafialovo-viditelné spektroskopie).

3.4.2 PCR – polymerázová řetězová reakce

Polymerázová řetězová reakce je v molekulárních technologiích rutinně využívána pro zmnožení a označení izolované DNA. Nejprve je nezbytné připravit roztok, který obsahuje primery (navržené specificky pro úsek, který chceme amplifikovat), směs dNTP (deoxyribonukleosidtrifosfátů všech čtyř bází), chlorid hořečnatý, termostabilní DNA polymerázu (oblíbená je například Taq polymeráza z horkovodní bakterie *Thermus aquaticus*, ale na základě požadavků konkrétní reakce se dá použít i řada jiných), purifikovanou vodu a reakční pufr (nejčastěji kombinace KCl a Tris HCl, upravuje pH a koncentraci solí v reakci) (Lorenz 2012).

Pro každou PCR je třeba dvou typů primerů – přední „forward“ primery, komplementární k sekvenci nacházející se ve směru 5' od zvoleného úseku, a zadní „reverse“ primery ve směru

3'. Primery se volí tak, aby nasedaly co nejbližší z obou stran zvoleného úseku. Mívají délku 15-30 bází a nesmějí vykazovat nežádoucí komplementaritu v rámci vlastního vlákna ani páru (Lorenz 2012). Pro jejich navržení lze využít specializovaných programů, například Primer3 (Untergasser *et al.* 2012) nebo Primer-BLAST (Ye *et al.* 2012).

PCR roztok s cílovou DNA je vystaven amplifikačnímu procesu o třech fázích, které se opakují do dosažení kýženého počtu kopií (obvykle třicetkrát) (Lorenz 2012):

1) Denaturace – roztok je po 30 až 60 sekund vystaven teplotě 95 °C, DNA se rozplétá (denaturuje) na dvě oddělená vlákna vlivem uvolnění vodíkových můstků.

2) Annealing (nasednutí) primerů – teplota se na 30 sekund sníží na hodnotu odpovídající parametrům primerů, obvykle mezi 50 a 60 °C. Primery se naváží na komplementární místa templátové DNA.

3) Extenze – teplota se zvýší na 72 °C, za působení DNA polymerázy se od primerů syntetizují komplementární vlákna.

Přístroj, který se pro tuto reakci používá, se nazývá termální cykler.

Metoda PCR je primárně určena pro amplifikaci krátkých fragmentů, při amplifikaci celého genomu, jako je potřeba pro GWAS, se přistupuje k různým způsobům, jak toto omezení obejít:

- LA-PCR (ligation-anchored PCR) – nescifická RNA ligáza je použita ke kovalentnímu navázání určitého oligonukleotidu ke koncům cílové DNA. Pro následnou PCR jsou použity primery komplementární k tomuto přidanému oligonukleotidu (Troutt *et al.* 1992).
- IRS-PCR (interspersed repetitive sequence-PCR) – metoda využívající mikrosatelitních repetit, které se u zkoumaného druhu často vyskytují, jakožto primerů (Phan *et al.* 2000).
- PEP-PCR (primer extension preamplification PCR) a DOP-PCR (degenerate oligonucleotide-primed PCR) – metody využívající směsi více degenerovaných (=

vzájemně podobných) či částečně degenerovaných primerů. Pro zpřesnění se navíc přidává i druhá DNA polymeráza s opravnou aktivitou (Arneson *et al.* 2008).

3.4.3 Detekce variability

3.4.3.1 Vazebná analýza

Genová vazba je tendence vloh dědit se společně, souvisí s jejich prostorovou blízkostí v genomu. Principem vazebné analýzy je identifikace chromozomální oblasti, která se dělí se sledovaným fenotypem v jedné či více příbuzných skupinách, pomocí celogenomového genotypování pravidelně rozložených markerů se známou pozicí a následného bioinformatického výpočtu spojitosti. Pro každý zkoumaný marker je vypočítána celková pravděpodobnost (podíl pravděpodobnosti, že jsou dva lokusy ve vazbě, a pravděpodobnosti, že nejsou). Logaritmus tohoto poměru se označuje jako LOD skóre. Markery, které soustavně korelují se sledovaným fenotypem (nerekombinují), tak vymezují kandidátní oblast velkou v řádu jednotek centimorganů (cM) pro hledaný gen (Stránecký 2015).

3.4.3.2 Analýza počtu změn kopií DNA

Strukturní změny o velikosti více než 50 bp se označují jako *copy number variations* (CNV). Mohou být děděné nebo vznikající *de novo*, lze je detekovat pomocí CGH čipů (komparativní genomová hybridizace) a byly spojeny s řadou fenotypů včetně polygenně podmíněných znaků, u holštýnského skotu například základními parametry produkce mléka (Liu *et al.* 2019).

3.4.3.3 Exomové sekvenování

Exomové sekvenování se používá u vzácných fenotypů, kde se předpokládá *de novo* kauzální mutace v oblastech genomu, které obsahují kódující oblasti pro vybrané (případně všechny) proteiny. Výsledek sekvenování se následně porovnává s exomovou knihovnou (Stránecký 2015).

3.4.3.4 Homozygoní mapování

Homozygotní mapování je metodou využívanou zejména pro hledání mutací podmiňujících vzácné fenotypy. Vychází z předpokladu příbuznosti rodičů jedince vykazujícího sledovaný znak a spočívá v detekci dlouhých autogenních (tj. pocházejících od společného předka) oblastí v genomu. Lze provádět v kombinaci s metodou DNA čipů (Stránecký 2015).

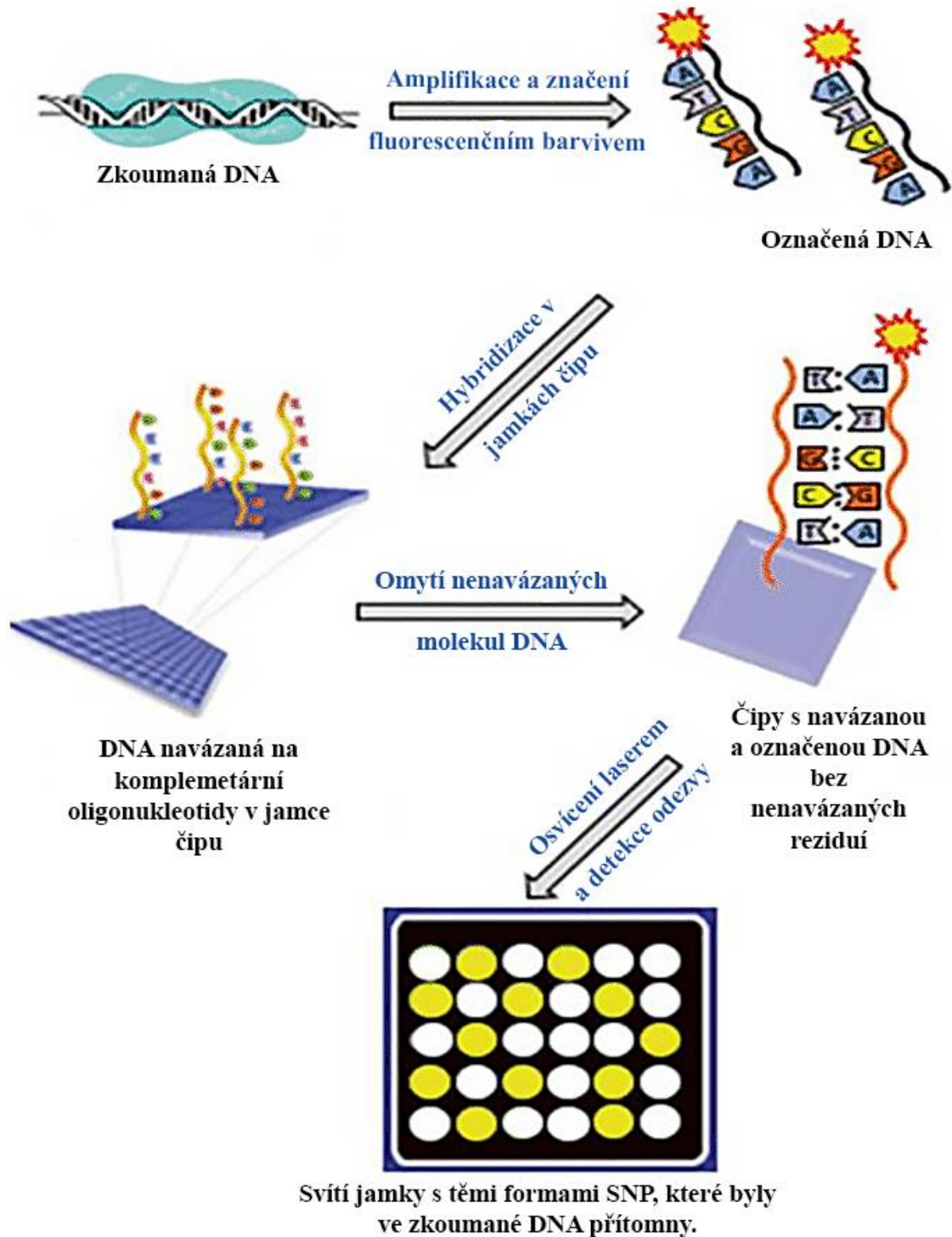
3.4.3.5 Analýza pomocí DNA čipů (*DNA microarray*)

DNA čipy jsou malé destičky (odtud jejich jméno) s velkým množstvím jamek, ve kterých jsou předem pevně navázané oligonukleotidové sondy. Každá jamka váže jeden unikátní úsek se známým jednonukleotidovým polymorfismem (*single-nucleotide polymorphism* neboli SNP; místo v genomu, kde se od sebe jedinci mohou lišit v jedné nukleové bázi). Postup analýzy je rámcově tento (LaFramboise 2009) (přehled na Obrázku 8):

1. Sledovaná DNA je předem namnožena metodou PCR (viz výše) a fluorescenčně označena.
2. Fluorescenčně značená DNA se nanese na čip a nechá se hybridizovat se sondami.
3. Nenavázaná DNA je z čipů odstraněna (washing).
4. Čipy jsou osvětleny UV lampou. Jamky s navázanou DNA svítí, to znamená, že se u zkoumaného jedince příslušný SNP vyskytuje. Pokud jamka nesvítí, znamená to, že ve zkoumané DNA tento polymorfismus přítomen není. Heterozygotní založení je detekováno na základě světelné odpovědi na dvě různé alely stejného SNP místa a nižší intenzity světla. Každý čip tedy obsahuje dvojnásobný počet jamek oproti počtu sledovaných SNP – jednu pro každou alelu.

DNA čipy lze nechat sestavit na zakázku u specializovaných laboratoří, v současné době je však pro analýzu SNP u skotu dostupná i řada komerčně nabízených sad – standardem pro nyní již rutinní genotypování v chovu skotu se stal Illumina BovineSNP50 DNA Analysis BeadChip (Illumina, Inc., San Diego, USA), který snímá data o 53 218 SNP. Pro rozsáhlejší genotypování například plemenných býků jsou k dispozici vysokokapacitní čipy Illumina BovineHD DNA Analysis Kit (777 962 SNP) a Affymetrix Axiom Genome-Wide BOS 1 Array

(648 874 SNP) (Affymetrix, Inc., Santa Clara, USA) (Rincon *et al.* 2011), bohužel zatím nejsou mezi chovateli mléčného skotu široce využívané především vzhledem ke své zvýšené finanční náročnosti.



Obrázek 8: Schéma detekce SNP pomocí DNA čipu (adaptováno z Mishra *et al.* 2017)

3.4.4 Zpracování dat z DNA čipů

3.4.4.1 Převod surových dat do pracovního formátu

Produktem SNP genotypování je obvykle soubor se surovými daty (.txt), který se k dalšímu zpracování upravuje pomocí specializovaného softwaru – pro čipy od firmy Illumina (Illumina, Inc., San Diego, USA) lze použít například Illumina GenCall, Illumina GenomeStudio nebo R s k tomu určeným balíčkem funkcí, jako je argyle (Morgan 2015) či Bioconductor (Huber *et al.* 2015). Kromě tohoto souboru je nutná také mapa sledovaných SNP, jež je obvykle dodávána výrobcem čipu, a soubor s daty o fenotypu, který je nutno sestavit zvlášť.

Požadovaný výstupní formát prvního kroku závisí na softwaru zvoleném pro další zpracování – mezi nejuniverzálněji využitelné patří formáty programu PLINK (kombinace .ped a .map nebo .bed, .fam a .bim, viz Obrázek 9), VCF nebo Hapmap (Bradbury *et al.* 2020).

*.ped									*.map			
FID	IID	PID	MID	Sex	P	rs1	rs2	rs3	Chr	SNP	GD	BPP
1	1	0	0	2	1	CT	AG	AA	1	rs1	0	870000
2	2	0	0	1	0	CC	AA	AC	1	rs2	0	880000
3	3	0	0	1	1	CC	AA	AC	1	rs3	0	890000

*.fam							*.bed	*.bim					
FID	IID	PID	MID	Sex	P	Contains binary version of the SNP info of the *.ped file. (not in a format readable for humans)		Chr	SNP	GD	BPP	Allele 1	Allele 2
1	1	0	0	2	1			1	rs1	0	870000	C	T
2	2	0	0	1	0			1	rs2	0	880000	A	G
3	3	0	0	1	1			1	rs3	0	890000	A	C

Covariate file				
FID	IID	C1	C2	C3
1	1	0.00812835	0.00606235	-0.000871105
2	2	-0.0600943	0.0318994	-0.0827743
3	3	-0.0431903	0.00133068	-0.000276131

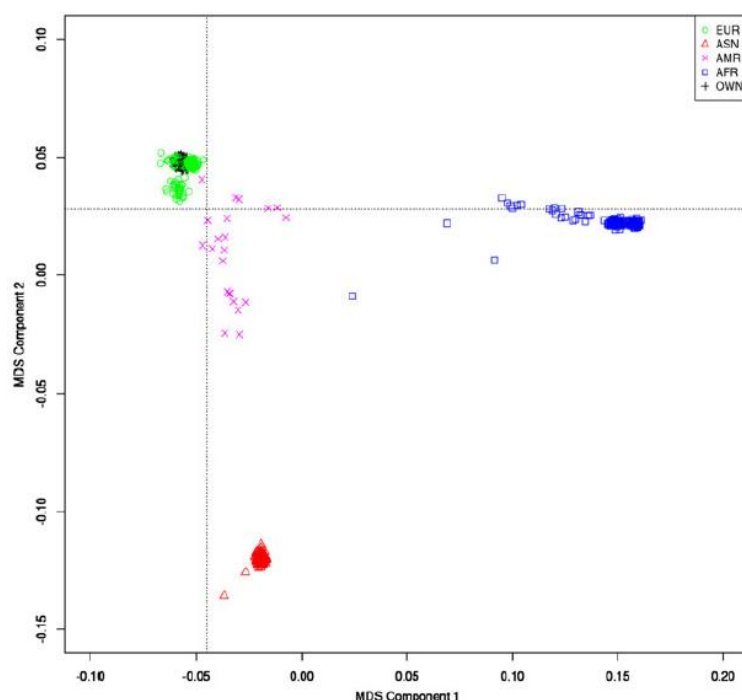
Legend			
FID	Family ID	rs{x}	Alleles per subject per SNP
IID	Individual ID	Chr	Chromosome
PID	Paternal ID	SNP	SNP name
MID	Maternal ID	GD	Genetic distance (morgans)
Sex	Sex of subject	BPP	Base-pair position (bp units)
P	Phenotype	C{x}	Covariates (e.g., Multidimensional Scaling (MDS) components)

Obrázek 9: Přehled formátů programu PLINK s legendou (Marees *et al.* 2018).

3.4.4.2 Čištění dat

Před vlastní asociační analýzou je nutné data očistit od jedinců a SNP, u nichž podíl chybějících dat přesahuje 10 %, a od jedinců s mendelisticou chybou nad 5 % (v případě dostupných dat o příbuznosti). Dále se vyřazují vzácné varianty pomocí MAF filtru (= *Minor Allele Frequency*, frekvence minoritní alely, typicky nastaveno na 0,01 nebo 0,05) (Illumina 2016).

V následujícím kroku je nutné provést stratifikační analýzu, která pomůže odhalit možné zkreslující efekty v případě, že se v souboru vyskytují oddělené shluky genetiky podobných jedinců (viz Obrázek 10). Statistické modely používané pro asociační analýzu totiž potřebují ke svému správnému fungování populaci co nejbližší normálnímu rozdělení (Rawlings *et al.* 1998). Tímto způsobem je také možno identifikovat a vyřadit z analýzy jednotlivce, kteří se zbytku souboru silně vymykají a taktéž by mohli výsledky zkreslovat (v anglicky psané literatuře označování jako *outliers*) (Marees *et al.* 2018).



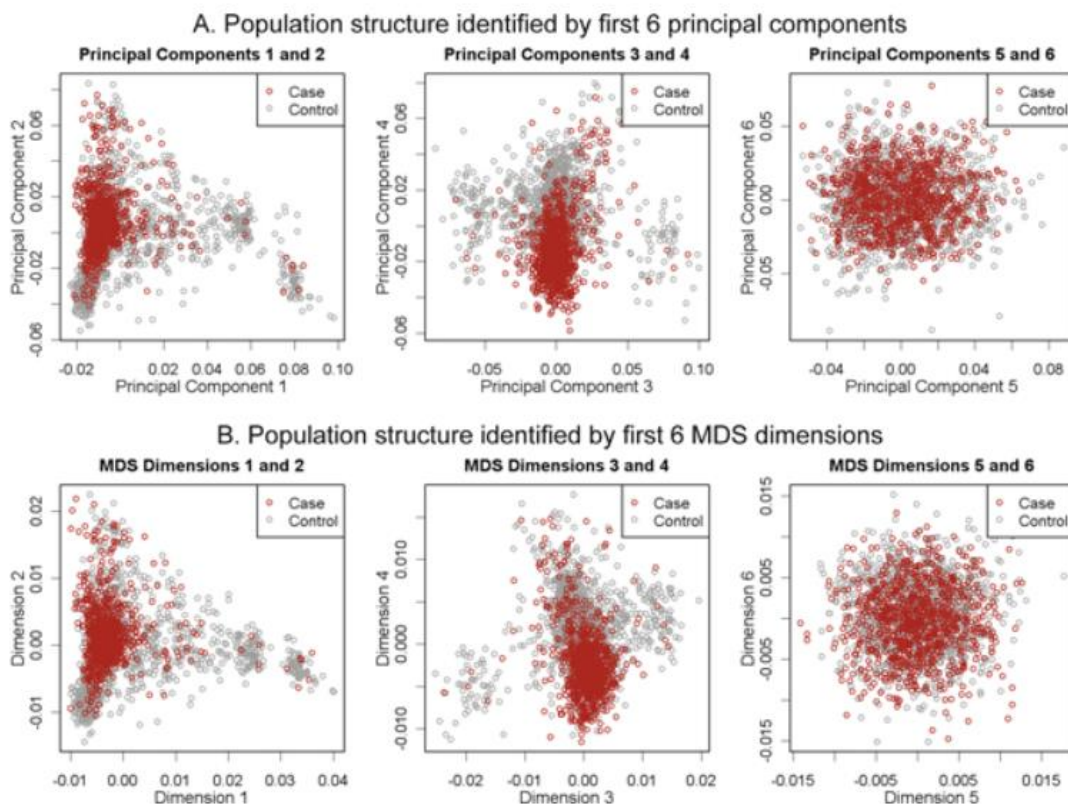
Obrázek 10: Graf genetické stratifikace z Marees *et al.* (2018), barvy značí příslušnost k etnikům severoamerického obyvatelstva. Mezi jednotlivými etniky je patrná vysoká stratifikace a lze v rámci nich identifikovat jedince, kteří se významně vymykají.

Ke kontrole stratifikace se obvykle používá jeden ze dvou přístupů: analýza hlavních komponent (*Principal Component Analysis*, PCA) nebo multidimenzionální škálování (*Multidimensional Scaling*, MDS).

Analýza hlavních komponent spočívá v tom, že se z původních proměnných (genotypových dat) každého jedince vytvoří lineární kombinací nové, umělé proměnné (komponenty), které vyjadřují, do jaké míry se jedinec podílí na celkové variabilitě populace. Čím více komponent je použito, tím vyšší procento variability vysvětlují. Jedinci s nejextrémnějšími hodnotami komponent jsou vyřazováni (Wang *et al.* 2009).

Multidimenzionální škálování využívá párové IBS (*Identity By State*) vzdálenosti získané na základě skrytého Markovova modelu (blíže v Purcell *et al.* 2007). Opět se jedná o kvantifikaci variability – čím více alel mají dva jedinci společných, tím blíže si v rámci páru jsou. Při analýze IBS vzdáleností je sestavena matice těchto párových statistik mezi každými dvěma jedinci (kde 0 = zcela identický a 1 = zcela odlišný). Na základě multidimenzionálního škálování pak lze data z této matice převést do podoby lineárních vzdáleností mezi jedinci v jednotlivých dimenzích – opět platí, že čím více dimenzí je posuzováno, tím větší procento variability vysvětlují. Jedinci s nejextrémnějšími pozicemi v dimenzích jsou vyřazováni.

Jednotlivé komponenty i dimenze lze proti sobě graficky promítat na bodovém grafu (srovnání obou metod na Obrázku 11). Čištění dat lze provést v programech PLINK (Purcell *et al.* 2007), TASSEL 5 (Bradbury *et al.* 2007) nebo vhodných knihovnách R.



Obrázek 11: Grafické srovnání výsledků stratifikační analýzy pomocí PCA o šesti komponentách (nahore) a MDS o šesti dimenzích (dole) (Wang et al. 2009). Je zřejmé, že použité metody jsou srovnatelně účinné.

3.4.4.3 Možnosti zápisu fenotypu

Fenotyp se z bioinformatického hlediska dělí do dvou hlavních kategorií – kvalitativní a kvantitativní. Nejjednodušším typem kvalitativního fenotypu je fenotyp binární – sledovaný fenotypový projev u jedince je či není pozorován. Binární fenotyp se zapisuje typicky [1,0], případně [1,2]. Nebinární kvalitativní fenotyp je pro potřeby analytických programů také nutno označit číselně (např. 0 může značit žlutou, 1 oranžovou a 2 modrou barvu duhovky). Kvantitativní fenotyp je vyjádřen číselnou hodnotou znaku.

3.4.4.4 Stanovení hladiny významnosti

Při mnohonásobném testování roste pravděpodobnost chyby prvního druhu (zamítnutí nulové hypotézy, když platí), je proto nutné upravit hladinu významnosti. Nejčastější metodou

je Bonferroniho korekce, která spočívá ve vydělení základní hladiny významnosti (u biologických dat rovna 0,05) počtem testování, tedy počtem sledovaných SNP (Marees *et al.* 2018). Hladina významnosti i naměřené p-hodnoty se v publikacích pro lepší přehlednost obvykle vyjadřují jejich záporným dekadickým logaritmem (vyšší hodnota záporného logaritmu odpovídá vyšší významnosti, nikoliv naopak jako u p-hodnot).

3.4.4.5 Vlastní asociační analýza

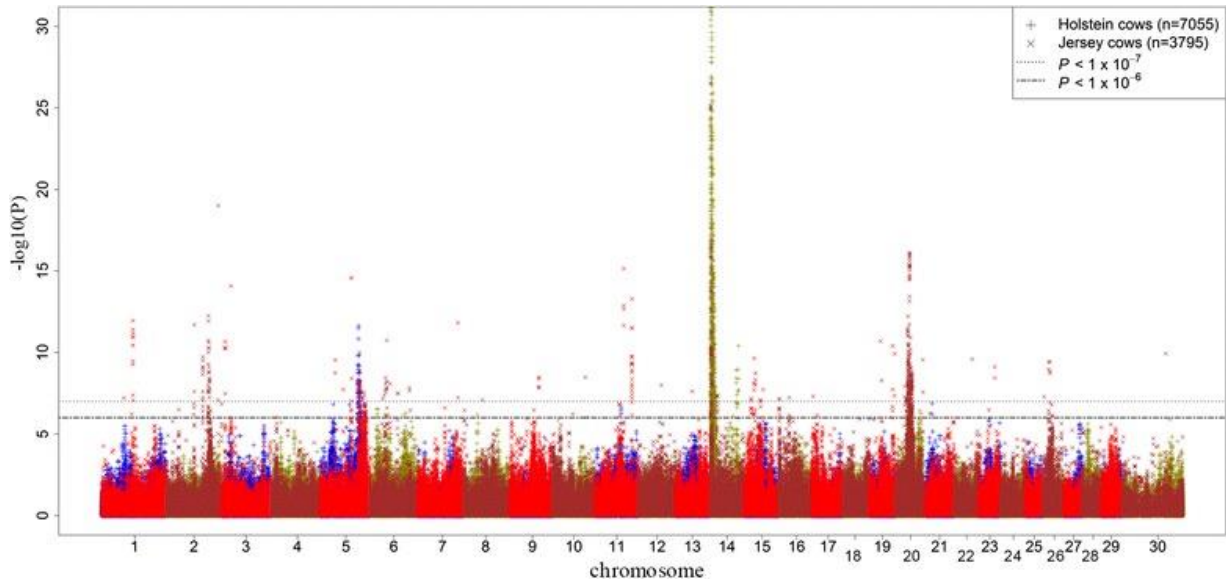
Vlastní asociační analýzu lze realizovat v jednom z řady dostupných specializovaných softwarů: PLINK (Purcell *et al.* 2007), BLINK (Huang *et al.* 2018), SNPTest (Marchini & Howie 2010), TASSEL 5 (Bradbury *et al.* 2007), GEMMA (Zhou & Stephens 2012) a mnoho dalších – tyto softwary jsou obvykle volně dostupné z internetových zdrojů, volba záleží na potřebách konkrétního výzkumu a povaze dat. Metodika, použité statistické modely a podoba výstupních dat se odvíjí od použitého softwaru, společným cílem však je stanovení signifikance (p-hodnoty) pro jednotlivé SNP markery.

3.4.4.6 Vyhodnocení výsledků

Ve finální fázi jsou SNP seřazeny podle p-hodnoty od nejnižší a porovnány s hladinou významnosti, která byla pro danou studii zvolena. SNP, jež byly zjištěny jako významné, jsou podrobeny bližšímu prozkoumání, dohledání nejbližších genů a diskusi s příslušnými publikacemi. Na základě odhadu efektu alel lze určit žádoucí alelické sestavy – pokud je patologický fenotyp zapsán jako 1 a fyziologický jako 0, vybíráme sestavu s nejnižším efektem jako nejvíce žádoucí. Protože však v chovu skotu není obvyklé šlechtit zvláště otcovské a mateřské linie, je v případě heterozygotní žádoucí sestavy vhodné vyhodnotit i to, která alela příznivěji ovlivňuje fenotyp ve své homozygotní sestavě. Žádoucí sestavu lze uplatnit pouze v rámci jednoho plemene – může se stát, že SNP reprezentuje jen místo v genové vazbě se skutečnou kauzální mutací a mezi plemeny se v něm vyskytuje odlišnost (Porto-Neto *et al.* 2014).

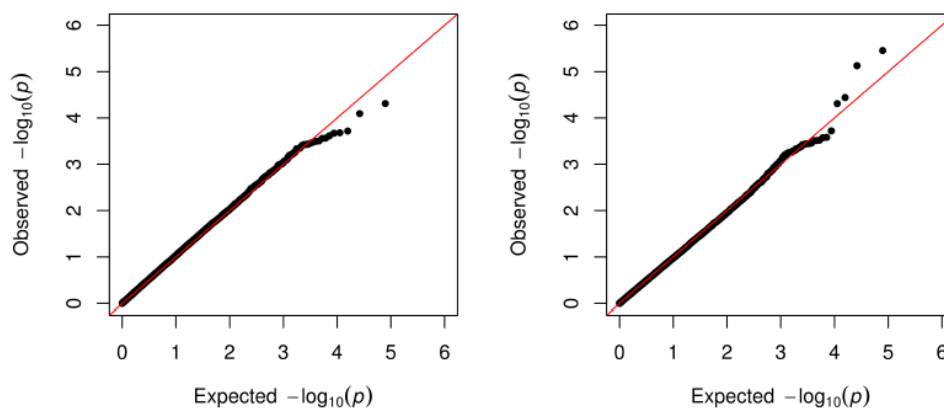
Výsledky lze vizualizovat typem grafu nazývaným podle svého vzezření „Manhattan“. Tento graf zachycuje $-\log_{10}(p)$ na ose y a pozici jednotlivých SNP na ose x, obvykle je doplněn

o jednu nebo více vodorovných přímek reprezentujících hladinu významnosti po úpravě a přepočtení na záporný dekadický logaritmus – příklad na Obrázku 12. Body nad hladinou významnosti reprezentují statisticky významné polymorfismy.



Obrázek 12: Manhattan graf pro dojivost u holštýnského a jerseykého skotu (Aliloo et al. 2015).

Druhým typem grafu, který se z výsledků asociační analýzy typicky sestavuje, je tzv. kvantil-kvantil neboli Q-Q graf (Obrázek 13). Srovnává teoretický (očekávaný) a naměřený kvantil p-hodnot, slouží pro kontrolu normálního rozdělení – očekáváme, že pouze velmi malá část sledovaných SNP bude vykazovat asociaci, tudíž většina křivky bude odpovídat normálnímu rozdělení (teoretickým kvantilům). Příliš výrazná odchylka od střední osy může u GWAS ukazovat na stratifikaci populace nebo vazbovou nerovnováhu (*linkage disequilibrium*, jev, kdy je měření zkresleno zvýšeným výskytem určité kombinace alel polymorfismů, které jsou spolu ve vazbě, přičemž pouze část z nich je skutečně odpovídá kauzální oblasti) (Marees et al. 2018).



Obrázek 13: Příklady uspokojivého průběhu Q-Q grafů pro GWAS (Welderufael et al. 2018).

3.5 Úskalí studia genetického založení odolnosti vůči mastitidě

Ačkoliv je mastitida intenzivně zkoumána, výsledky předchozích asociačních analýz se často neshodují. Genetický výzkum odolnosti vůči mastitidě je komplikován několika faktory:

- **Nízká dědivost** – výsledky pro znaky s nízkou dědivostí mohou být snadno zkresleny vlivy prostředí. Tento efekt se dá minimalizovat pouze částečně pečlivou prioritizací asociovaných SNP a zpětnou kontrolou jejich biologické významnosti. Například Cai et al. (2018) ověřovali výsledky asociační analýzy následnými testy exprese přílehlých genů ve tkáni mléčné žlázy při napadení bakterií *Escherichia coli*. Pro zjednodušení celého procesu se může místo konkrétních informací o fenotypu využívat odhad plemenné hodnoty pro sledovaný znak, který bývá u hospodářských zvířat dostupnější a s výskytem mastitid koreluje.
- **Náročnost na kvalitu i kvantitu dat** – z povahy asociační analýzy vyplývá, že se její přesnost zlepšuje s velikostí vzorku. Pro vložku s takto nízkou dědivostí je vhodné analyzovat data od nejméně stovek jedinců, přičemž všichni musí mít co nejpřesněji známý fenotyp a co nejkonstantnější podmínky (potenciální zkreslující faktory). Laboratorní zpracování takového množství vzorků je finančně i časově náročné. V chovech dobytka je ale často již zavedeno genotypování plemenných zvířat pomocí

SNP čipů, je proto možné využít již existujících dat z databází chovatelských organizací. Ačkoliv však aktuálně nejvyužívanější čip pokrývá přes 50 000 SNP markerů, jedná se pouze o nepatrný zlomek bezmála 30 000 000 SNP, které byly při opakovaném sekvenování genomu skotu zaznamenány (Snelling *et al.* 2019).

- **Nedokonalé zmapování genomu** – mnohé SNP jsou identifikovány programy pro asociační analýzu jako kauzální, nenachází se ale v blízkosti žádného známého genu. Aktuálně není možno s jistotou říct, že jejich oblast nemá dosud nezjištěný biologický význam (mimo geny se mohou nacházet například enhancer/silencer sekvence ovlivňující transkripci) (Halfon 2020). Pro prioritizaci se v okolí kandidátních SNP dohledávají vazná místa pro transkripční faktory (TFBS) a signální proteiny. K tomu se využívá specializovaných predikčních programů, například PhysBinder (Broos *et al.* 2013) nebo GeneMANIA (Warde-Farley *et al.* 2010) (Suravajhala & Benso 2017).

3.6 Dosavadní asociační studie odolnosti vůči mastitidě

Rané asociační studie využívaly mikrosatelitních markerů a jejich závěry byly pouze přibližné předpovědi, na kterých chromozomech by se měly kauzální mutace koncentrovat (například Lund *et al.* 2007).

Sodeland *et al.* (2010) provedli laboratorní analýzu 2589 plemenných býků norského červeného skotu a pro posouzení souvislosti s rezistencí vůči mastitidě sledovali jejich dcery. Sledováno bylo 17 349 SNP. Data byla zpracována souhrnně i v rámci souborů dat podle pořadí laktace, ve které byla mastitida u dcer zaznamenána, přičemž pro každý SNP bylo vypočítáno skóre věrohodnostního poměru (LRT). Významná souvislost s prevalencí mastitidy byla zjištěna u celkem 29 SNP (Obrázek 14). Nejsilnější asociace byla zjištěna u SNP *BTA-119376* na chromozomu *BTA6*.

BTA	SNP	Position (bp)	D1 LRT	D2 LRT	CD LRT	TP
2	BTA-47902	68 074 185	4.92	3.26	15.50	CM3
2	BTA-120621	103 854 622	3.29	3.05	12.72	CM1
2	BTA-120624	103 892 096	5.03	3.26	15.52	CM1
2	rs29025784	112 396 633	4.06	4.10	14.62	CM1
4	rs29020694	90 418 076	5.65	3.23	20.14	CM1
6	BTA-119376	90 670 190	3.28	12.74	32.36	CM1
6	BTA-119376	90 670 190	2.76	3.16	15.46	CM6
6	BTA-77136	94 544 954			18.80	CM4
6	BTA-109071	95 256 811			14.18	CM6
6	BTA-77356	96 189 520	3.08	5.18	15.60	CM1
7	BTA-78563	22 841 729			16.92	CM5
7	BTA-99486	26 347 331			16.80	CM5
9	BTA-84619	88 416 414	2.95	2.71		CM5
10	BTA-79349	92 455 424			14.20	CM6
13	rs29022774	62 928 524			13.90	CM4
13	rs29022775	62 928 533			15.90	CM4
14	rs29012803	17 289 087			14.90	CM5
14	BTA-34796	40 769 096			15.36	CM2
14	BTA-34923	45 153 040			18.34	CM2
14	BTA-111421	47 425 522			15.96	CM2
16	BTA-38543	33 335 777			14.64	CM1
17	BTA-103789	34 861 876	3.02	2.76	12.84	CM5
20	BTA-25160	31 659 731			15.16	CM2
20	BTA-50239	35 530 051	2.96	3.84	16.02	CM1
20	BTA-50236	35 861 339			17.10	CM1
20	rs29021255	38 261 470			14.88	CM3
20	BTA-19985	43 267 496			20.76	CM1
20	BTA-22852	50 780 080			14.36	CM5
29	rs29027496	45 602 144			14.36	CM1

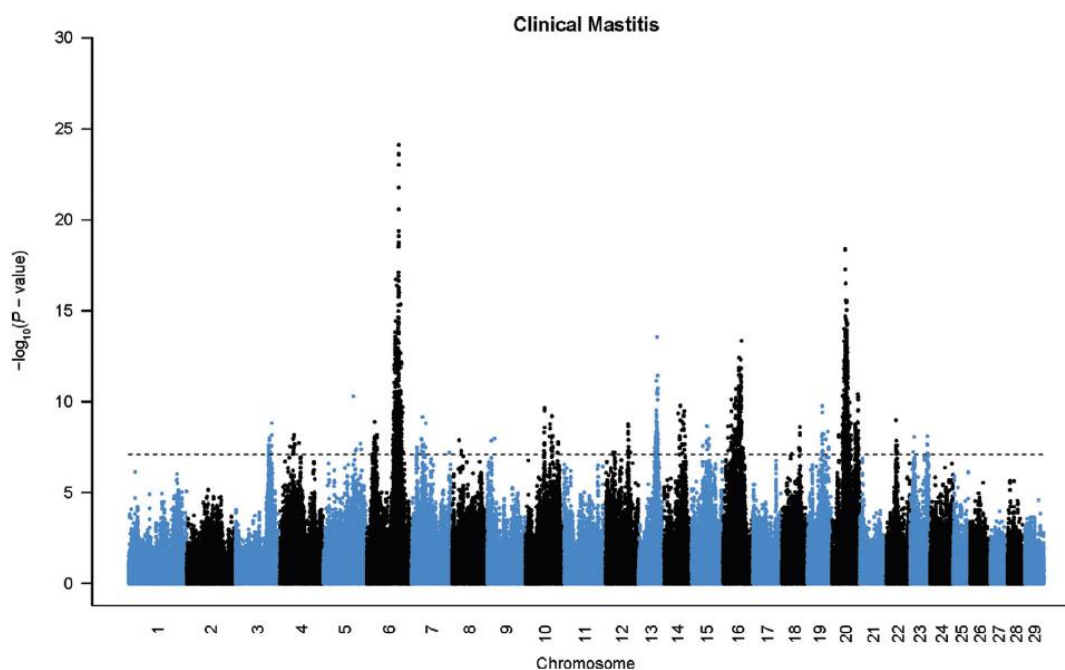
CM, clinical mastitis; GWAS, genome-wide association studies; LRT, likelihood ratio test.

Obrázek 14: Kandidátní polymorfismy zjištěné Sodeland et al. (2010).

Tiezzi et al. (2014) využili podobné metodiky, zahrnuli však pouze data z prvních laktací dcer plemenných býků v severoamerické populaci holštýnského skotu. Data dále zpracovali pro identifikaci obsáhlejších oblastí vlivu a jejich mechanismů účinku. Detekovali asociaci v genech *LY6K*, *LY6D*, *LYNX1*, *LYPD2*, *SLURP1*, *PSCA* na chromozomu BTA14, které jsou součástí lymfocytárního antigenního komplexu 6 (*LY6*), známého vlivem na regulaci činnosti neutrofilů. Dále také na chromozomu 2 v genech *IFIH1*, *LY75*, *DPP4*, *ITGB6* a *NR4A2*, které se rovněž účastní regulace specifické imunity. Kandidátní geny na chromozomu BTA20 pak ovlivňují metabolismus mléčné žlázy (*GHR*, *OXCT1*), produkci protilátek a fagocytózu (*C6*, *C7*, *C9*, *CIQTNF3*), supresi nádorového bujení (*DAB2*) nebo regulaci tvorby cytokinů (*PRLR*).

Ve stejném roce provedli asociační studii pomocí SNP čipů také Sahana et al. (2014). Analyzovali vzorek celkem 10 560 jedinců napříč třemi plemeny (holštýnský, norské červený

a jerseyjský skot). Zjistili, že některé SNP vykazují asociaci specificky pro jednotlivá plemena – u holštýnského skotu byla nejsilnější na chromozomu 6 (viz Obrázek 15), zejména v genech *GC* a *NPFRR2*, zatímco u norského červeného byly významnější oblasti na chromozomech BTA13 a BTA19, u jerseyjského zase BTA16 a BTA20.



Obrázek 15: Manhattan graf výsledků asociační analýzy rezistence vůči mastitidě u holštýnského skotu (Sahana *et al.* 2014).

Poměrně nedávná studie Szyda *et al.* (2019) se soustředila na propojení dvou výše zmíněných strategií – analýze SNP a analýze počtu změn kopií (*copy number variations*, CNV). Pracovali se skromnějším vzorkem čítajícím 13 párů nevlastních sester, z nichž jedna jeví náchylnost vůči mastitidě (opakovaný výskyt) a druhá nikoliv (žádný výskyt). Zjistili 191 CNV, v nichž byla delece typická (detekce alespoň v devíti zkoumaných párech) pro náchylné jedince, zatímco u zdravých se vyskytovala delší varianta. Oblasti těchto CNV korespondovaly s exony 46 genů, mezi nimi *APP* (BTA1), *FOXL2* (BTA1), *SSFA2* (BTA2), *OTUD3* (BTA2), *ADORA2A* (BTA17), *TXNRD2* (BTA17) a *NDUFS6* (BTA20). SNP, které vykazovaly obdobnou míru asociace v rámci příbuzných páru, bylo zjištěno pouze 17, nejvýznamněji v intronech *RNF122* (BTA27), *MET* (BTA4) a *WRN* (BTA27). Ze závěrů této práce tedy vyplývá možný zásadnější význam vícečetných delecí (CNV) ve srovnání s významem SNP.

Z výše uvedeného lze soudit, že mezi jednotlivými publikacemi panuje pouze omezená shoda a je nutné sjednocení poznatků prostřednictvím podrobnějšího studia této problematiky.

3.7 Jiná odvětví asociačních studií u skotu

Vzhledem k ekonomickému významu skotu a rychlému rozšíření genotypování pomocí DNA čipů jistě nepřekvapí, že se celogenomové asociační studie staly v posledních letech velmi populárními nejen ve věci rezistence vůči mastitidě. Zájem v případě mléčného skotu všeobecně padl na významné produkční faktory jako je dojivost, obsah tuku a bílkovin v mléce či ukazatele reprodukce.

K prvním patřili Kolbehdari *et al.* (2009), kteří takto studovali dojivost u holštýnského plemene. Jejich práce se zakládala na pouhých 1 536 SNP sledovaných u 462 holštýnských býků, přičemž jako fenotyp sloužily ukazatele užitkovosti jejich dcer. Detekovali tři statisticky významné SNP, po jednom na chromozomech 1, 21 a 23.

Širší genomický pohled na mléčnou užitkovost přinesli o rok později Jiang *et al.* (2010), kteří používali tehdy nově vyvinutý Illumina BovineSNP50 BeadChip (Illumina, Inc, San Diego, USA) čítající 54 000 SNP, pomocí něhož zpracovali vzorky populace o 2 180 kusech holštýnského skotu (14 býčích a jejich dcerách). Pro všechny sledované produkční ukazatele, tj. celková dojivost, obsah tuku a bílkovin v mléce (procentuální i absolutní) detekovali velké množství velmi silných asociací (106), zejména v relativně úzké oblasti na chromozomu 14.

Na práci těchto průkopníků v následujících letech navázaly další – Aliloo *et al.* (2015) provedli asociační analýzu dat pocházejících ze zřejmě dosud nejrozsáhlejšího vzorku, patřila 3 513 757 kusů holštýnského a jerseykého skotu. Detekovali pro dojivost významné SNP na chromozomech 5, 14 (zde suverénně nejvíce) a 20. Sermyagin *et al.* (2016) se zabývali ruskou populací holštýnského skotu, detekovali 14 asociovaných polymorfismů na chromozomech 1, 5, 9, 13, 14, 17, 20 a 27. Korkuč *et al.* (2021) sledovali mléčnou užitkovost u holandsko-fríského skotu, přičemž analýzu rozdělili podle fáze laktace. Pro ukazatele užitkovosti detekovali jiné SNP v rané fázi (chromozomy 8 a 9) než v pozdější fázi (především chromozom 6). Všechny publikace zmíněné v tomto odstavci těžily data pomocí Illumina BovineSNP50 BeadChip.

Jednu z nejširších studií co do počtu analyzovaných SNP (15 551 021) provedli Cai *et al.* (2019), kteří se zabývali prioritizací kandidátních genů pro ukazatele plodnosti u populace holštýnského skotu v severských státech – sledovali vazbové fáze asociovaných SNP, asociované úseky vyhledali ve funkční genomové mapě, díky čemuž vytipovali celkem deset kandidátních genů na chromozomu 1 (*DIRC*, *SEMA5B*, *ADCY5*, *MYLK*, *KALRN*, *UMPS*, *ITGB5*, *HEG1*, *PSMG1*, *MPD*, *BRWD1*) a po jednom genu na chromozomech 6 (*FRAS1*), 10 (*OTX2*), 13 (*KLF6*), 17 (*LIF*) a 24 (*SPIRE1*).

Jiné výzkumné týmy se mezitím zabývaly asociačními studiemi pro další ekonomicky významný znak – plodnost. Sahana *et al.* (2010) ji sledovali u švédské a dánské populace holštýnského skotu (2 531 plemenných býků), také oni používali Illumina BovineSNP50 BeadChip. Nejvýznamnější asociace našli na chromozomu 14, o něco slabší na chromozomech 3, 5, 10, 13, 19, 20 a 24. O rok později sledovali Schulman *et al.* (2011) plodnost ayrshirského skotu, zjistili celkem 11 celogenomově významných polymorfismů, a to rozprostřené na chromozomech 1, 2, 4, 12, 13, 20, 24 a 27.

Asociační studie u masných plemen jsou obecně prováděny méně z toho důvodu, že informace o hlavních produkčních parametrech lze získávat na základě fenotypu snadněji, s nižší potřebou čekání na data o užitkovosti dcer, než je tomu u býků mléčných plemen. Jsou však stále výhodné pro několik *in vivo* obtížně zjistitelných vlastností, jako je kvalita masa nebo konverze krmiva. Na ty se tudíž GWAS u masného skotu začaly soustřeďovat nedříve, počínaje Abe *et al.* (2008), kteří se konkrétně zabývali růstem a kvalitou masa japonského černého a limousinského skotu. Rozsah a metodika této práce odpovídala ranému datu její publikace – sledovali 318 mikrosatelitních markerů u 186 zvířat. Výsledkem bylo 9 asociovaných QTL na chromozomech 5 a 19.

Jako příklady současnějších studií na masném skotu lze uvést Bhuiyan *et al.* (2018) a Bedhane *et al.* (2019), kteří se zabývali ukazateli růstu, konverze krmiva a kvality masa u místního korejského plemene Hanwoo (2 109 a 2 110 kusů). Bhuiyan *et al.* detekovali 27 celogenomově významných SNP na chromozomech 4, 6 a 14, Bedhane *et al.* detekovali 107 SNP napříč čtrnácti chromozomy.

Na výsledky výše uvedených asociačních studií navázala řada dalších publikací, které se věnovaly bližšímu studiu konkrétních kandidátních oblastí, například Paredes-Sánchez *et al.* (2015), kteří vyhledávali kandidátní geny pro růstové parametry masného skotu, nebo Olsen *et al.* (2016), kteří se zaměřovali na mapování vytipovaného úseku chromozomu 6 pro odolnost vůči mastitidě a určili *GC* (*group specific component*) jako kandidátní gen zapříčiňující asociace detekované v této oblasti.

3.8 Genomická selekce plemenných zvířat

Genomická selekce plemenných zvířat je v současnosti již široce rozšířenou metodou ve šlechtitelství, která slouží pro zvýšení genetického zisku a zkrácení doby potřebné pro relativně spolehlivý odhad plemenné hodnoty nejen u příslušníků mléčných plemen skotu. Je založena na výpočtu genomické plemenné hodnoty (GEBV, *genomic enhanced breeding value*) pro jednotlivé ukazatele užitekosti (Hayes *et al.* 2009).

Pro odhad těchto hodnot se využívá statistické metody nejlepší nevychýlené předpovědi (BLUP, *best linear unbiased prediction*), která odhaduje GEBV relativně k aktuální referenční populaci. Zpracování dat referenční populace sdílí řadu podobností s metodou GWAS (také se jedná o lineární regresi jednonukleotidových polymorfismů a fenotypu), na rozdíl od ní však není jeho primárním účelem prioritizace malého množství nejvýznamnějších markerů, ale odhad efektu (kladného či záporného) pro každý posuzovaný marker či haplotyp – spoléhá se na rovnoměrné rozmístění sledovaných SNP v genomu takovým způsobem, aby se každý nacházel ve vazbové fázi s alespoň jedním dalším (Goddard & Hayes 2007). GEBV jedince pro danou vlastnost je následně vypočteno jako suma efektů jeho genotypu.

Výsledky asociačních studií mohou být užitečné pro zpřesnění odhadu genomických plemenných hodnot, jak pozorovali například Abo-Ismael *et al.* (2017), když efekt SNP detekovaných v GWAS zohlednili při posuzování genomických plemenných hodnot, které následně porovnávali se skutečnými fenotypovými hodnotami.

4 Metodika

4.1 Sběr dat

Genotypová data 1258 holštýnských krav byla poskytnuta Svazem chovatelů holštýnského skotu. Pocházela z čipu BovineSNP50 DNA Analysis BeadChip od společnosti Illumina (Illumina, Inc., San Diego, USA), který snímá celkem 53 218 SNP a je v současnosti dostupný chovatelům pro komerční testování. Data byla následně zkompileována do formátu programu PLINK (Purcell *et al.* 2007). Chovatelé dotyčných krav dále poskytli pro účely asociační studie binární fenotypová data – zda dané zvíře v průběhu života trpělo či netrpělo mastitidou. Data byla pro potřeby analýzy v programu TASSEL 5 (Bradbury *et al.* 2007) formátována do prostého textového souboru (.txt) následujícím způsobem):

<Phenotype>	
taxa data	
Taxa	Mastitis
1	1
2	1
3	1
4	0
5	0
6	1
7	0
8	1
9	0

Nadpis <Phenotype> je vyžadován, aby byl soubor identifikován jako fenotypový. Následující řádek obsahuje identifikátory povahy dat ve sloupcích pod nimi – může se jednat o data typu *taxa* (identifikační číslo či jméno jedinců, musí se shodovat s jejich označením v genotypovém souboru), *data* (numericky vyjádřený fenotyp), *covariate* (kovariát, tj. vysvětlující proměnná, jejíž vliv ale není předmětem zkoumání) nebo *factor* (seskupující proměnná, například rodinná příslušnost). Třetí sloupec obsahuje názvy proměnných a všechny nižší řádky již samotná data, v tomto případě číslo zvířete v prvním řádku a binární fenotyp ve druhém (0 značí mastitidy prostou krávu, 1 krávu, která mastitidou někdy trpěla).

4.2 Čištění dat

Primární čištění bylo provedeno v softwaru PLINK (Purcell *et al.* 2007). Hraniční hodnoty byly zvoleny na základě doporučení Marees *et al.* (2018) a kritérií použitých v jiných asociačních studiích podobného zaměření (Welderufael *et al.* 2018; Sodeland *et al.* 2010 aj.) následovně:

MAF (Minor Allele Frequency), četnost minoritní alely, byla stanovena na minimálně 5 %. Tato metoda asociační analýzy není vhodná pro zachycení vlivu vzácných alel, proto je žádoucí takové předem vyřadit.

Maximální přípustný podíl chybějících genotypových dat po individuálních SNP (Missingness Per SNP) byl nastaven na 10 %. Maximální přípustný podíl chybějících genotypových dat sledovaného jedince (Missingness Per Individual) byl nastaven na 10 %.

Příkaz pro realizaci výše uvedeného se zapisuje jako:

```
plink --file ../data --cow --geno 0.1 --maf 0.05 --mind 0.1  
--recode --out ../clean_data
```

kde:

--*plink* je obecným aktivačním příkazem programu PLINK.

--*file* indikuje cestu k datasetu (dvojice ze souborů .ped a soubor .map se stejnou přepnou).

--*cow* specifikuje, že se jedná o genom skotu (výchozí je lidský).

--*geno* určuje hodnotu filtru pro Missingness Per SNP.

--*mind* určuje hodnotu filtru pro Missingness Per Individual.

--*maf* určuje hodnotu filtru MAF.

--*recode* je pokyn pro naformátování očištěných dat do výstupního datasetu.

--*out* specifikuje cestu k výstupnímu souboru/datasetu.

Celkem bylo při tomto primárním čištění dat vyřazeno 1 661 SNP a 34 jedinců, k další analýze pokračovalo 51 557 SNP a 1 224 jedinců.

Následně byla provedena kontrola stratifikace populace pomocí multidimenzionálního škálování. Maximální přípustná odchylka stratifikační pozice jedince od průměru populace byla stanovena jako dvojnásobek směrodatné odchylky ($2\text{ SD} = \text{Standard Deviation}$).

V prvním kroku byl v PLINK připraven soubor s párovými IBS vzdálenostmi (koncovka `.genome`), v druhém kroku (na novém řádku) potom tabulka s čtyřdimenzionálními stratifikačními pozicemi (*Multidimensional Distribution Scaling*, MDS):

```
plink --file ../clean_data -cow --genome --out clean_data  
  
plink --file ../clean_data -cow --read-genome ../clean_data.genome  
--cluster --mds-plot 4 --out ../mds_result
```

kde:

`../clean_data` je cesta k souboru s očištěnými daty z předchozího kroku.

`--genome` je příkaz pro sestavení IBS matice.

`--read-genome` označuje adresu.

`--cluster` je příkaz pro sestavení shlukovací matice z IBS dat.

`--mds-plot` je příkaz pro mnohorozměrné škálování, číslo za příkazem specifikuje žádaný počet analyzovaných dimenzí, v tomto případě čtyři.

Výsledkem tohoto procesu je textový soubor s pozicemi jedinců v rámci sledovaných dimenzí (průměr je vždy 0):

FID	IID	SOL	C1	C2	C3	C4
CZE	1	0	-0,0122334	-0,0253612	0,00331427	-0,00683347
CZE	2	0	-0,0041215	-0,0348395	-0,00176654	0,00362091
CZE	3	0	-0,0148686	-0,0279226	-0,00740302	0,000703083
CZE	4	0	-0,00767081	-0,004864	-0,0189934	-0,0275164
CZE	5	0	-0,00684364	-0,0226117	-0,0140508	0,00643552
CZE	6	0	0,0139906	-0,0267655	-0,0127792	0,0112631
CZE	7	0	-0,00245785	-0,0200715	-0,0231544	0,0123148
CZE	8	0	-0,0117453	-0,0344446	-0,00392325	-0,00271828
CZE	9	0	-0,0084003	-0,0154251	0,0199151	0,0085606

kde:

FID značí identifikační označení rodiny, stáda či jiné skupiny (v této práci bez rozlišení)

IID značí individuální označení jedince.

SOL značí kód ze shlukovací matice, byl-li přidělen.

C1 značí pozici v první dimenzi.

C2 značí pozici v druhé dimenzi.

C3 značí pozici ve třetí dimenzi.

C4 značí pozici ve čtvrté dimenzi.

V závislosti na zvolených parametrech (zvolené citlivosti stratifikační analýzy) je možno posuzovat dvě až čtyři dimenze.

Výsledky je vhodné vizualizovat v podobě korelačního diagramu, což lze provést například pomocí R (R Core Team 2020):

```
data <- read.csv("../mds_result.mds", sep="")

x <- data$C1

y <- data$C2

plot(x, y, xlab = "Dimenze 1", ylab = "Dimenze 2", frame =
FALSE, pch = 19, col = c("hotpink4"))
```

kde:

data je tabulka obsahující sloupce (C1,...) s pozicemi v dimenzích, které jsou vykreslovány.

plot je příkaz pro sestavení grafu.

xlab specifikuje popisec osy x grafu.

ylab specifikuje popisec osy y grafu.

frame specifikuje orámování grafu (TRUE pro orámování, FALSE bez něj).

pch specifikuje symbol, který bude reprezentovat bod grafu (z galerie symbolů R).

col specifikuje barvu grafu (z galerie barev R).

V následujícím kroku byli z analýzy vyřazeni všichni jedinci, jejichž odchýlení od průměru populace, pozice 0, přesáhlo dvojnásobek směrodatné odchylky dané dimenze (viz Příloha 1, která obsahuje tabulku s dimenzemi, výpočet směrodatných odchylek a zvýraznění jedinců k vyřazení). Celkem se jednalo o 216 zvířat.

Identifikační čísla vyřazených zvířat byla zaznamenána do samostatného textového souboru obsahujícího sloupec s jejich FID a IID (viz výše), aby mohla být vyjmuta z analýzy pomocí:

```
plink --file ...\clean_data --cow --remove ...\excluded.txt --recode
--out ...\clean_data_without_excluded
```

kde:

...\clean_data je cesta k datasetu s daty očištěnými v předchozím kroku (.ped a .map).

...\excluded.txt je cesta k textovému souboru s identifikačními čísly jedinců k vyřazení.

--remove je příkaz k vyřazení daných jedinců.

Finální data pro asociační analýzu tedy čítala 51 557 SNP a 1 042 jedinců s průměrnou úspěšností genotypování > 99,88 %.

4.3 Stanovení hladiny významnosti

Vzhledem k tomu, že asociační analýza je ve svém základu mnohonásobným testováním, je třeba náležitě upravit hladinu významnosti. Výchozí hladina významnosti pro biologická data je všeobecně přijímána jako 0,05. Počet jednotlivých testů v asociační analýze odpovídá počtu SNP, tedy 51 557.

Hladinu významnosti lze Bonferroniho korekcí stanovit jako:

$$\alpha = \frac{0,05}{51\,557} = \mathbf{0,0000097}$$

$$-\log_{10}(\alpha) = \mathbf{6,0134}$$

Dále byla na základě předchozích studií (Welderufael *et al.* 2018; Banos *et al.* 2017 aj.) přijata ještě benevolentnější sugestivní prahová hodnota, při jejímž překonání byly SNP přijaty k bližšímu prozkoumání, přestože nedosáhly celogenomové významnosti. Tato hodnota byla stanovena jako $1E-04 = \mathbf{0,0001}$, $-\log_{10}(0,0001) = \mathbf{4}$.

4.4 Asociační analýza

Vlastní asociační analýza byla realizována ve volně dostupném programu TASSEL 5 (Bradbury *et al.* 2007) s využitím obecného lineárního modelu (General Linear Model, GLM) (Rawlings *et al.* 1998):

$$\mathbf{Y} = \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

kde \mathbf{Y} je matice s vícerozměrnými měřeními (jedna závislá proměnná na sloupec), \mathbf{X} je matice s nezávislými proměnnými (každý sloupec je výčet pozorování na jednu nezávislou proměnnou), $\boldsymbol{\beta}$ je matice s odhadnutými parametry (tj. jak silný vliv na závislou proměnnou má nezávislá proměnná) a $\boldsymbol{\varepsilon}$ je matice s náhodnými chybami (nekorelované, v normálním rozdělení). Závislou proměnnou je v tomto případě fenotyp, nezávislou je genotyp, tedy data o jednotlivých SNP.

Program TASSEL 5 byl zvolen mimo jiné proto, že kromě standardních výsledků se sílami asociací poskytuje i odhad efektů jednotlivých alelických sestav. Asociační analýza pomocí obecného lineárního modelu byla provedena následujícími příkazy v příkazovém řádku prostředí Windows:

```
./start_tassel.bat  
  
./run_pipeline.bat -fork1 -plink -ped <clean_data_without_excluded.ped> -  
map <clean_data_without_excluded.map> -endPlugin -fork2 -r <phenotype.txt>  
-combine5 -input1 -input2 -intersect -FixedEffectLMPlugin -endPlugin -  
export association_result
```


kde:

./start_tassel.bat je příkaz pro vyvolání programu TASSEL 5.

./run_pipeline.bat je příkaz pro zahájení sady instrukcí („pipeline“).

-fork je vlajka specifikující segment sady instrukcí, které mají být provedeny postupně.

-plink je příkaz pro aktivaci pluginu pro import genotypových dat ve formátu PLINK.

-ped specifikuje importovaný soubor *.ped*.

-map specifikuje importovaný soubor *.map*.

-endPlugin je příkaz pro ukončení pluginu.

-r je příkaz pro import textového souboru s fenotypem (viz bod 4.1).

-combine je příkaz pro sjednocení genotypového (zde jako *input1*) a fenotypového (zde jako *input2*) do jednoho datasetu pro GLM.

-intersect je příkaz pro vytvoření průniku genotypu a fenotypu na základě oddílu *taxa* (identifikačních čísel/jmen jedinců, odstraní se tak z fenotypové části datasetu jedinci, kteří byli v předchozích krocích vyřazeni z genotypového).

-FixedEffectLMPPlugin je příkaz pro vyvolání pluginu, který na datasetu uplatní GLM.

-export je příkaz pro exportování textového souboru s výsledky asociační analýzy.

Výsledky asociační analýzy (Příloha 2) byly seřazeny podle p-hodnoty od nejnižší po nejvyšší (respektive podle $-\log_{10}(p)$ od nejvyššího po nejnižší) a dále naformátovány pro grafické znázornění (viz níže). Na základě vygenerovaného odhadu efektů alel (Příloha 3) byla sestavena tabulka s nejvýhodnějšími alelickými sestavami a výhodnějšími alelami.

4.5 Grafické znázornění výsledků

4.5.1 Manhattan graf

Pro grafické znázornění výsledků byla použita volně dostupná platforma R (R Core Team 2020) s balíčkem qqman (Turner 2018). Po instalaci balíčku byl výsledný Manhattan graf získán pomocí následujících příkazů:

```
data <- read.delim2("../data.txt")

library(qqman)

manhattan(data, col = c("hotpink4", "palevioletred2"),
suggestiveline = -log10(1E-04), genomewideline = 6.01331)
```

kde:

read.delim2 je příkaz pro načtení daného textového souboru s tabulkou (hodnoty oddělené tabulátory).

library je příkaz pro vyvolání specifického balíčku funkcí, v tomto případě qqman.

manhattan je příkaz balíčku qqman pro sestavení Manhattan grafu.

data je tabulka obsahující sloupce nadepsané BP (pozice SNP na chromozomu), CHR (příslušná čísla chromozomů) a P (příslušné naměřené p-hodnoty).

col je příkaz pro specifikaci barvy grafu (z galerie R).

suggestiveline = příkaz specifikující úroveň sugestivní prahové hodnoty (viz bod 4.3).

genomewideline = příkaz specifikující celogenomovou hladinu významnosti (viz bod 4.3).

4.5.2 Kvantil-quantil (Q-Q) graf

Následně byl v rámci kontroly kvality analýzy vytvořen kvantil-quantil graf:

```
data <- read.delim2("../data.txt")

library(qqman)

qq(data$P, col = c("hotpink4"))
```

kde:

data je tabulka obsahující sloupec nazvaný P (naměřené p-hodnoty).

qq je příkaz pro sestavení Q-Q grafu, první položka za závorkou specifikuje dataset (*data*) a sloupec (*P*), z nichž se má vycházet.

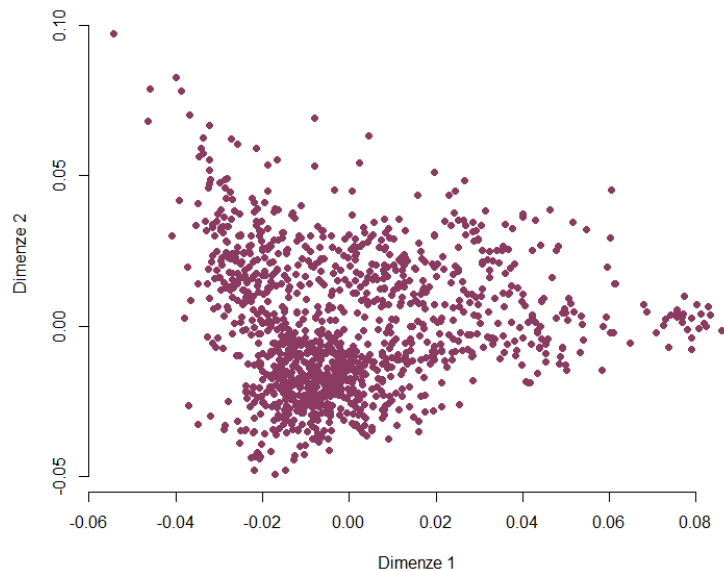
col je znakový vektor specifikující barvy grafu (z galerie barev R).

4.6 Anotace nejvýznamnějších SNP

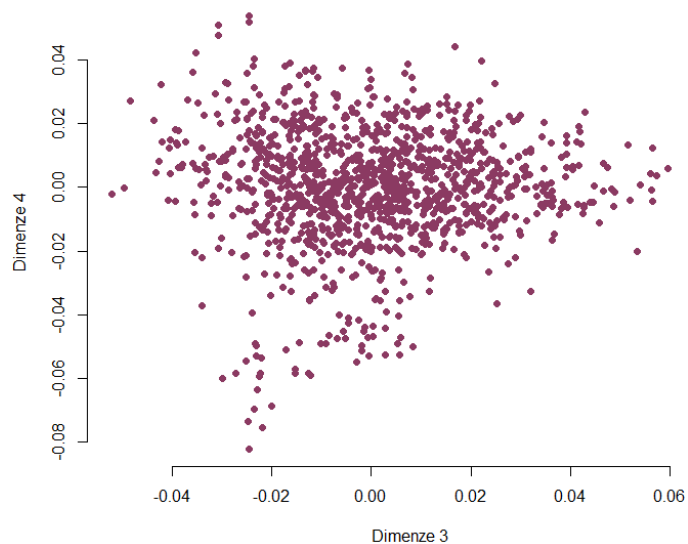
Polymorfismy s nejsilnější naměřenou asociací byly blíže prozkoumány za použití mezinárodních genomických databází, předně databáze Ensembl (Yates *et al.* 2020). Absolutní pořadí báze, kde se daný SNP nachází, se mění v závislosti na konkrétní použitém referenčním genomu (různí jedinci se v délce genomu mírně liší, mají tedy pozice genů různě posunuté). Z tohoto důvodu byla umístění SNP identifikována nikoliv podle bodových pozic (v bp) udávaných výrobcem DNA čipu, ale podle sekvencí jejich lokusů (lze dohledat v dokumentaci výrobce či vyžádat). Tyto sekvence byly zasazeny do kontextu referenčního genomu *Bos_taurus ARS-UCDI.2* v databázi Ensembl pomocí nástroje BLAST (*Basic Local Alignment Search Tool*). Na základě výsledků vyhledávání byla posouzena jejich relativní pozice vůči nejbližším genům.

5 Výsledky

5.1 Výsledky stratifikační analýzy

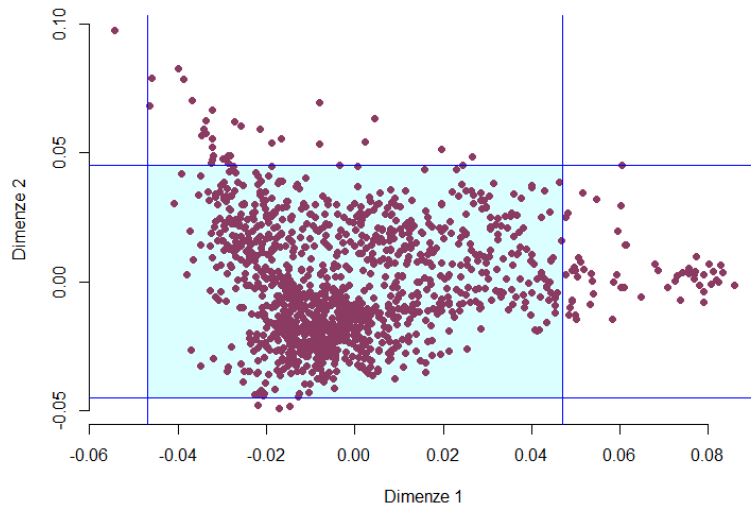


Obrázek 16: Bodový graf první a druhé dimenze MDS pro analýzu populační stratifikace sledovaného souboru.

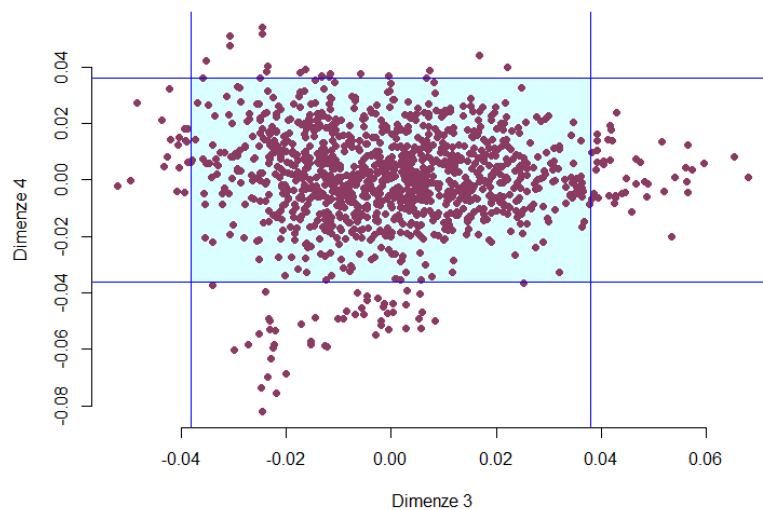


Obrázek 17: Bodový graf třetí a čtvrté dimenze MDS pro analýzu populační stratifikace sledovaného souboru.

Na Obrázcích 16 a 17 je patrný výskyt odlehlých jedinců (*outliers*), na kladné straně Dimenze 1 dokonce drobného shluku. Pro další analýzu byli přijati pouze jedinci, kteří se v žádné z dimenzí nevytékali průměru populace o více než dvě směrodatné odchylky (Obrázek 18 a 19):



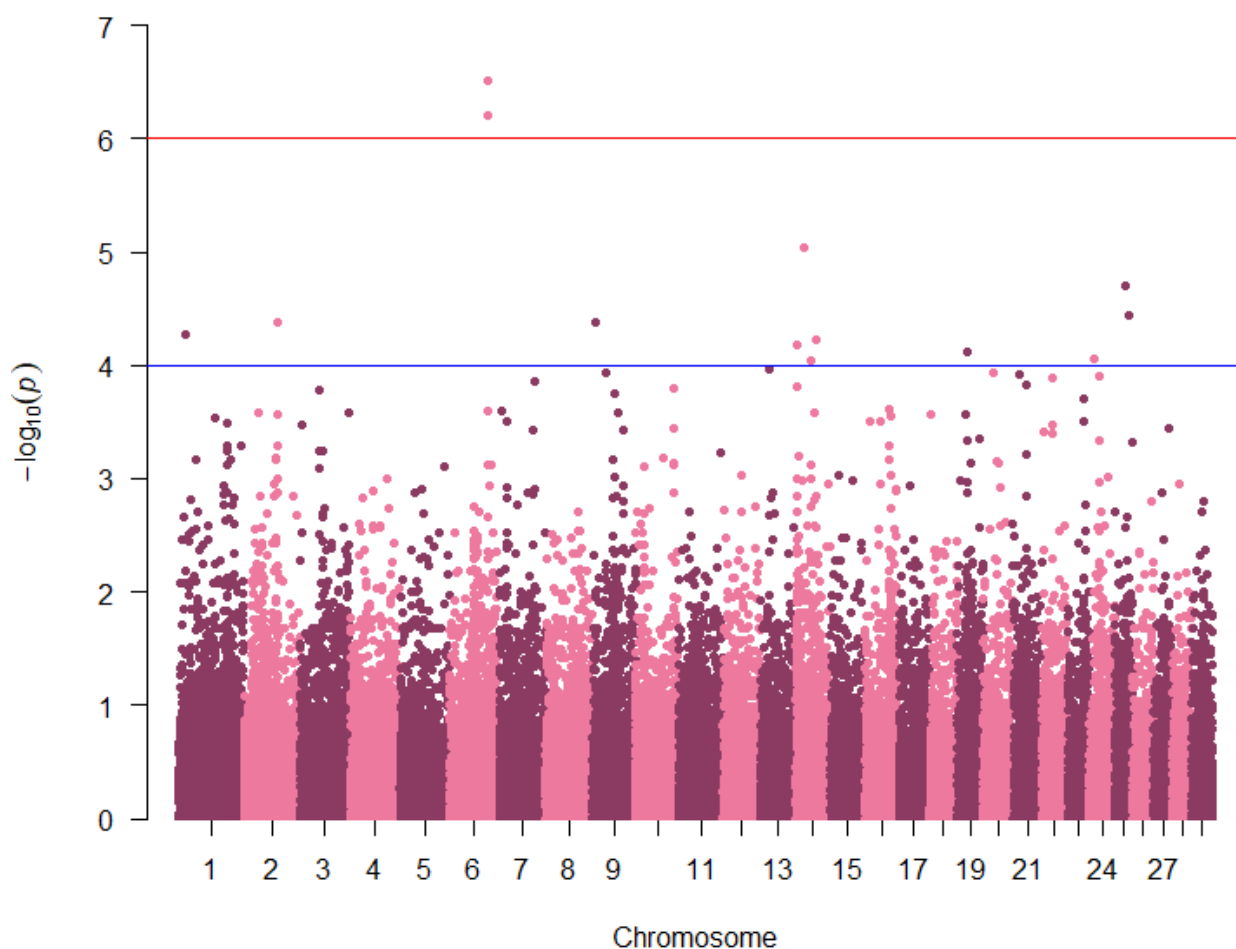
Obrázek 18: Bodový graf první a druhé dimenze MDS s bleděmodře označenou množinou jedinců, kteří byli přijati k další analýze. Modré linky reprezentují dvojnásobky směrodatných odchylek.



Obrázek 19: Bodový graf třetí a čtvrté dimenze MDS s bleděmodře označenou množinou jedinců, kteří byli přijati k další analýze. Modré linky reprezentují dvojnásobky směrodatných odchylek.

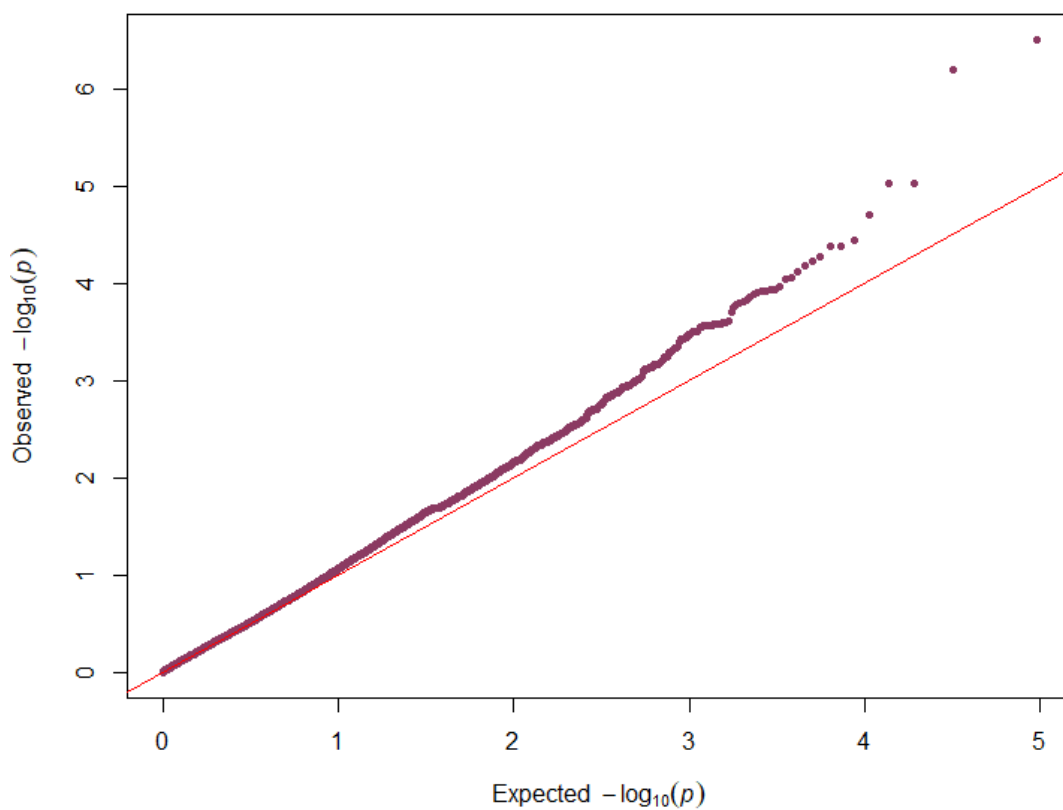
5.2 Souhrn výsledků asociační analýzy

Z výsledného Manhattan grafu (Obrázek 20) je patrné, že nejméně dva SNP přesáhly celogenomovou hladinu významnosti a nejméně jedenáct dalších sugestivní prahovou hodnotu:



Obrázek 20: Manhattan graf zachycující $-\log_{10}(p)$ jednotlivých sledovaných SNP (svíslá osa) a jejich pozici na chromozomech (vodorovná osa) Červená vodorovná přímka značí celogenomovou hladinu významnosti, modrá sugestivní prahovou hodnotu.

Průběh kvantil-kvantil grafu kontroly distribuce $-\log_{10}(p)$ (Obrázek 21) byl přijat jako uspokojivý – většinu průběhu se blíží ose normálního rozdělení a vymyká se až u nejsilnějších asociací na konci.



Obrázek 21: Kvantil-kvantil (Q-Q) graf, svislá osa zachycuje pozorovanou $-\log_{10}(p)$ hodnotu, vodorovná osa očekávanou $-\log_{10}(p)$ hodnotu při normálním rozdělení.

5.3 Významné SNP

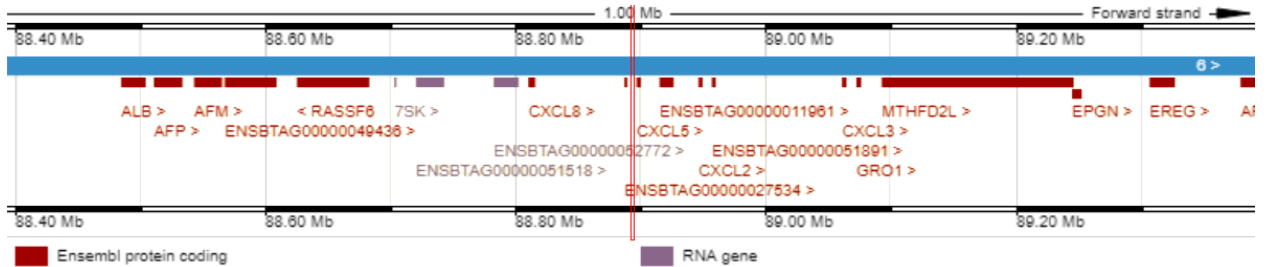
Čtyřicet polymorfismů s nejsilnějším vztahem k odolnosti proti mastitidě je zaznamenáno v následující tabulce. Fialově jsou zvýrazněny SNP dosahující celogenomové významnosti a tučně SNP dosahující sugestivní významnosti – jejich pozice je níže zasazena do kontextu genomu.

Název SNP	Chromozom	Pozice (bp)	p-hodnota	$-\log_{10}(p)$
BTB-00265951	6	90 642 682	3,12E-07	6,5058
BTA-121769-no-rs	6	89 912 196	6,25E-07	6,204
BTB-01283269	14	20 196 617	9,25E-06	5,034
BTA-36062-no-rs	14	20 216 601	9,25E-06	5,034
ARS-BFGL-NGS-109683	25	23 825 593	1,99E-05	4,7014
ARS-BFGL-NGS-37189	25	32 912 638	3,65E-05	4,4374
ARS-BFGL-NGS-104096	9	7 010 227	4,11E-05	4,386
Hapmap43710-BTA-86183	2	79 846 105	4,12E-05	4,3851
BTA-28028-no-rs	1	14 551 165	5,40E-05	4,2678
ARS-BFGL-BAC-23887	14	49 905 143	5,98E-05	4,2234
ARS-BFGL-NGS-115947	14	5 494 654	6,50E-05	4,1869
ARS-BFGL-NGS-1837	19	26 556 279	7,61E-05	4,1185
Hapmap50627-BTA-23969	24	13 582 752	8,84E-05	4,0533
UA-IFASA-7226	14	36 276 655	8,97E-05	4,0473
ARS-BFGL-NGS-109707	13	19 167 543	1,08E-04	3,9663
Hapmap38180-BTA-83144	9	32 484 636	1,14E-04	3,943095
ARS-BFGL-NGS-102895	20	24 228 836	1,15E-04	3,939302
Hapmap31201-BTA-162197	21	16 172 731	1,22E-04	3,91364

Hapmap43878-BTA-53506	21	16 272 399	1,22E-04	3,91364
ARS-BFGL-NGS-99694	24	25 861 246	1,23E-04	3,910095
BTA-53899-no-rs	22	24 689 372	1,31E-04	3,882729
BTB-01455682	7	87 625 506	1,39E-04	3,856985
ARS-BFGL-NGS-24846	21	34 037 563	1,47E-04	3,832683
UA-IFASA-6329	14	5 096 531	1,53E-04	3,815309
Hapmap54410-rs29021343	10	92 034 186	1,59E-04	3,798603
Hapmap50807-BTA-86717	3	47 407 040	1,67E-04	3,777284
BTB-01215320	9	53 313 101	1,75E-04	3,756962
BovineHD2300011339	23	39 241 545	1,96E-04	3,707744
BTB-00649471	16	56 489 594	2,41E-04	3,617983
UA-IFASA-6709	6	93 381 472	2,50E-04	3,60206
BTB-00292673	7	4 953 801	2,56E-04	3,59176
ARS-BFGL-NGS-44341	2	33 944 781	2,58E-04	3,58838
ARS-BFGL-NGS-7542	3	118 963 574	2,61E-04	3,583359
ARS-BFGL-NGS-35676	9	64 109 435	2,63E-04	3,580044
BTB-01709766	14	47 337 155	2,64E-04	3,578396
ARS-BFGL-NGS-34770	19	21 145 204	2,67E-04	3,573489
ARS-BFGL-NGS-35513	18	4 397 257	2,67E-04	3,573489
ARS-BFGL-NGS-94630	2	79 893 482	2,73E-04	3,563837
Hapmap35064-BES1_Contig468_700	2	79 923 716	2,73E-04	3,563837

5.3.1 BTB-00265951

Sekvence oblasti obsahující SNP s označením *BTB-00265951* byla vyhledána v databázi Ensembl (Yates *et al.* 2020) na chromozomu 6:

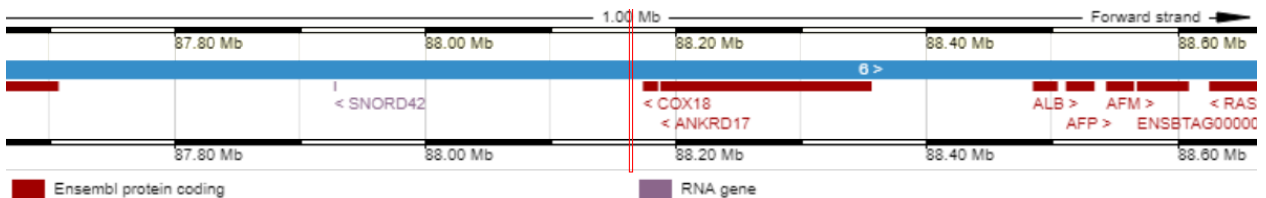


Obrázek 22: Pozice SNP *BTB-00265951* (vertikální červená linka) (*Mb* je zkratka pro megabázi = 1 000 000 bp).

Je patrné, že se nachází mimo sekvenci známých genů, ale v rámci oblasti s vysokou koncentrací genů. Nejbližšími protein kódujícími geny jsou *CXCL2*, *CXCL5*, *CXCL8*, *ENSBTAG00000027534* a *ENSBTAG00000011961*.

5.3.2 BTA-121769-no-rs

Sekvence oblasti obsahující SNP s označením *BTA-121769-no-rs* byla vyhledána v databázi Ensembl (Yates *et al.* 2020) na chromozomu 6:

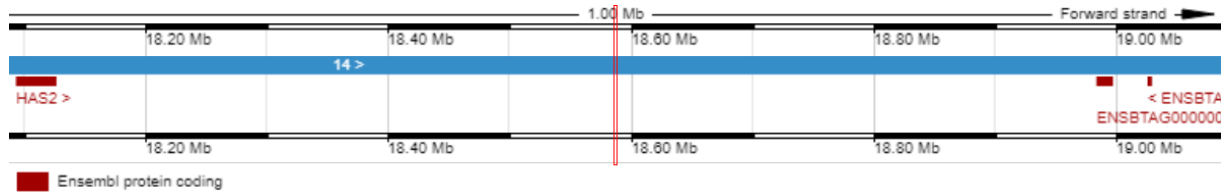


Obrázek 23: Pozice SNP *BTA-121769-no-rs* (vertikální červená linka).

Je patrné, že se nachází mimo sekvenci známých genů. Nejbližšími protein kódujícími geny jsou *COX18* a *ANKRD17*.

5.3.3 BTB-01283269

Sekvence oblasti obsahující SNP s označením *BTB-01283269* byla vyhledána v databázi Ensembl (Yates *et al.* 2020) na chromozomu 14:

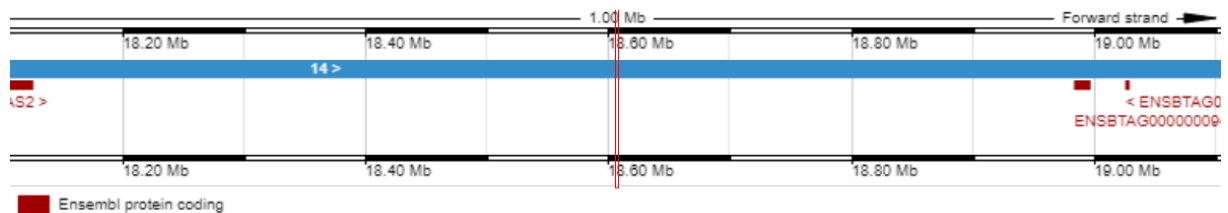


Obrázek 24: Pozice SNP *BTB-01283269* (vertikální červená linka).

Je patrné, že se nachází mimo sekvenci známých genů a žádné se nenachází ani v jeho blízkosti.

5.3.4 BTA-36062-no-rs

Sekvence oblasti obsahující SNP s označením *BTA-36062-no-rs* byla vyhledána v databázi Ensembl (Yates *et al.* 2020) na chromozomu 14:

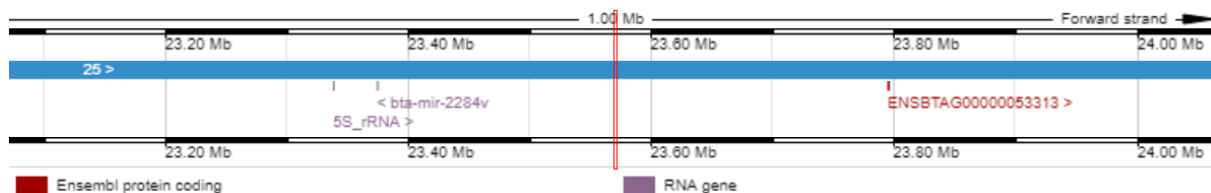


Obrázek 25: Pozice SNP *BTA-36062-no-rs* (vertikální červená linka).

Je patrné, že se nachází mimo sekvenci známých genů a žádné neleží ani v jeho blízkosti.

5.3.5 ARS-BFGL-NGS-109683

Sekvence oblasti obsahující SNP s označením *ARS-BFGL-NGS-109683* byla vyhledána v databázi Ensembl (Yates *et al.* 2020) na chromozomu 25:

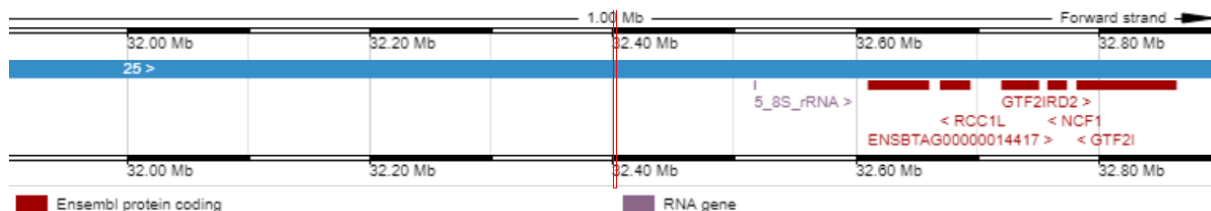


Obrázek 26: Pozice SNP ARS-BFGL-NGS-109683 (vertikální červená linka).

Je patrné, že se nachází mimo oblast známých genů. Nejbližším protein kódujícím genem je *ENSBTAG00000053313*, nejbližším funkční RNA kódujícím genem je *bta-mir-2284v*.

5.3.6 ARS-BFGL-NGS-37189

Sekvence oblasti obsahující SNP s označením *ARS-BFGL-NGS-37189* byla vyhledána v databázi Ensembl (Yates *et al.* 2020) na chromozomu 25:

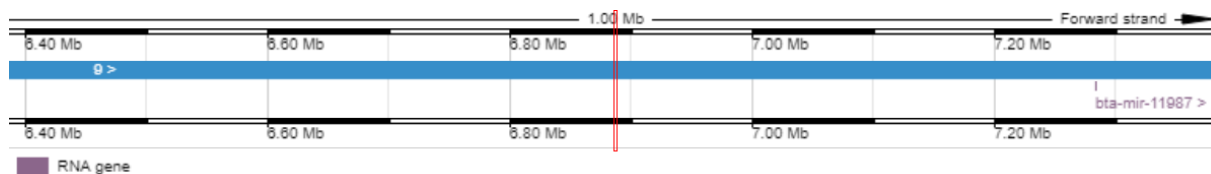


Obrázek 27: Pozice SNP ARS-BFGL-NGS-37189 (vertikální červená linka).

Je patrné, že se nachází mimo oblast známých genů. Nejbližším genem je gen pro funkční RNA *58S_rRNA*, nejbližšími protein kódujícími geny jsou *ENSBTA00000014417* a *RCC1L*.

5.3.7 ARS-BFGL-NGS-104096

Sekvence oblasti obsahující SNP s označením *ARS-BFGL-NGS-104096* byla vyhledána v databázi Ensembl (Yates *et al.* 2020) na chromozomu 9:

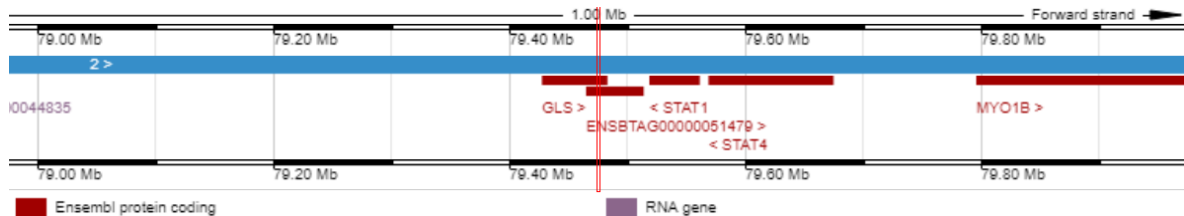


Obrázek 28: Pozice SNP ARS-BFGL-NGS-104096 (vertikální červená linka).

Je patrné, že se nachází mimo sekvenci známých genů a žádné neleží ani v jeho blízkosti.

5.3.8 Hapmap43710-BTA-86183

Sekvence oblasti obsahující SNP s označením *Hapmap43710-BTA-86183* byla vyhledána v databázi Ensembl (Yates *et al.* 2020) na chromozomu 2:

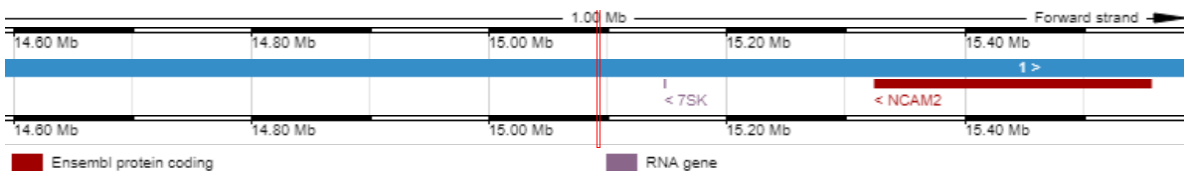


Obrázek 29: Pozice SNP *Hapmap43710-BTA-86183* (vertikální červená linka).

Je patrné, že se nachází na průniku hned dvou genů (oba na vedoucím vlákně, začátek druhého tudíž leží v rámci sekvence prvního), *GLS* a *ENSBTAG00000051479*.

5.3.9 BTA-28028-no-rs

Sekvence oblasti obsahující SNP s označením *BTA-28028-no-rs* byla vyhledána v databázi Ensembl (Yates *et al.* 2020) na chromozomu 1:

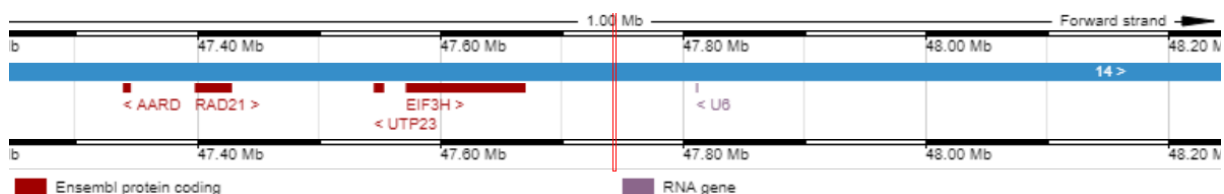


Obrázek 30: Pozice SNP *BTA-28028-no-rs* (vertikální červená linka).

Je patrné, že se nachází mimo sekvenci známých genů. Nejbližšími protein kódujícím genem je *NCAM2*, poblíž se nachází také funkční RNA kódující gen *7SK*.

5.3.10 ARS-BFGL-BAC-23887

Sekvence oblasti obsahující SNP s označením *ARS-BFGL-BAC-23887* byla vyhledána v databázi Ensembl (Yates *et al.* 2020) na chromozomu 14:

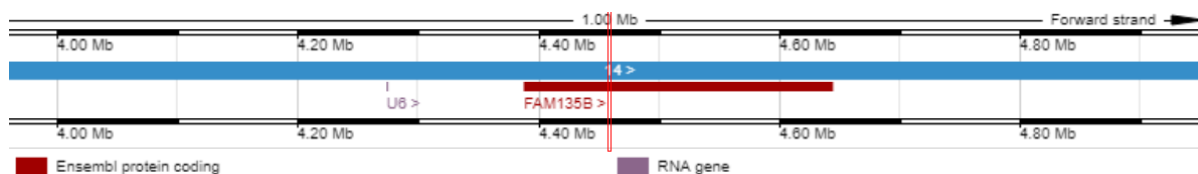


Obrázek 31: Pozice SNP ARS-BFGL-BAC-23887 (vertikální červená linka).

Je patrné, že se nachází mimo sekvenci známých genů. Nejbližšími protein kódujícími geny jsou *EIF3H* a *UTP23*, poblíž se nachází také funkční RNA kódující gen *U6*.

5.3.11 ARS-BFGL-NGS-115947

Sekvence oblasti obsahující SNP s označením *ARS-BFGL-NGS-115947* byla vyhledána v databázi Ensembl (Yates *et al.* 2020) na chromozomu 14:

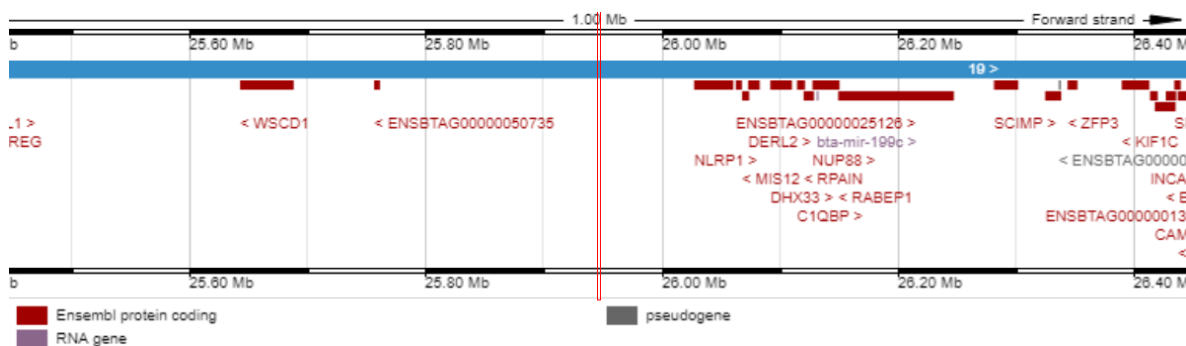


Obrázek 32: Pozice SNP ARS-BFGL-NGS-115947 (vertikální červená linka).

Je patrné, že se nachází na průniku s genem *FAM135B*.

5.3.12 ARS-BFGL-NGS-1837

Sekvence oblasti obsahující SNP s označením *ARS-BFGL-NGS-1837* byla vyhledána v databázi Ensembl (Yates *et al.* 2020) na chromozomu 19:

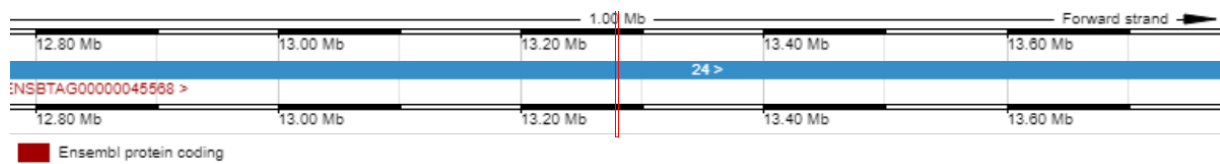


Obrázek 33: Pozice SNP ARS-BFGL-NGS-1837 (vertikální červená linka).

Je patrné, že se nachází mimo sekvenci známých genů, leží však nedaleko genového shluku s nejbližšími protein kódujícími geny *NLRP1*, *ENSBTAG00000025126* a *DERL2*. Z druhé strany leží nejbliže gen *ENSBTAG00000050735*.

5.3.13 Hapmap50627-BTA-23969

Sekvence oblasti obsahující SNP s označením *Hapmap50627-BTA-23969* byla vyhledána v databázi Ensembl (Yates *et al.* 2020) na chromozomu 24:

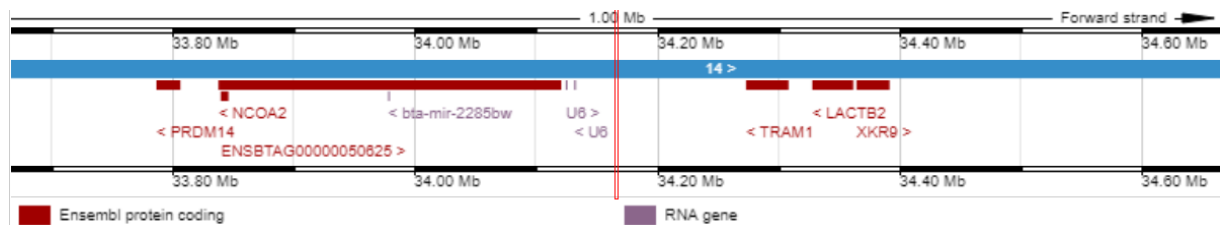


Obrázek 34: Pozice SNP *Hapmap50627-BTA-23969* (vertikální červená linka).

Je patrné, že se nachází mimo sekvenci známých genů a žádné neleží ani v jeho blízkosti.

5.3.14 UA-IFASA-7226

Sekvence oblasti obsahující SNP s označením *UA-IFASA-7226* byla vyhledána v databázi Ensembl (Yates *et al.* 2020) na chromozomu 14:



Obrázek 35: Pozice SNP *UA-IFASA-7226* (vertikální červená linka).

Je patrné, že se nachází mimo sekvenci známých genů, leží však mezi dvěma genovými shluky s nejbližšími protein kódujícími geny *TRAM1* a *NCOA2*, nejbližšími funkční RNA kódujícími geny *U6*.

5.4 Žádoucí alelické sestavy významných SNP

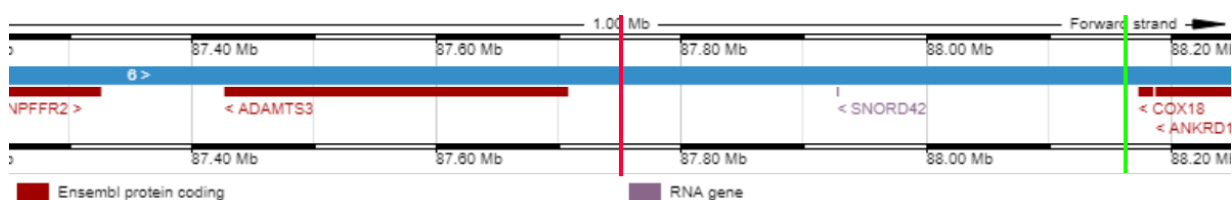
V následující tabulce jsou zaznamenány možné alely nejvýznamnějších polymorfismů a alelické sestavy, které byly spojeny se zvýšenou odolností vůči mastitidě. Žádoucí sestavu lze uplatnit pouze v rámci jednoho plemene, v tomto případě holštýnského.

Název SNP	Alela 1	Alela 2	Nejvýhodnější sestava	Výhodnější alela
BTB-00265951	T	C	C/C	C
BTA-121769-no-rs	A	G	G/G	G
BTB-01283269	A	G	A/A	A
BTA-36062-no-rs	T	C	T/T	T
ARS-BFGL-NGS-109683	T	C	C/C	C
ARS-BFGL-NGS-37189	A	C	A/C	C
ARS-BFGL-NGS-104096	A	C	C/C	C
Hapmap43710-BTA-86183	A	G	A/A	A
BTA-28028-no-rs	T	C	T/T	T
ARS-BFGL-BAC-23887	T	C	C/C	C
ARS-BFGL-NGS-115947	T	C	C/C	C
ARS-BFGL-NGS-1837	T	C	C/C	C
Hapmap50627-BTA-23969	A	G	A/G	G
UA-IFASA-7226	T	C	C/C	C

6 Diskuse

6.1 Metody a významné SNP v předchozích asociačních analýzách

Sodeland *et al.* (2010) použili pro asociační analýzu genotypová data (17 349 SNP) plemenných býků ($n = 2\,589$) a rezistenci vůči mastitidě posuzovali podle jejího výskytu u dcer ($n = 1\,389\,776$). Uvádí, že statisticky významnou souvislost s prevalencí mastitidy detekovali u celkem 29 SNP, je však třeba podotknout, že analýzu provedli pomocí smíšeného lineárního modelu, pro který přijali hladinu významnosti $p < 0,0001$, což odpovídá $-\log_{10}(p) = 4$, ale za použití Bonferroniho korekce by při 17 349 SNP odpovídala $p < 0,0000029$, $-\log_{10}(p) = 5,54$. Velmi silnou asociaci nicméně detekovali na chromozomu 6 (SNP *BTA-119376*), relativně nedaleko pozice *BTA-121769-no-rs*, který zaznamenala jako významný tato práce:



Obrázek 36: Pozice *BTA-119376* (vertikální červená linka) a *BTA-121769-no-rs* (vertikální zelená linka).

Pozorovali také několik asociací na chromozomu 14, ač poměrně vzdálených od SNP zaznamenaných v této práci (> 1 Mb). Dále pozorovali významné asociace na chromozomech 4 a 20, kde tato práce nedetekovala žádnou.

Tiezzi *et al.* (2014) studovali data býků na základě fenotypu jejich dcer, konkrétně 1 361 býků a 103 585 dcer na první laktaci. Pracovali s daty z Illumina BovineSNP50 BeadChip (Illumina, Inc., San Diego, USA) stejně jako tato práce. Po očištění dat měli k dispozici pro asociační analýzu soubor 39 004 SNP, měli však k dispozici také rodokmenová data, stádovou příslušnost a nebinární fenotyp (plemenné hodnoty dcer), použili tudíž analýzu pomocí smíšeného lineárního modelu (MLM). Nejsilnější asociace detekovali na chromozomu 14, opět velmi vzdálené od SNP zaznamenaných v této práci (> 1 Mb), a o něco slabší na chromozomech 2, 8, 11 a 24.

Sahana *et al.* (2014) taktéž používali genotypová data plemenných býků a fenotypová data o plemenných hodnotách jejich dcer – celkem 10 560 býků, z toho 4 992 holštýnského plemene. Používali převážně data z Illumina BovineSNP50 BeadChip (Illumina, Inc., San Diego, USA), ale pro část býků ($n = 557$) byla dostupná také širší data z Illumina BovineHD Genotyping BeadChip (777 000 SNP) a pro 89 býků byly vypracovány celogenomové sekvence. Oblast nejsilnějších asociací pozorovali na chromozomu 6 – v rámci ní se nachází oba SNP, které v této práci dosáhly celogenomové významnosti (*BTA-121769-no-rs* i *BTB-00265951*).

Fang *et al.* (2017) ve své asociační studii mléčné produkce a rezistence vůči mastitidě kombinovali analýzu lineárním smíšeným modelem (blíže popsán v Kang *et al.* 2010) s analýzou exprese pomocí technologie RNA-Seq, nedetekovali pro mastitidu žádné SNP s významnou asociací. Pro GWAS u holštýnského skotu používali data o celkem 15 355 382 SNP získaná od 5 056 krav. Hladinu významnosti stanovili Bonferroniho korekcí na $p = 3,3 \times 10^{-9}$, což odpovídá $-\log_{10}(p) = 8,48$.

Welderufael *et al.* (2018) analyzovali data o 39 378 SNP od 993 holštýnských krav. Pro výskyt mastitidy nezaznamenali žádnou asociaci dosahující hladiny významnosti po Bonferroniho korekci ($-\log_{10}(p) = 5,90$). Používali podobně jako tato práce binární fenotypová data, genotypy z Illumina BovineSNP50 BeadChip (Illumina, Inc., San Diego, USA), software PLINK (Purcell *et al.* 2007), obecný lineární model (GLM) a anotaci prostřednictvím databáze Ensembl (Yates *et al.* 2020), ale navíc se zabývali také vývojem stavu (zotavitelností) zvířat, která mastitidou onemocněla.

Za povšimnutí stojí také recentní asociační studie Korcuć *et al.* (2021), ačkoliv se nezabývala přímo mastitidami – doживost, kterou sledovali, s nimi úzce souvisí. Na této studii je zajímavý aspekt rozdělení sledovaných zvířat do skupin podle fáze a pořadí laktace. Zjistili totiž odlišné asociované SNP pro jednotlivé kategorie, přičemž právě odolnost vůči mastitidě by s tímto jevem mohla souviset. Tomu napovídá i výsledek studie, který u období se zvýšeným výskytem mastitid (zejména na začátku laktace) vykazuje vyšší sílu asociace pro SNP na chromozomu 6 (ve stejné oblasti jako tato a některé výše zmíněné práce).

Z diskuse metod a výsledků vyplývá, že ačkoliv významných SNP nebylo touto prací identifikováno mnoho, její výsledky se potkávají se závěry předchozích asociačních studií rezistence vůči mastitidě v detekci asociované oblasti na chromozomu BTA6.

6.2 Funkce genů v blízkosti významných SNP

Jak je patrné z úryvků genomové mapy, oba celogenomově významné polymorfismy se nacházejí blízko (vzdálenost < 0,1 Mb) nejméně jednoho genu.

BTB-00265951 se nachází uvnitř shluku genů, z nichž řada patří ke skupině *CXCL*, což je zkratka pro C-X-C motiv obsahující chemokinový ligand (Chen *et al.* 2020). Proteiny, které kóduje, slouží při infekci na plazmatické membráně jako chemoatraktanty pro buňky imunitního systému, zejména neutrofilů (Kraan *et al.* 2001) (Jeyaseelan *et al.* 2005). Gilbert *et al.* (2013) se zabývali studiem exprese vybraných proteinů při bakteriální mastitidě u skotu a pozorovali intenzivní expresi několika genů z rodiny *CXCL*, mezi nimi i *CXCL5*, který je polymorfismu *BTB-00265951* nejbližší. Tyto geny jsou tedy přímo spojeny s imunitní odpovědí, což nadále zvyšuje pravděpodobnost vlivu detekovaného SNP na rezistenci vůči mastitidě.

Gen nejbližší *BTB-00265951* v opačném směru je v databázi Ensembl (Yates *et al.* 2020) veden pod názvem *ENSBTAG00000027534*. Protein, který kóduje, se podle této databáze nazývá cytosolická 5'-nukleotidáza 3A. U skotu zřejmě zatím nebyla zkoumána, u člověka byla popsána jako enzym katalyzující defosforylaci nukleosidových 5'-monofosfátů (Aksoy *et al.* 2009). Je klíčová pro udržování stavu nukleotidů v buňce – žádná role v imunitní odpovědi dosud nebyla popsána.

Druhý nejvýznamnější polymorfismus, *BTA-121769-no-rs*, má ve své bezprostřední blízkosti gen *COX18* a *ANKRD17*. Gen *COX18*, který je *BTA-121769-no-rs* nejbližší, kóduje faktor pro sestavení enzymu cytochrom c oxidázy (*cytochrome c oxidase assembly factor*). Je nezbytný také pro integraci transmembránových proteinů do membrány (Bourens & Barrientos 2017), což je proces, jehož narušení by mohlo vést ke zhoršení odolnosti, které bylo zaznamenáno v této práci.

Gen *ANKRD17* kóduje protein 17 s ankyrinovou repetiční doménou (*ankyrin repeat domain protein 17*) a byl u něj pozorován především význam pro diferenciaci hepatocytů (Jones

et al. 2002), ale Menning a Kufer (2013) identifikovali i roli v kaskádách imunitní reakce proti bakteriální infekci – taktéž proces, jehož narušení by mohlo snížit odolnost vůči mastitidě.

Diskuse genů nalezených v oblastech asociovaných polymorfismů ukazuje na vysokou pravděpodobnost jejich zapojení do imunitní reakce, a tudíž i vlivu na náchylnost vůči mastitidě, což je v souladu se závěry této práce.

7 Závěr

V rámci literárního přehledu byla sestavena rešerše českých i mezinárodních odborných publikací seznamující s problematikou zánětlivých onemocnění mléčné žlázy (mastitid), jejich etiologií, epidemiologií a závažnými ekonomickými dopady v chovu mléčného skotu. Následně byly prezentovány aktuální poznatky o možném genetickém založení odolnosti vůči těmto zánětům a bioinformatické metody vhodné k jeho studiu.

V experimentální části práce byla vstupní data o 53 218 SNP od celkem 1 258 dojených krav holštýnského plemene zpracována výše zdokumentovaným způsobem a podrobena asociační analýze.

Z výsledků asociační analýzy vyplývá, že mezi celogenomově významné faktory pro odolnost vůči mastitidě u skotu patří polymorfismy *BTA-121769-no-rs* a *BTB-00265951* na chromozomu *BTA6*. Při diskusi s dosavadními studiemi stejného zaměření byla zaznamenána omezená shoda, pro praktickou aplikaci těchto poznatků ve šlechtitelství je tudíž vhodné bližší studium asociované oblasti. Pro tento účel byly identifikovány nejbližší ležící geny a jejich funkce v rámci imunitní reakce.

Úvodní hypotéza, že asociační studie na základě SNP čipů umožní detekovat oblasti genomu skotu zodpovědné za výskyt mastitidy, byla přijata jako platná.

8 Seznam použité literatury

Abe T, Saburi J, Hasebe H, Nakagawa T, Kawamura T, Saito K, Nade T, Misumi S, Okumura T, Kuchida K, Hayashi T, Nakane S, Mitsuhasi T, Nirasawa K, Sugimoto Y, Kobayashi E. 2008. Bovine quantitative trait loci analysis for growth, carcass, and meat quality traits in an F2 population from a cross between Japanese Black and Limousin. *Journal of Animal Science* **86**(11):2821-32.

Abebe R, Hatiya H, Abera M, Megersa B, Asmare K. 2016. Bovine mastitis: prevalence, risk factors and isolation of *Staphylococcus aureus* in dairy herds at Hawassa milk shed, South Ethiopia. *BMC Veterinary Research* **12**(1):270.

Abo-Ismael MK, Brito LF, Miller SP, Sargolzaei M, Grossi DA, Moore SS, Plastow G, Stothard P, Nayeri S, Schenkel FS. 2017. Genome-wide association studies and genomic prediction of breeding values for calving performance and body conformation traits in Holstein cattle. *Genetics, Selection, Evolution* **49**(1):82.

Agropress. 2017. Tvorba (syntéza) mléčných složek. Available from <https://www.agropress.cz/synteza-mlecnych-slozek/> (accessed May 2020).

Aksoy P, Zhu MJ, Kalari KR, Moon I, Pelleymounter LL, Eckloff BW, Wieben ED, Yee VC, Weinshilboum RM, Wang L. 2009. Cytosolic 5'-nucleotidase III (NT5C3): gene sequence variation and functional genomics. *Pharmacogenetics and Genomics* **19**(8):567-76.

Aliloo H, Pryce JE, González-Recio O, Cocks BG, Hayes BJ. 2015. Validation of markers with non-additive effects on milk yield and fertility in Holstein and Jersey cows. *BMC Genetics* **22**(16):89.

Arneson N, Hughes S, Houlston R, Done S. 2008. Whole-Genome Amplification by Improved Primer Extension Pre-amplification PCR (I-PEP-PCR). *Cold Spring Harbor Protocols* **2008**(1):pdb.prot4921.

Banos G, Bramis G, Bush SJ, Clark EL, McCulloch MEB, Smith J, Schulze G, Arsenos G, Hume DA, Psifidi A. 2017. The genomic architecture of mastitis resistance in dairy sheep. *BMC Genomics*. **18**(1):624.

Bansal BK, Gupta D. 2009. Economic analysis of bovine mastitis in India and Punjab- A review. *Indian Journal of Dairy Science* **62**:337-345.

Bedhane M, van der Werf J, Gondro C, Duijvesteijn N, Lim D, Park B, Park MN, Hee RS, Clark S. 2019. Genome-Wide Association Study of Meat Quality Traits in Hanwoo Beef Cattle Using Imputed Whole-Genome Sequence Data. *Frontiers in Genetics* **10**:1235.

Bhuiyan MSA, Lim D, Park M, Lee S, Kim Y, Gondro C, Park B, Lee S. 2018. Functional Partitioning of Genomic Variance and Genome-Wide Association Study for Carcass Traits in Korean Hanwoo Cattle Using Imputed Sequence Level SNP Data. *Frontiers in Genetics* **9**:217.

Bourens M, Barrientos A. 2017. Human mitochondrial cytochrome c oxidase assembly factor COX18 acts transiently as a membrane insertase within the subunit 2 maturation module. *Journal of Biological Chemistry*. **292**(19):7774-7783.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**:2633-2635.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. 2020. Tassel 5 User Manual. Available from <https://bitbucket.org/tasseladmin/tassel-5-source/wiki/UserManual> (accessed February 2021).

Broos S, Soete A, Hooghe B, Moran R, van Roy F, De Bleser P. 2013. PhysBinder: Improving the prediction of transcription factor binding sites by flexible inclusion of biophysical properties. *Nucleic Acids Research* **41**:W531-4.

Cai Z, Guldbrandtsen B, Lund MS, Sahana G. 2018. Prioritizing candidate genes post-GWAS using multiple sources of data for mastitis resistance in dairy cattle. *BMC Genomics* **19**: 656.

Cai Z, Gulbrandsen B, Lund MS, Sahana G. 2019. Prioritizing candidate genes for fertility in dairy cows using gene-based analysis, functional annotation and differential gene expression. *BMC Genomics* **20**:255.

ČSÚ – Český statistický úřad. 2020. Výroba a užitkovost. Available from https://vdb.czso.cz/vdbvo2/faces/cs/index.jsf?page=vystup-objekt&pvo=ZEM09&skupId=2889&z=T&f=TABULKA&katalog=30840&pvo=ZEM09&u=v150__VUZEMI__97__19#w= (accessed May 2020).

Drbohlav J, Vodičková M. 2001. Tabulky látkového složení mléka a mléčných výrobků. Ústav zemědělských a potravinářských informací, Praha.

Goddard ME, Hayes BJ. 2007. Genomic selection. *Journal of Animal Breeding and Genetics* **124**(6):323-30.

Halfon MS. 2020. Silencers, Enhancers, and the Multifunctional Regulatory Genome. *Trends in Genetics* **36**(3):149-151.

Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. 2009. Invited review: Genomic selection in dairy cattle: progress and challenges. *Journal of Dairy Science* **92**(2):433-43.

Hofírek B & Haas D. 2003. Kategorizace zdraví mléčné žlázy, klinické formy mastitid a jejich terapie. Sborník referátů odborného semináře mastitidy skotu ČBS a VFU **1**:10-22.

Huang M, Liu X, Zhou Y, Summers R, Zhang Z. 2018. BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *GigaScience* **8**(2) 2047-2170.

Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L, Girke T, Gottardo R, Hahne F, Hansen KD, Irizarry RA, Lawrence M, Love MI, MacDonald J, Obenchain V, Ole's AK, Pag'es A, Reyes A, Shannon P, Smyth GK, Tenenbaum D, Waldron L, Morgan M. 2015. Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods* **12**(2):115-121.

Chen X, Chen R, Jin R, Huang Z. 2020. The role of CXCL chemokine family in the development and progression of gastric cancer. *International Journal of Clinical Experimental Pathology* **13**(3):484-492.

Iannuzzi L. G- and R-banded prometaphase karyotypes in cattle (*Bos taurus*). 1996. *Chromosome Research* **4**(6):448-56.

Illumina. 2016. Data Analysis: GWAS Processing. Available from: <https://www.illumina.com/content/dam/illumina-marketing/documents/products/technotes/technote-gwas-processing.pdf> (Accessed February 2021).

Jeyaseelan S, Manzer R, Young SK, Yamamoto M, Akira S, Mason RJ, Worthen GS. 2005. Induction of CXCL5 during inflammation in the rodent lung involves activation of alveolar epithelium. *American Journal of Respiratory Cell and Molecular Biology* **32**(6):531-9.

Jiang L, Liu J, Sun D, Ma P, Ding X, Yu Y, Zhang Q. 2010. Genome wide association studies for milk production traits in Chinese Holstein population. *PLoS One*. **5**(10):e13661.

Jones EA, Tosh D, Wilson DI, Lindsay S, Forrester LM. 2002. Hepatic differentiation of murine embryonic stem cells. *Experimental Cell Research* **272**(1):15-22.

Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, Sabatti C, Eskin E. 2010. Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics* **42**: 348–354.

Kašná E, Fleischer P, Zavadilová L, Šlosárková S, Krupová Z, Staněk S. 2018. First national recording of health traits in dairy cows in the Czech Republic. *ICAR Technical Series* **23**:173-177.

Kobayashi K, Chinatsu K, Oyama S, Kumura H. 2016. Pro-inflammatory cytokine TNF- α is a key inhibitory factor for lactose synthesis pathway in lactating mammary epithelial cells. *Experimental Cell Research* **340**(2):295-304.

Kolbehdari D, Wang Z, Grant JR, Murdoch B, Prasad A, Xiu Z, Marques E, Stothard P, Moore SS. 2009. A whole genome scan to map QTL for milk production traits and somatic cell score in Canadian Holstein bulls. *Journal of Animal Breeding Genetics*. **126**(3):216-27.

Kopřiva V. 2011. Mléko a mlezivo – hlavní rozdíly a nutriční význam mléka ve výživě, doplňkový studijní materiál. Veterinární a farmaceutická univerzita Brno. Available from [www: http://cit.vfu.cz/ivbp/wp-content/uploads/2011/07/VY_04_03.pdf](http://cit.vfu.cz/ivbp/wp-content/uploads/2011/07/VY_04_03.pdf) (accessed June 2020).

Korkuč P, Arends D, May K, König S, Brockman GA. 2021. Genomic Loci Affecting Milk Production in German Black Pied Cattle (DSN). *Frontiers in Genetics* **12**: 640039.

Kraan MC, Patel DD, Haringman JJ, Smith MD, Weedon H, Ahern MJ, Breedveld FC, Tak PP. 2001. The development of clinical signs of rheumatoid synovial inflammation is associated with increased synthesis of the chemokine CXCL8 (interleukin-8). *Arthritis Research* **3**(1):65-71.

Kvapilík J. 2014. Mastitidy dojených krav a ekonomické ztráty. *Veterinářství* **64**(12): 946-955.

LaFramboise T. 2009. Single nucleotide polymorphism arrays: A decade of biological, computational and technological advances. *Nucleic acids research*. **37**(13): 4181-4193.

Laven R. 2016. Mastitis Part 4 - Detecting and Treating Clinical Mastitis. Available from <https://www.nadis.org.uk/disease-a-z/cattle/mastitis/mastitis-part-4-detecting-and-treating-clinical-mastitis/> (accessed May 2020).

Liu M, Fang L, Liu S, Pan MG, Seroussi E, Cole JB, Ma L, Chen H, Liu GE. 2019. Array CGH-based detection of CNV regions and their potential association with reproduction and other economic traits in Holsteins. *BMC Genomics*. **20**(1):181.

Lorenz TC. 2012. Polymerase chain reaction: basic protocol plus troubleshooting and optimization strategies. *Journal of Visualized Experiments* **63**:3998.

Lund MS, Sahana G, Andersson-Eklund L, Hastings N, Fernandez A, Schulman N, Thomsen B, Viitala S, Williams JL, Sabry A, Viinalass H, Vilkki J. 2007. Joint Analysis of

Quantitative Trait Loci for Clinical Mastitis and Somatic Cell Score on Five Chromosomes in Three Nordic Dairy Cattle Breeds. *Journal of Dairy Science* **90**(11): 5282-5290.

Marees AT, de Kluiver H, Stringer S, Vorspan F, Curis E, Marie-Claire C, Derks EM. 2018. A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *International Journal of Methods in Psychiatric Research*. **27**(2):e1608.

Marchini J, Howie B. 2010. Genotype imputation for genome-wide association studies. *Nature Review Genetics* **11**(7):499-511.

Menning M, Kufer TA. 2013. A role for the Ankyrin repeat containing protein Ankrd17 in Nod1- and Nod2-mediated inflammatory responses. *FEBS Letters* **587**(14):2137-42.

Mishra S, Mishra C, Mishra D, Bhuyan C, Rosalin B. 2017. Application of advanced molecular marker technique for improvement of animal: A critical review. *Journal of entomology and zoology studies*. **5**: 1283-1295.

Morgan AP. 2015. argyle: an R package for analysis of Illumina genotyping arrays. *G3* **6**: 281-286.

Nickerson SC, Akers RM. 2011. Mammary Gland Anatomy. *Encyclopedia of Dairy Sciences*, **3**:328–337.

Nielsen C. 2009. Economic Impact of Mastitis in Dairy Cows [Ph.D. Thesis]. Swedish University of Agricultural Sciences, Uppsala.

Nosková A. 2016. Genetická analýza mléčné užitkovosti [BSc. Thesis]. Česká zemědělská univerzita, Praha.

Olsen HG, Knutsen TM, Lewandowska-Sabat AM, Grove H, Nome T, Svendsen M, Arnyasi M, Sodeland M, Sundsaasen KK, Dahl SR, Heringstad B, Hansen HH, Olsaker I, Kent MP, Lien S. 2016. Fine mapping of a QTL on bovine chromosome 6 using imputed full sequence data suggests a key role for the group-specific component (GC) gene in clinical mastitis and milk production. *Genetics, Selection, Evolution* **48**(1):79.

Østerås O, Sølverød L, Reksen O. 2006. Milk culture results in a large Norwegian Survey – effects of season, parity, days in milk, resistance and clustering. *Journal of Dairy Science* **89**: 1010-1023.

Paredes-Sánchez FA, Sifuentes-Rincón AM, Segura Cabrera A, García Pérez CA, Parra Bracamonte GM, Ambriz Morales P. 2015. Associations of SNPs located at candidate genes to bovine growth traits, prioritized with an interaction networks construction approach. *BMC Genetics* **16**:91.

Phan J, Reue K, Péterfy M. 2000. MS-IRS PCR: a simple method for the isolation of microsatellites. *Biotechnique* **28**(1):18-20.

Porto-Neto LR, Kijas JW, Reverter A. 2014. The extent of linkage disequilibrium in beef cattle breeds using high-density SNP genotypes. *Genetics Selection Evolution* **46**(1):22.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ & Sham PC 2007. PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*, **81**.

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Available from: <https://www.R-project.org/> (accessed February 2021).

Rawlings JO, Pantula SG, Dickey DA. 1998. *Applied Regression Analysis*. Springer, New York.

Rincon G, Weber KL, Eenennaam AL, Golden BL, Medrano JF. 2011. Hot topic: performance of bovine high-density genotyping platforms in Holsteins and Jerseys. *Journal of Dairy Science* **94**(12):6116-21.

Rollin E, Dhuyvetter KC, Overton MW. 2015. The cost of clinical mastitis in the first 30 days of lactation: An economic modeling tool. *Preventive Veterinary Medicine* **122**(3):257-264.

Sahana G, Guldbbrandtsen B, Bendixen C, Lund MS. 2010. Genome-wide association mapping for female fertility traits in Danish and Swedish Holstein cattle. *Animal Genetics*. **41**(6):579-88.

Sahana G, Guldbbrandtsen B, Thomsen B, Holm LE, Panitz F, Brøndum RF, Bendixen C, Lund MS. 2014. Genome-wide association study using high-density single nucleotide polymorphism arrays and whole-genome sequences for clinical mastitis traits in dairy cattle. *Journal of Dairy Science* **97**(11):7258-75.

Sermyagin A, Gladyr E, Kharitonov SN, Ermilov AN, Strekozov NI, Brem G, Zinovieva NA. 2016. Genome-wide association study for milk production and reproduction traits in Russian Holstein cattle population. *Sel'skokhozyaistvennaya Biologiya* **51**: 182-193.

Schulman NF, Sahana G, Iso-Touru T, McKay SD, Schnabel RD, Lund MS, Taylor JF, Virta J, Vilkki JH. 2011. Mapping of fertility traits in Finnish Ayrshire by genome-wide association analysis. *Animal Genetics* **42**(3):263-9.

Smola J. 2003. Nové aspekty v etiologii mastitid. Sborník referátů odborného semináře mastitidy skotu ČBS a VFU **1**:7-9.

Snelling WM, Bennett GL, Keele JW, Kuehn LA, McDanel TG, Smith TP, Thallman RM, Kalbfleisch TS, Pollak EJ. 2019. A survey of polymorphisms detected from sequences of popular beef breeds. *Journal of Animal Science* **93**(11):5128-43.

Sodeland M, Kent MP, Olsen HG, Opsal MA, Svendsen M, Sehested E, Hayes BJ, Lien S. 2011. Quantitative trait loci for clinical mastitis on chromosomes 2, 6, 14 and 20 in Norwegian Red cattle. *Animal Genetics* **42**(5):457-65.

Stránecký V. 2015. Současné metody analýzy genomu a jejich využití v hledání genetických příčin nemocí [Ph.D. Thesis]. Karlova Univerzita, Praha.

Suravajhala P, Benso A. 2017. Prioritizing single-nucleotide polymorphisms and variants associated with clinical mastitis. *Advances and applications in bioinformatics and chemistry* **10**:57–64.

Szyda J, Mielczarek M, Frąszczak M, Minozzi G, Williams J, Wojdak-Maksymiec K. 2019. The genetic background of clinical mastitis in Holstein-Friesian cattle. *Animal* **13**(10): 2156-2163.

Tiezzi F, Parker-Gaddis KL, Cole JB, Clay JS, Maltecca C. 2015. A genome-wide association study for clinical mastitis in first parity US Holstein cows using single-step approach and genomic matrix re-weighting procedure. *PLoS One* (e0114919) DOI: 10.1371/journal.pone.0114919.

Troutt AB, McHeyzer-Williams MG, Pulendran B, Nossal GJ. 1992. Ligation-anchored PCR: a simple amplification technique with single-sided specificity. *Proceedings of the National Academy of Sciences* **89**(20): 9823-9825.

Turner S. 2018. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *Journal of Open Source Software* **3**(25):731.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M and Rozen SG. 2012. Primer3 – new capabilities and interfaces. *Nucleic Acids Research* **40**(15):e115.

VFU – Veterinární a farmaceutická univerzita Brno. 2018. Klinické vyšetření mléčné žlázy a diagnostika mastitid. Available from https://www.vfu.cz/files/1680_28_vystup.pdf (accessed May 2020).

Wang D, Sun Y, Stang P, Berlin JA, Wilcox MA, Li Q. 2009. Comparison of methods for correcting population stratification in a genome-wide association study of rheumatoid arthritis: principal-component analysis versus multidimensional scaling. *BMC Proceedings* **7**(Suppl7):S109.

Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT, Maitland A, Mostafavi S, Montojo J, Shao Q, Wright G, Bader GD, Morris Q. 2010. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Research* **38**:W214-20.

Weisberg I. 2013. Mastitis prevention and control: A prevention methodology. Available from <https://www.progressivedairy.com/topics/herd-health/mastitis-prevention-and-control-a-prevention-methodology> (accessed May 2020).

Welderufael BG, Løvendahl P, de Koning DJ, Janss LLG, Fikse WF. 2018. Genome-Wide Association Study for Susceptibility to and Recoverability From Mastitis in Danish Holstein Cows. *Frontiers in Genetics* **9**:141.

Yates AD, Achuthan P, Akanni W, Allen J, Alvarez-Jarreta J, Amode RM, Armean IM, Azov AG, Bennett R, Bhai J, Billis K, Boddu S, Marugán JC, Cummins C, Davidson C, Dodiya K, Fatima R, Gall A, Giron CG, Gil L, Grego T, Haggerty L, Haskell E, Hourlier T, Izuogu OG, Janacek SH, Juettemann T, Kay M, Lavidas I, Le T, Lemos D, Martinez JG, Maurel T, McDowall M, McMahon A, Mohanan A, Moore B, Nuhn M, Oheh DN, Parker A, Parton A, Patricio M, Sakthivel MP, Salam AIA, Schmitt BM, Schuilenburg H, Sheppard D, Sycheva M, Szuba M, Taylor K, Thormann A, Threadgold G, Vullo A, Walts B, Winterbottom A, Zadissa A, Chakiachvili M, Flint B, Frankish A, Hunt SE, Iisley G, Kostadima M, Langridge N, Loveland JE, Martin FJ, Morales J, Mudge JM, Muffato M, Perry E, Ruffier M, Trevanion SJ, Cunningham F, Howe KL, Zerbino DR, Flicek P. 2020. Ensembl 2020, *Nucleic Acids Research*, **48**(D1):D682–D688.

Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**:134.

Zhou X, Stephens M. 2012. Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*. **44**(7):821-824.

Seznam příloh

Příloha 1 – Tabulka s výsledky MDS analýzy

Příloha 2 – Tabulka s výsledky asociační analýzy

Příloha 3 – Tabulka s odhadem efektu alelických sestav