

การศึกษาความสัมพันธ์เชื่อมโยงในจีโนมสำหรับลักษณะสุขภาพเต้านมของโคนมไทยโดยใช้วิธีการแบบขั้นตอนเดียวที่มีการถ่วงน้ำหนักด้วยโมเดลวันทดสอบการถดถอยแบบสุ่ม

Genome-wide association study for udder health traits of Thai dairy cattle using weighted single-step approach with random regression test-day model

สายันห์ บัวบาน^{1*}, เกียรติศักดิ์ เหล็งหนูดำ¹, วุฒิไกร บุญคุ้ม² และ เพราพิลาส ภักดีดินแดน³

Sayan Buaban^{1*}, Keaitisak Lengnoodum¹, Wuttigai Boonkum² and Paopilas Phakdeedindan³

¹ สำนักเทคโนโลยีชีวภาพการผลิตปศุสัตว์ กรมปศุสัตว์ ต. บางกะดี อ.เมือง จ.ปทุมธานี 12000

¹ Bureau of Biotechnology in Livestock Production, Department of Livestock Development, Bangkokadee, Pathumthani, 12000

² ภาควิชาสัตวศาสตร์ คณะเกษตรศาสตร์ มหาวิทยาลัยขอนแก่น อำเภอเมือง จังหวัดขอนแก่น 40002

² Department of Animal Science, Khon Kaen University, Meaung, Khon Kaen 40002

³ ภาควิชาสัตวบาล คณะสัตวแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ปทุมวัน กรุงเทพฯ 10330

³ Department of Animal Science, Chulalongkorn University, Pathum Wan, Bangkok 10330

บทคัดย่อ: การศึกษาความสัมพันธ์ทั่วทั้งจีโนม (GWAS) เป็นเครื่องมือที่มีประสิทธิภาพในการระบุบริเวณดีเอ็นเอบนจีโนม และความผันแปรทางพันธุกรรมที่เกี่ยวข้องกับลักษณะฟีโนไทป์ การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อจำแนกบริเวณดีเอ็นเอบนจีโนมที่เกี่ยวข้องกับลักษณะสุขภาพเต้านมในประชากรโคนมไทย และใช้บริเวณดีเอ็นเอบนจีโนมที่ได้ระบุหาและกลไกทางชีววิทยาของยีนที่มีอิทธิพลต่อลักษณะเหล่านี้ ชุดข้อมูลที่ศึกษาประกอบด้วยข้อมูลคะแนนเซลล์โซมาติกในวันทดสอบ (TD-SCS) รายเดือน จำนวน 82,378 บันทึก ข้อมูลจีโนมไทป์ได้จากการตรวจหาข้อมูลความแตกต่างทางพันธุกรรมในระดับจีโนม (SNPs) ด้วย Illumina BovineSNP50 BeadChip (Illumina Inc., San Diego, CA, USA) มีจำนวน SNPs ที่เข้าวิเคราะห์ทั้งหมด 41,930 SNPs จากสัตว์ที่มีข้อมูล SNPs และฟีโนไทป์จำนวน 632 ตัว ประเมินอิทธิพลของ SNPs (SNP effect) โดยใช้วิธีการแบบขั้นตอนเดียวที่มีการถ่วงน้ำหนัก (WssGWAS) ด้วยโมเดลวันทดสอบการถดถอยแบบสุ่ม เพื่อระบุหาพื้นที่สำคัญที่เกี่ยวข้องกับลักษณะที่ศึกษาจะใช้บริเวณดีเอ็นเอบนจีโนมที่ได้อธิบายความแปรปรวนทางพันธุกรรมทั้งหมดร้อยละ 0.5 หรือมากกว่า บริเวณดีเอ็นเอบนจีโนมหลักที่เกี่ยวข้องกับคะแนนเซลล์โซมาติก (SCS) พบอยู่บนโครโมโซม BTA 11, BTA 16 และ BTA 21 บริเวณจีโนมหลายแห่งได้อธิบายถึงความแปรปรวนทางพันธุกรรมเพียงเล็กน้อย ซึ่งบ่งชี้ถึงการถ่ายทอดทางพันธุกรรมของลักษณะที่ศึกษาที่ควบคุมด้วยยีนหลายยีน เมื่อเปรียบเทียบกับผลลัพธ์กับฐานข้อมูล (NCBI, Genecards และ UniProt) พบว่ามี 21 ยีนที่มีรายงานเกี่ยวข้องกับภูมิคุ้มกัน และมีอิทธิพลต่อ SCS อย่างมีนัยสำคัญ เช่น ยีนในกลุ่ม Interleukin superfamily (IL1A, IL1B, IL1F10, IL36A, IL36B, IL36G, IL37) ตัวรับไซโตไคน์ IL1 (IL36RN, IL1RN) รวมทั้งยีนอื่น ๆ (MIA3, RCOR1, TRAF3, CDC42BPB, EXOC3L4, TNFAIP2, PPP1R13B, TAF1A, TRNAT-UGU, TRNAC-GCA, HHIPL2, NCK2) เครือข่ายและกลไกทางชีววิทยาของยีนสำหรับ SCS ส่วนใหญ่ยืนยันการค้นพบยีนที่เกี่ยวข้องกับระบบภูมิคุ้มกัน กระบวนการอักเสบ การติดเชื้อ และการแบ่งเซลล์ ยีนที่ระบุได้ในการศึกษานี้สามารถใช้เป็นยีนเป้าหมายในการศึกษาการแสดงออกของยีน

คำสำคัญ: คะแนนเซลล์โซมาติก; โคนมไทย; โมเดลวันทดสอบการถดถอยแบบสุ่ม; WssGWAS

* Corresponding author: buaban_ai@hotmail.com

ABSTRACT: Genome-wide association studies (GWAS) are a powerful tool to identify genomic regions and variants associated with phenotypes. However, only limited mutual confirmation from different studies is available. The objectives of this study were to identify genomic regions associated with udder health trait in Thai dairy cattle, and to identify genes and pathways that may influence this trait. The studied data set contained 82,378 monthly test-day somatic cell score (TD-SCS). A density of single nucleotide polymorphisms (SNPs) panel (BovineSNP50 BeadChip, Illumina Inc., San Diego, CA, USA) was used for genotyping. A total of 41,930 SNPs from 632 animals that had both genotypes and phenotypes were used for analysis. Effects of SNPs were estimated by a weighted single-step GWAS (WssGWAS), which back-solved the genomic BLUP from single-step genomic best linear unbiased prediction (ssGBLUP) using single-trait random regression test-day models. Genomic regions that explained 0.5 percent or more of the total genetic variance were selected for further analyses of candidate genes. The main genomic regions associated with SCS were located on chromosomes BTA 11, BTA 16 and BTA 21. Many genomic regions explained a small fraction of the genetic variance, indicating polygenic inheritance of the studied trait. We compared the results with databases (NCBI, Genecards and UniProt) and found 21 reported QTLs related to SCS. A large member of Interleukin superfamily (IL1A, IL1B, IL1F10, IL36A, IL36B, IL36G and IL37) IL1 receptors (IL36RN and IL1RN) as well as other genes (MIA3, RCOR1, TRAF3, CDC42BPB, EXOC3L4, TNFAIP2, PPP1R13B, TAF1A, TRNAT-UGU, TRNAC-GCA, HHIPL2 and NCK2) related to immunity significantly influencing SCS were identified. The biological networks including the immunological pathway such as lymphocyte activation are closely related to SCS. The candidate genes identified in this study can be used as target genes in studies of gene expression.

Keywords: somatic cell score; Thai dairy cattle; random regression model; WssGWAS

บทนำ

สุขภาพเต้านม เป็นลักษณะที่มีความสำคัญทางเศรษฐกิจ และมีผลต่อการทำกำไรจากอุตสาหกรรมเลี้ยงโคนมทั้งโดยตรงและโดยอ้อม การลดลงของจำนวนเซลล์โซมาติกจะช่วยให้โคนมเป็นโรคเต้านมอักเสบลดลง ซึ่งจะช่วยลดผลกระทบจากความสูญเสียทางเศรษฐกิจที่เกิดจากการให้ผลผลิตน้ำนมลดลง (Seegers et al., 2003) จำนวนเซลล์โซมาติกเป็นลักษณะที่ควบคุมด้วยยีนหลายคู่ (polygenic) ซึ่งได้รับผลกระทบจากปัจจัยหลายอย่าง และหลายยีน ซึ่งแต่ละยีนมีผลกระทบเพียงเล็กน้อยต่อการแสดงออกของลักษณะ (Snelling et al., 2013) การปรับปรุงด้านการจัดการและโภชนาการ พร้อมกับการคัดเลือกพันธุ์กรรมอย่างเข้มข้นสามารถจำกัดการเกิดโรคเต้านมอักเสบได้ (Rupp and Boichard, 2003)

ในทศวรรษที่ผ่านมาการคัดเลือกด้วยจีโนม (genomic selection, GS) ซึ่งหมายถึงการคัดเลือกจากค่าการผสมพันธุ์จีโนม (genomic estimated breeding values, GEBVs) ของสัตว์แต่ละตัวที่ได้จากการประเมินพันธุ์กรรมจีโนม (genomic evaluation) จากการใช้ข้อมูลจีโนม ร่วมกับข้อมูลฟีโนไทป์ และข้อมูลพันธุ์ประวัติ (Meuwissen et al., 2001) ได้กลายมาเป็นวิธีการมาตรฐาน และเครื่องมือที่สำคัญสำหรับการปรับปรุงทางพันธุกรรมของโคนมจนเกิดความสำเร็จในการพัฒนาแผนการคัดเลือกแบบดั้งเดิมในหลายประเทศในปัจจุบัน เช่น ประเทศสหรัฐอเมริกา แคนาดา สหราชอาณาจักร นิวซีแลนด์ ออสเตรเลีย ฝรั่งเศส เนเธอร์แลนด์ เยอรมนี และประเทศสแกนดิเนเวีย (Silva et al., 2014) ทั้งนี้ เนื่องจากความก้าวหน้าอย่างรวดเร็วที่เกิดจากเทคโนโลยีไมโครอะเรย์ โดยการจำแนกความแตกต่างทางพันธุกรรมของโคนมรายตัวสามารถทำได้ครั้งละจำนวนมากและมีราคาถูกอย่างต่อเนื่อง (จากราคามากกว่า 500 ดอลลาร์ต่อตัวในปีพ.ศ. 2533 เป็นประมาณ 50 ดอลลาร์ต่อตัวในปัจจุบัน) สามารถจำแนกรูปแบบทางพันธุกรรมของโคนมแต่ละตัวได้ตั้งแต่ 2,900 SNPs (Bovine3K Genotyping Beadchip, Illumina®) ถึง 777,000 SNPs (BovineHD Genotyping BeadChip, Illumina®) ครอบคลุมทั่วทั้งจีโนมของโคนม ดังนั้น การศึกษาความสัมพันธ์เชื่อมโยงในจีโนม (genome-wide association study, GWAS) ซึ่งหมายถึงการหาความสัมพันธ์ระหว่างความผันแปรของเครื่องหมายทางพันธุกรรม (เช่น SNPs) ทั่วทั้งจีโนมกับฟีโนไทป์ที่สนใจของตัวสัตว์ (Hayes and Goddard, 2010) โดยการตรวจหาความแปรปรวนทางพันธุกรรม หรือตำแหน่งของยีนบริเวณดีเอ็นเอบนจีโนมที่มีส่วนเกี่ยวข้องทำให้เกิดการเปลี่ยนแปลงของฟีโนไทป์ (Johansson et al., 2010) จึงกลายเป็นจริงในทางปฏิบัติสำหรับการระบุตำแหน่ง SNPs ที่เกี่ยวข้องกับ quantitative trait loci (QTL) หรือยีนที่มีอิทธิพลที่สำคัญ (Zhang et al., 2016) เมื่อเปรียบเทียบกับกลยุทธ์การทำแผนที่ QTL แบบ

ดั้งเดิมแล้ว การศึกษาความสัมพันธ์ทั่วทั้งจีโนมมีข้อได้เปรียบที่สำคัญทั้งในด้านอำนาจการตรวจสอบความแปรปรวนเชิงสาเหตุ (causal variants) ที่มีอิทธิพลเพียงเล็กน้อย และในการกำหนดบริเวณดีเอ็นเอบนจีโนมซึ่งมีการแปรผันเชิงสาเหตุที่แคบลง (Hirschhorn and Daly, 2005) จึงได้รับการยอมรับอย่างกว้างขวางว่าเป็นเครื่องมือหลักในการระบุ QTL และบริเวณดีเอ็นเอบนจีโนมที่เกี่ยวข้องกับฟีโนไทป์ และมีความเป็นไปได้ที่เราจะแยกลักษณะเชิงปริมาณที่ถ่ายทอดได้ เช่นลักษณะสุขภาพเต้านม โดยการหาแผนที่บริเวณดีเอ็นเอบนจีโนม หรือ QTL

เมื่อเร็ว ๆ นี้การศึกษาความสัมพันธ์เชื่อมโยงในจีโนมโดยใช้วิธีการแบบขั้นตอนเดียวที่มีการถ่วงน้ำหนัก (weighted single-step genome-wide association, WssGWAS) ที่นำเสนอโดย Wang et al. (2012) เป็นวิธีการที่ช่วยให้การประเมินอิทธิพลของ SNPs จาก GEBV ที่ได้จากการทำนายเชิงเส้นแบบไม่มีอคติในขั้นตอนเดียวที่ดีที่สุด (single-step genomic best linear unbiased prediction, ssGBLUP) (Aguilar et al., 2010) โดยใช้ข้อมูลฟีโนไทป์ ข้อมูลจีโนไทป์ และพันธุ์ประวัติของสัตว์ที่เกี่ยวข้องพร้อมกันขั้นตอนเดียว วิธีการนี้สามารถนำไปประยุกต์ใช้กับโมเดลที่มีความซับซ้อน เช่น โมเดลการถดถอยแบบสุ่ม ได้อย่างมีประสิทธิภาพ โดยเฉพาะการทำนายพันธุ์กรรมจีโนมของลักษณะที่มีการวัดฟีโนไทป์ต่อเนื่อง (Kang et al., 2017) และสามารถใช้ได้กับ SNPs ที่มีความแปรปรวนที่ไม่เท่ากัน ซึ่งส่งผลให้เกิดความแม่นยำในการประมาณค่าอิทธิพลของ SNPs มากยิ่งขึ้น (Wang et al., 2012) ดังนั้นเมื่อขนาดประชากรของสัตว์ที่มีทั้งฟีโนไทป์และจีโนไทป์มีขนาดเล็ก และมีการควบคุมลักษณะฟีโนไทป์โดย QTL ที่มีอิทธิพลขนาดใหญ่แล้ว WssGWAS อาจทำงานได้ดีกว่าวิธี GWAS แบบดั้งเดิม การศึกษาล่าสุดที่ได้ใช้วิธีการนี้สำหรับลักษณะที่มีความสำคัญทางเศรษฐกิจในปศุสัตว์ได้แก่ ลักษณะอัตราการเจริญเติบโต อัตราแลกเนื้อ น้ำหนักตัว และปริมาณเนื้อในไก่อกระตัง (Fragomeni et al., 2014; Wang et al., 2014) อัตราการเจริญเติบโตและอัตราแลกเนื้อในสุกร (Howard et al., 2015) องค์ประกอบโปรตีนในน้ำมันของโคนม (Zhou et al., 2019) กรดไขมันในโคเนื้อ (Lemos et al., 2016) ลักษณะซากในโคเนื้อ (Silva et al., 2017, 2019) ลักษณะทางการสืบพันธุ์ในไก่อกระตัง (Fragomeni et al., 2014) ลักษณะทางการสืบพันธุ์ในโคเนื้อ (Irano et al., 2016; Melo et al., 2017) ลักษณะน้ำเชื้อและขนาดครอกในสุกร (Daniele et al., 2018) โรคเต้านมอักเสบ (Tiezzi et al., 2015) และความเร็วในการต่อสู้ในโคเนื้อ (Valente et al., 2016)

การประเมินค่าทางพันธุกรรมระดับประเทศสำหรับลักษณะสุขภาพเต้านมของโคนมไทยได้ดำเนินการโดยใช้โมเดลวันทดสอบการถดถอยแบบสุ่ม (random regression test-day model, RR-TDM) โดยใช้ข้อมูลฟีโนไทป์และข้อมูลพันธุ์ประวัติ (สายพันธ์ และคณะ, 2563) และเมื่อมีการผสมผสานระหว่าง RR-TDM เข้ากับวิธีการแบบขั้นตอนเดียวซึ่งเรียกว่า single-step random regression test-day model (SS-RR-TDM) ในการทำนายค่าทางพันธุกรรมจีโนม น่าจะเป็นวิธีที่จะทำให้ได้การทำนายที่แม่นยำมากขึ้น มีอคติลดลง และส่งผลให้การศึกษาความสัมพันธ์เชื่อมโยงในจีโนมมีประสิทธิภาพที่ดีขึ้น

ข้อมูลบริเวณดีเอ็นเอบนจีโนมที่เกี่ยวข้องกับลักษณะสุขภาพเต้านมของประชากรโคนมไทยในปัจจุบันยังมีอยู่อย่างจำกัด การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อจำแนกบริเวณดีเอ็นเอบนจีโนมที่เกี่ยวข้องกับลักษณะสุขภาพเต้านม ได้แก่ คะแนนเซลล์โซมาติก (somatic cell score, SCS) ในประชากรโคนมไทย และใช้บริเวณดีเอ็นเอบนจีโนมที่ได้รับจากหาลายนิ้วมือและกลไกทางชีววิทยาของยีนที่มีอิทธิพลต่อลักษณะ SCS

วิธีการศึกษา

ข้อมูลฟีโนไทป์ พันธุ์ประวัติ และจีโนไทป์

ข้อมูลที่ใช้ในการศึกษานี้เป็นข้อมูลเซลล์โซมาติกในวันทดสอบ (test-day somatic cell count, TD-SCC, 1000เซลล์/ซีซี) รายตัวของแม่โคนมไทยในรอบการให้นมครั้งแรกที่ได้สุ่มเก็บตัวอย่างจากฟาร์มเกษตรกรผู้เลี้ยงโคนมทั่วประเทศเป็นรายเดือนใน ภาคเหนือ ภาคตะวันออกเฉียงเหนือ ภาคตะวันออก ภาคกลาง และภาคตะวันตก จำนวน 132, 294, 219, 451 และ 245 ฟาร์ม ตามลำดับ แล้วนำไปตรวจวิเคราะห์ด้วยเครื่องวิเคราะห์องค์ประกอบน้ำนม และเซลล์โซมาติกอัตโนมัติ (MilkoScan™, FOSS) คลอดลูกระหว่างเดือนพฤศจิกายน 2536 ถึงเดือนมีนาคม 2560 ที่รวบรวมไว้ในระบบฐานข้อมูลโคนมของสำนักเทคโนโลยีชีวภาพการผลิตปศุสัตว์ กรมปศุสัตว์ หลังจากนั้นได้จัดการข้อมูลเพื่อให้เป็นไปตามเงื่อนไขดังต่อไปนี้คือ อายุคลอดลูกครั้งแรกอยู่ระหว่าง 20 ถึง 48 เดือน วันให้นมจำกัดให้อยู่

ระหว่าง 5 และ 305 วัน วันทดสอบครั้งแรกอยู่ระหว่าง 5 ถึง 60 วันหลังคลอด ปริมาณน้ำนมในแต่ละวันอยู่ระหว่าง 1 ถึง 45 กิโลกรัม แม่โคแต่ละตัวต้องมีข้อมูลในวันทดสอบอย่างน้อย 5 บันทึก และกำหนดให้มีข้อมูลอยู่ในกลุ่มการจัดการเดียวกัน (contemporary groups) อย่างน้อย 5 บันทึก นอกจากนี้แม่โคทั้งหมดจะต้องทราบพ่อพันธุ์ และข้อมูลพันธุ์ประวัติจะถูกสืบทอดกลับไป 3 ชั่วอายุ ภายหลังการจัดการข้อมูลจะมีข้อมูล TD-SCC สำหรับการศึกษาจำนวน 82,378 ข้อมูลจากแม่โคนม 13,737 ตัว รวมกับข้อมูลพันธุ์ประวัติของโคที่เกี่ยวข้องกับโคที่ให้ผลผลิตจำนวน 32,743 ตัว เนื่องจากจำนวน SCC ค่อนข้างที่จะมีการแจกแจงแบบเบ้ (skewed distribution) จึงถูกแปลงเป็นคะแนนเซลล์โซมาติกในวันทดสอบ (test-day somatic cell score, TD-SCS) ตามสมการ $TD-SCS = \log_2(TD-SCC/100) + 3$ (Ali and Shook 1980) เพื่อให้ข้อมูลมีความเหมาะสมในการวิเคราะห์

Table 1 Number of animals and records used in this study

Items	Number of animals and records
Animals in pedigree	32,743
Animals with records	13,737
Test-day records	82,378
Genotyped animals	632
Bulls	142
Cows	490
SNP information	41,930

ข้อมูลจีโนมได้จากการตรวจหาข้อมูลความแตกต่างทางพันธุกรรมในระดับจีโนมด้วย Illumina BovineSNP50 BeadChip (Illumina Inc., San Diego, CA, USA) ที่เป็นเวอร์ชัน 2 (มีจำนวน 54,609 SNPs) หรือ เวอร์ชัน 3 (มีจำนวน 53,218 SNPs) โดยทั้งสองเวอร์ชันมี SNPs ที่เหมือนกันจำนวน 50,908 SNPs โดยข้อมูล SNPs ที่ได้จะต้องผ่านการควบคุมคุณภาพ (quality control) ซึ่งพิจารณาด้วย call rate > 0.9 สำหรับทั้ง SNPs และจีโนมสัตว์ ความถี่อัลลีลย่อย (minor allele frequency, MAF) > 0.05 การออกจากสมดุล Hardy-Weinberg (ความแตกต่างระหว่างความถี่ที่คาดหวัง และความถี่ที่สังเกต) < 0.15 การทดสอบความขัดแย้งระหว่างคู่พ่อแม่ลูกหลาน (parent-progeny conflict) นอกจากนี้ SNP ที่ไม่ทราบตำแหน่งหรืออยู่ในโครโมโซมเพศจะไม่ได้ถูกนำมาใช้ในการวิเคราะห์ ภายหลังจากการควบคุมคุณภาพ มีข้อมูล SNPs ที่เหลือสำหรับการวิเคราะห์ จำนวน 41,930 SNPs จากสัตว์จำนวน 632 ตัว (Table 1) การควบคุมคุณภาพ SNPs และตัวอย่างนั้นดำเนินการโดยใช้โปรแกรม PREGSF90 (Misztal, 2018) รายละเอียดของชุดข้อมูลที่ใช้ดังแสดงรายละเอียดไว้ใน Table 1

การวิเคราะห์ข้อมูล

1. การทำนายค่าการผสมพันธุ์จีโนม (GEBV)

ค่าสัมประสิทธิ์การถดถอยสุ่มของ GEBV (random regression coefficient of GEBV) ของสัตว์แต่ละตัวสำหรับแต่ละลักษณะที่ได้นั้น เริ่มต้นจากการวิเคราะห์หาค่าประกอบความแปรปรวนและความแปรปรวนร่วม เพื่อใช้ในการประเมินค่าการผสมพันธุ์จีโนม โดยใช้วิธีการแบบขั้นตอนเดียวจากโมเดลวันทดสอบการถดถอยแบบสุ่ม (SS-RR-TDM) วิเคราะห์ที่ละลักษณะด้วย BLUPF90 software family (Misztal et al., 2018) ซึ่ง SS-RR-TDM สามารถเขียนให้อยู่ในรูปของเมตริกซ์ได้ดังนี้:

$$y = X_1b_1 + X_2b_2 + Vh + Za + Wp + e, \quad (1)$$

โดยที่ $\mathbf{y}$ เป็นเวกเตอร์ของบันทึก SCS ในวันทดสอบ; $\mathbf{b}_1$ เป็นเวกเตอร์ของอิทธิพลที่เป็นระบบซึ่งประกอบด้วยปัจจัยคงที่ของ ปี-ฤดูกาลตลอด; $\mathbf{b}_2$ เป็นเวกเตอร์ของค่าสัมประสิทธิ์การถดถอยคงที่ของ Legendre polynomials (LPs) ที่ซ่อนอยู่ในกลุ่มพันธุ์-กลุ่มอายุที่ตลอด; $\mathbf{h}$ เป็นเวกเตอร์ของอิทธิพลสุ่มเนื่องจาก ฝูง-ปี-เดือนทดสอบ; $\mathbf{a}$ และ $\mathbf{p}$ เป็นเวกเตอร์ของสัมประสิทธิ์การถดถอยแบบสุ่มของ LPs ที่ซ่อนกันภายในอิทธิพลเนื่องจากพันธุกรรม และอิทธิพลเนื่องจากสิ่งแวดล้อมตามลำดับ; และ $\mathbf{e}$ เป็นเวกเตอร์ของอิทธิพลของความคลาดเคลื่อน เมทริกซ์ $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{V}$, $\mathbf{Z}$ และ $\mathbf{W}$ เป็น incidence matrices ที่สอดคล้องกับอิทธิพลที่กล่าวมาข้างต้น ลำดับของ LPs ซึ่งเป็นค่าแปรผันร่วมของปัจจัยคงที่ และปัจจัยสุ่ม ที่คำนวณจากวันให้นม (DIM) ตามที่กำหนดโดย Gengler และคณะ (1999) การศึกษาครั้งนี้ใช้ LPs ลำดับที่ 2 (constant, linear, and quadratic) ซึ่งได้จากเปรียบเทียบกับลำดับที่แตกต่างกัน โดยพิจารณาจากค่า log-likelihoods สูงที่สุด ค่า Akaike's information criteria (AIC: Burnham and Anderson 2002) และความแปรปรวนของความคลาดเคลื่อนต่ำสุด และใช้ LPs ลำดับเดียวกันสำหรับอิทธิพลการถดถอยแบบคงที่และแบบสุ่ม

ตัวแปรของอิทธิพลของความคลาดเคลื่อนถูกสมมติให้มีค่าเท่ากันตลอดการให้นมเพื่อลดความซับซ้อนของโมเดล สมมติให้อิทธิพลแบบสุ่มว่ามีกระจายแบบปกติด้วยค่าเฉลี่ยเท่ากับ 0 และโครงสร้างความแปรปรวนร่วมของโมเดลกำหนดให้เป็นดังนี้

$$\text{Var} \begin{bmatrix} \mathbf{h} \\ \mathbf{a} \\ \mathbf{p} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{I}\sigma_{\text{htm}}^2 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{H} \otimes \mathbf{G}_0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \otimes \mathbf{P}_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}\sigma_e^2 \end{bmatrix} \quad (2)$$

โดยที่ $\mathbf{G}_0$ และ $\mathbf{P}_0$ เป็นเมทริกซ์ 3×3 ความแปรปรวนร่วมของสัมประสิทธิ์การถดถอยสุ่มสำหรับอิทธิพลเนื่องจากพันธุกรรม และอิทธิพลเนื่องจากสภาพแวดล้อมตามลำดับ, $\mathbf{H}$ เป็นเมทริกซ์ความสัมพันธ์ทางพันธุกรรมที่ดัดแปลงจากการรวมกันของความสัมพันธ์ระหว่างสัตว์ในพันธุ์ประวัติ (numerator relationship matrix, $\mathbf{A}$) กับความสัมพันธ์ทางจีโนม (genomic-based relationship matrices, $\mathbf{G}$), $\mathbf{I}$ เป็นเมทริกซ์เอกลักษณ์ (Identity matrix), $\otimes$ เป็น Kronecker product operation, σ_{htm}^2 เป็นความแปรปรวนร่วมของฝูง-ปี-เดือนทดสอบ และ σ_e^2 เป็นความแปรปรวนร่วมของความคลาดเคลื่อน

ค่าผกผันของ $\mathbf{H}$ ซึ่งจำเป็นสำหรับการคำนวณแบบผสมสามารถเขียนได้ดังนี้:

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \tau(\alpha\mathbf{G} + \beta\mathbf{A}_{22})^{-1} - \omega\mathbf{A}_{22}^{-1} \end{bmatrix}, \quad (3)$$

โดยที่ $\mathbf{A}$ คือเมทริกซ์ความสัมพันธ์ระหว่างสัตว์ในพันธุ์ประวัติสำหรับสัตว์ทุกตัว; $\mathbf{A}_{22}$ เป็นเมทริกซ์ความสัมพันธ์ระหว่างสัตว์ในพันธุ์ประวัติสำหรับสัตว์ที่มีจีโนมไทป์ α , β , ω และ τ เป็นปัจจัยถ่วงน้ำหนัก α และ β เป็นค่าที่ให้เพื่อหลีกเลี่ยงปัญหา singularity โดยมีค่าเท่ากับ 0.95 และ 0.05 ตามลำดับ (VanRaden, 2008) เมทริกซ์ $\mathbf{G}$ ถูกสร้างขึ้นตามวิธีการของ VanRaden (2008) ดังนี้:

$$\mathbf{G} = \frac{\mathbf{Z}\mathbf{D}\mathbf{Z}}{2 \sum_{i=1}^m p_i(1-p_i)}, \quad (4)$$

โดยที่ $\mathbf{Z}$ คือเมทริกซ์ของ SNPs ที่ปรับค่าตามความถี่อัลลีล โดย AA, Aa และ aa มีค่า -1, 0, และ 1 ตามลำดับ; $\mathbf{D}$ คือเมทริกซ์แนวทแยงมุมของค่าถ่วงน้ำหนักสำหรับความแปรปรวนของ SNPs (เริ่มแรก $\mathbf{D} = \mathbf{I}$); m คือจำนวนเครื่องหมาย SNPs และ p_i คือความถี่อัลลีลของ SNPs ที่ i

ผลลัพธ์จากการแก้สมการโมเดลผสมจาก SS-RR-TDM (ค่าสัมประสิทธิ์การถดถอยแบบสุ่มของ GEBV) จะนำไปคำนวณค่าการผสมพันธุ์จีโนมซึ่งจะระบุว่าเป็นพื้นฐานการผลิตที่ 305 วัน ดังนั้นค่าการผสมพันธุ์จีโนมที่ 305 วันจึงกำหนดให้เป็นค่าเฉลี่ยของค่าการผสมพันธุ์ระหว่างวันให้นมที่ 5 ถึง 305 ตามที่อธิบายโดย Jamrozik et al (1997)

2. การศึกษาความสัมพันธ์เชื่อมโยงในจีโนม (GWAS)

การศึกษาความสัมพันธ์เชื่อมโยงในจีโนมวิเคราะห์โดยใช้วิธีการแบบขั้นตอนเดียวที่มีการถ่วงน้ำหนัก (WssGWAS) อิทธิพลของ SNPs หรือ SNP effect ($\mathbf{u}$) นั้นคำนวณมาจากการวนการวนซ้ำ (iteration) คล้ายกับที่อธิบายโดย Wang et al (2014) ด้วยซอฟต์แวร์ postGSf90 (Aguilar et al., 2014) โดยใช้แบบจำลองเชิงเส้นเดียวกันกับที่ใช้ในการประเมินองค์ประกอบความแปรปรวน โดยสรุปแล้ว

ซอฟต์แวร์ postGSf90 จะแก้สมการเปลี่ยนค่าการผสมพันธุ์โนม ($\mathbf{a}_g$: GEBV สำหรับ SCS ที่ 305 วันเฉลี่ย) ที่ได้คำนวณมาจากค่าสัมประสิทธิ์การถดถอยแบบสุ่มของ GEBV ของสัตว์แต่ละตัวจาก SS-RR-TDM ไปเป็นอิทธิพลของ SNPs ที่พิจารณาใช้ความแปรปรวนทางจีโนม (σ_u^2) ร่วมกันตามที่อธิบายไว้ในสมการ:

$$\begin{bmatrix} \mathbf{a}_g \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{ZDZ}' & \mathbf{ZD}' \\ \mathbf{DZ}' & \mathbf{D} \end{bmatrix} \sigma_u^2 \quad (5)$$

อิทธิพลของแต่ละ SNPs ได้มาจากการแก้สมการ:

$$\mathbf{u} = \mathbf{DZ}'\mathbf{G}^{-1}\mathbf{a}_g \quad (6)$$

เมื่อ $\mathbf{u}$ เป็นเวกเตอร์ของอิทธิพลของ SNPs; $\mathbf{a}_g$ เป็นเวกเตอร์ของ GEBVs และข้อกำหนดอื่น ๆ ได้กำหนดไว้ก่อนหน้านี้แล้ว

ในรอบแรกของการวนซ้ำ ค่าความแปรปรวนที่ถูกปรับโดยแต่ละ SNPs มีค่าเท่ากับ $2p_i(1 - p_i)$ และ $\mathbf{D} = \mathbf{I}$ ในการวนซ้ำรอบถัดไป มีการสร้างค่าถ่วงน้ำหนักเมตริกซ์ $\mathbf{G}$ ซึ่งการมีส่วนร่วมของ SNPs ที่คาดหวังไว้จะถูกแทนที่ด้วยความแปรปรวนที่เกิดขึ้นจริงเพื่อให้บริเวณดีเอ็นเอจีโนมที่มีอิทธิพลสูงต่อความแปรปรวนทางพันธุกรรมของลักษณะเห็นได้เด่นชัดมากยิ่งขึ้น ดังนั้นองค์ประกอบของ $\mathbf{D}$ จะถูกแทนค่าด้วย $D_{ii} = \frac{2p_i(1-p_i)u^2}{m}$ โดยที่ $\mathbf{u}$ คืออิทธิพลของ SNPs ในกระบวนการวนซ้ำครั้งก่อนหน้า จากนั้น อิทธิพลของ SNPs ครั้งใหม่ได้รับการพิจารณาจากค่าถ่วงน้ำหนักเมตริกซ์ $\mathbf{G}$ ตามสมการ (6) ที่ได้รายงานข้างต้น สำหรับคำอธิบายโดยละเอียดของขั้นตอนการวนซ้ำสามารถดูได้ตามขั้นตอนจาก ‘Scenario 1’ ใน Wang et al. (2012)

สำหรับกระบวนการวนซ้ำรอบแรก ค่าถ่วงน้ำหนักสำหรับแต่ละ SNPs มีค่าเท่ากับ 1 ซึ่งหมายความว่า SNPs ทั้งหมดมีน้ำหนักเท่ากัน (เช่น GBLUP ขั้นตอนเดียวแบบมาตรฐาน) สำหรับการทำการซ้ำต่อไป (รอบที่ 2, รอบที่ 3, ฯลฯ) ค่าถ่วงน้ำหนัก เป็นความแปรปรวนของแต่ละ SNPs ที่คำนวณโดยใช้อิทธิพลจากการแทนที่เบสในอัลลีลที่ประเมินได้จากการวนซ้ำก่อนหน้านี้ และความถี่อัลลีลนั้นๆ (Wang et al. 2012) ในการทำการซ้ำแต่ละครั้งค่าถ่วงน้ำหนักของ SNPs จะนำไปใช้ในการสร้างเมตริกซ์ $\mathbf{G}$ เพื่อคำนวณค่า GEBV ใหม่ และประมาณค่าอิทธิพลของ SNPs อีกครั้ง

ในการใช้งานครั้งแรกของ WssGBLUP Wang et al. (2012) แนะนำว่าควรคำนวณค่าถ่วงน้ำหนัก SNPs จากสมการ $d_{i(t+1)}^* = \hat{u}_{(t)}^2 2p_i(1 - p_i)$ ซึ่งอ้างอิงจากสมการสำหรับการคำนวณความแปรปรวนทางพันธุกรรมของ additive locus (Falconer and Mackay, 1996) อย่างไรก็ตาม Lourenco et al. (2017) แสดงให้เห็นว่าวิธีนี้ไม่สามารถเข้าถึง convergence ในกรณีที่มีลักษณะมีการควบคุมด้วยยีนจำนวนมาก หรือ polygenic trait เนื่องจากมีถ่วงน้ำหนักมากเกินไป ดังนั้นการถ่วงน้ำหนัก SNPs ที่ใช้ในการศึกษาครั้งนี้จึงใช้วิธี non-linear A ตามที่อธิบายไว้โดย VanRaden (2008):

$$d_{i(t+1)}^* = CT \frac{|\hat{a}_j|}{sd(\hat{a})} - 2 \quad (7)$$

โดยที่ CT เป็นค่าคงที่ควบคุมการเบี่ยงเบนจากการกระจายแบบปกติ; $|\hat{a}_j|$ เป็นค่าสัมบูรณ์ของอิทธิพลของ SNP ที่ประมาณได้สำหรับ SNPs ที่ j และ $sd(\hat{a})$ คือค่าเบี่ยงเบนมาตรฐานของเวกเตอร์ของอิทธิพลของ SNPs ที่ประมาณได้ การถ่วงน้ำหนักด้วยวิธี non-linear A ทำให้การคำนวณเข้าสู่ convergence ได้ดี และหลีกเลี่ยงค่าที่สูญขีด (Garcia et al., 2018) เนื่องจากการเปลี่ยนแปลงค่าถ่วงน้ำหนักสูงสุดถูกจำกัดโดยค่าต่ำสุดในช่วงระหว่าง 5 และเลขชี้กำลังของ CT โดยในการศึกษาครั้งนี้ได้กำหนดค่า CT เท่ากับ 1.125 ตาม Legarra et al. (2018) และ VanRaden (2008)

3. การจำแนกบริเวณดีเอ็นเอบนจีโนม (Genomic region identification)

เนื่องจากการตรวจหาบริเวณหรือตำแหน่งที่อยู่บนพื้นฐานของ SNPs เพียงๆ อาจส่งผลให้เกิดการรบกวนหรือการประเมินค่าต่ำเกินไป เพราะว่อัตราส่วนระหว่างจำนวน SNPs และจำนวนสัตว์ที่มีจีโนไทป์มีค่าสูง (Wang et al., 2014) นอกจากนี้ SNPs ที่อยู่ตำแหน่งติดกันอาจมีค่า Linkage disequilibrium (LD) ที่สูงต่อ quantitative trait locus (QTL) เดียวกันในแผง SNPs ที่มีความหนาแน่นสูง (High-density SNPs panel) ซึ่งจะส่งผลทำให้อิทธิพลของ QTL จะกระจายไปทั่ว SNP ทั้งหมดในบริเวณที่มีค่า LD สูง (Fan et al., 2011)

ด้วยเหตุนี้บริเวณหน้าต่างจีโนมที่ไม่ได้ซ้อนทับกันขนาด 1-Mb (non-overlapping window) ซึ่งเป็นสัดส่วนของความแปรปรวนทางพันธุกรรมในแต่ละบริเวณที่ประกอบด้วยหน้าต่างจีโนม กว้าง 1 Mb จะถูกคำนวณและใช้เพื่อระบุบริเวณดีเอ็นเอบนจีโนม เนื่องจากมีความเหมาะสมกว่าการใช้ SNPs เดียวๆ ในการพิจารณาจำนวน SNPs ที่เหมาะสม และจำนวนรอบในกระบวนการวนซ้ำ ได้มีการทดสอบหาค่าเฉลี่ยของบริเวณหน้าต่างจีโนมที่ไม่ทับซ้อนกันจากกระบวนการวนซ้ำด้วยจำนวน SNPs ที่แตกต่างกัน (10, 15, 20, 30 และ 40 SNPs) ตามวิธี WssGWAS โดยพิจารณาจากค่าตามความถูกต้องที่แท้จริง (realized accuracies) ของค่าความสามารถในการทำนาย (predictive ability, $\text{Cor}(\text{GEBV}, \text{TBV})$) และค่าอคติ (bias, $\text{TBV} = b_0 + b_1 \times \text{GEBV}$) ของ GEBV (Legarra et al., 2008; Garcia et al., 2018) และความเสถียรของการประมาณอิทธิพลของ SNPs (Wang et al., 2014) โดยเบื้องต้น ในการศึกษาครั้งนี้เราตัดสินใจที่จะใช้ผลลัพธ์จากการวนซ้ำครั้งที่ 2 ด้วยบริเวณหน้าต่างจีโนมที่มีจำนวน SNP ที่อยู่ติดกันอย่างต่อเนื่อง 20 SNP เพราะให้ความสามารถในการคาดการณ์มากที่สุดและเกิด inflation น้อยที่สุด และเกิดความเสถียรของการประมาณอิทธิพลของ SNP มากที่สุด

ร้อยละของความแปรปรวนทางพันธุกรรมที่อธิบายโดยบริเวณหน้าต่างจีโนมที่ประกอบด้วย SNPs ที่ต่อเนื่องกันที่ i ตามที่อธิบายโดย Wang et al (2014) i คำนวณได้ดังนี้:

$$\frac{\text{var}(a_i)}{\sigma_a^2} = \frac{\text{var} \sum_{j=1}^{20} z_j \hat{u}_j}{\sigma_a^2}, \quad (8)$$

โดยที่ a_i คือค่าทางพันธุกรรมของบริเวณหน้าต่างจีโนมที่ i ที่ประกอบไปด้วย SNP ที่อยู่ติดกันอย่างต่อเนื่อง 20 SNPs, σ_a^2 คือความแปรปรวนทางพันธุกรรมทั้งหมด, Z_j คือเวกเตอร์ของ SNP ที่ปรับค่าตามความถี่ยีนที่ j สำหรับสัตว์ทั้งหมด และ $\hat{u}_j$ คืออิทธิพลของ SNP ที่ j ภายในบริเวณหน้าต่างจีโนมที่ i (Zhang et al., 2010)

เพื่อที่จะจำแนกบริเวณดีเอ็นเอบนจีโนมที่มีความสำคัญเกี่ยวข้องกับลักษณะที่ได้ศึกษา จะเลือกบริเวณหน้าต่างจีโนมที่ไม่ทับซ้อนกัน (Non-overlapping windows) ที่มีจำนวน 20 SNPs ซึ่งสามารถอธิบายความแปรปรวนทางพันธุกรรมทั้งหมดได้ร้อยละ 0.5 หรือมากกว่า ตามเกณฑ์การศึกษาของ Fragomeni et al. (2014), Irano et al. (2016), Lemos et al. (2016), Oliveira et al. (2017), Silva et al. (2017), Lee et al., (2019), Zhou et al. (2019) และ Oliveira et al. (2019) ซึ่งพิจารณาแล้วว่าเป็นเกณฑ์ที่เหมาะสมสำหรับวัตถุประสงค์ของการศึกษาครั้งนี้ และแสดงรูปแบบกราฟ Manhattan plot ของความแปรปรวนทางพันธุกรรมของเครื่องหมายที่อธิบายโดยบริเวณดีเอ็นเอบนจีโนมเหล่านี้โดยใช้ gnuplot 5.2 (Williams and Kelley, 2019)

4. การระบุยีนที่เกี่ยวข้อง (Gene annotation)

การระบุยีนที่อยู่ภายในแต่ละบริเวณดีเอ็นเอบนจีโนมที่ได้เลือกแล้ว (จุดเริ่มต้นและจุดสิ้นสุดของหน้าต่าง) จะใช้เครื่องมือจาก Map Viewer ของฐานข้อมูล NCBI ด้วยชุดประกอบ UMD 3.1 เป็นแผนที่อ้างอิง

(https://www.ncbi.nlm.nih.gov/projects/mapview/map_search.cgi?taxid=9913&build=104.0) และยีนที่ระบุได้ทั้งหมดจะตรวจสอบผ่านการสืบค้นโดยใช้ฐานข้อมูลเอกสาร และฐานข้อมูลสาธารณะอื่นๆ (NCBI-<https://www.ncbi.nlm.nih.gov/>, Genecards-<https://www.genecards.org/> และ Uniprot-<https://www.uniprot.org/>) เพื่อให้ทราบถึงหน้าที่ของยีนที่เกี่ยวข้องต่อลักษณะที่ศึกษา

5. การวิเคราะห์เครือข่ายและกลไกทางชีววิทยาของยีน (Gene network and pathway analysis)

รายชื่อของยีนที่ระบุได้จะนำมาใช้เพื่อทำการวิเคราะห์เครือข่ายของยีนโดยใช้แหล่งข้อมูลออนไลน์ GeneMania (Warde-Farley et al., 2010) ส่วนการวิเคราะห์หน้าที่และกลไกทางชีววิทยาของยีนที่มีอิทธิพลต่อลักษณะที่ศึกษาจะดำเนินการภายใต้ฐานข้อมูลสาธารณะ DAVID (The Database for Annotation, Visualization and Integrated Discovery (Huang et al., 2009a, 2009b) เลือกการแปลผลโดยใช้ Gene ontology (Go) ซึ่งประกอบด้วย 3 กระบวนการ คือ กระบวนการทางชีวภาพ (Biological processes) องค์ประกอบของเซลล์ (Cellular components) และหน้าที่ระดับโมเลกุล (Molecular function)

ผลการศึกษาและวิจารณ์

การจำแนกบริเวณดีเอ็นเอบนจีโนม และการระบุยีนที่เกี่ยวข้อง

การศึกษาความสัมพันธ์เชื่อมโยงในจีโนมที่มีการถ่วงน้ำหนักแบบขั้นตอนเดียว (WssGBLUP) ในครั้งนี้ ทำให้เราสามารถตรวจสอบและระบุหาบริเวณดีเอ็นเอบนจีโนม และใช้บริเวณดีเอ็นเอบนจีโนมที่ได้ระบุหาตำแหน่งของยีนที่เกี่ยวข้องกับลักษณะของสุขภาพเต้านม คือ ค่า SCS โดยใช้แผน SNP ความหนาแน่นปานกลางในประชากรโคนมไทยสำหรับทุกโครโมโซมร่างกาย กราฟแมนฮัตตันได้แสดงร้อยละความแปรปรวนทางพันธุกรรมของเครื่องหมาย SNPs ที่อธิบายโดยมีบริเวณดีเอ็นเอบนจีโนมที่ไม่ทับซ้อนกันและประกอบด้วย 20 SNP ตาม Figure 1 และได้สรุปบริเวณดีเอ็นเอบนจีโนมที่อธิบายสัดส่วนของความแปรปรวนทางพันธุกรรมที่มากที่สุด (ร้อยละ 0.5 หรือมากกว่า) และยีนที่สำคัญเกี่ยวข้องกับลักษณะที่ศึกษาไว้ใน Table 2 บริเวณดีเอ็นเอบนจีโนมสามารถอธิบายความแปรปรวนทางพันธุกรรมได้มากถึง 0.6% อย่างไรก็ตามบริเวณดีเอ็นเอบนจีโนมส่วนใหญ่อธิบายได้น้อยกว่า 0.5% (Figure 1) และบริเวณเหล่านี้กระจายไปทั่วทั้งจีโนม ซึ่งบ่งชี้ว่าลักษณะนี้เป็นลักษณะที่ควบคุมด้วยยีนหลายยีน (polygenic) จากทั่วทั้งจีโนม ซึ่งมีส่วนร่วมสำคัญที่ทำให้เกิดการแปรปรวนทางพันธุกรรม บริเวณดีเอ็นเอบนจีโนมหลักที่เกี่ยวข้องกับ SCS พบอยู่บนโครโมโซม BTA 11, BTA 16 และ BTA 21

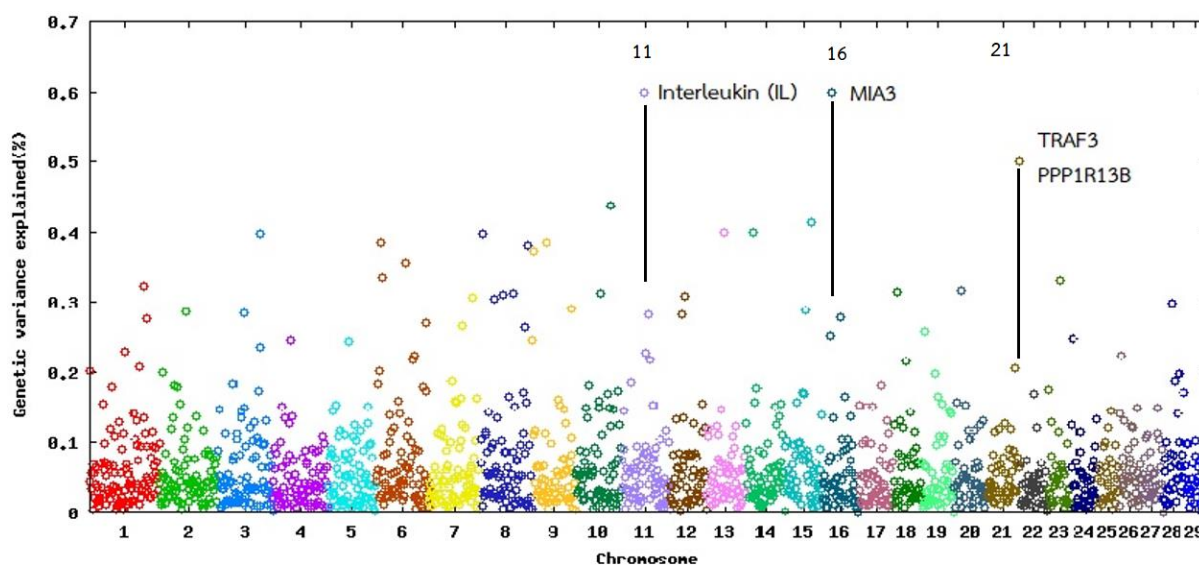


Figure 1 Manhattan plots of the additive genetic variance explained by windows of 20 adjacent SNPs for SCS in Thai dairy cattle

Table 2 Summary of the windows that explained the most of genetic variance for somatic cell score in Thai dairy cattle, with a list of annotated genes in the proximity of each window

Window Regions (nonoverlapping)	Var (%) ^{1/}	Chr	Start, bp	Stop (bp)	Genes ^{2/}
27928-27947	0.6	16	26671488	27692042	TRNAT-UGU , HHLPL2 , TAF1A , MIA3 , AIDA, BROX, FAM177B, DISP1, LOC100336644, TLR5, SUSP4, CCDC185, CAPN8
20925-20944	0.6	11	45736640	46794995	NCK2 , LOC101903923, LOC786288, TTL, POLR1B, CHCHD5, LOC107132936, SLC20A1, LOC101903687, NT5DC4, CKAP2L, IL1A , LOC107132933, LOC101904177, IL1B , IL37 , LOC104973371, IL36G , IL36A , IL36B , IL36RN , IL1F10 , IL1RN , PSD4, PAX8
34820-34839	0.5	21	68751694	70050042	LOC100847464, TRNAI-AAU, LOC787175, TRNAC-GCA , ZNF839, CINP, LOC101902499, TECPR2, ANKRD9, RCOR1 , LOC783220, TRAF3 , AMN, CDC42BPB , LOC101902932, LOC104975451, EXOC3L4 , LOC615559, TNFAIP2 , LOC100301250, LOC107131330, LOC509029, LOC100139514, EIF5, MARK3, CKB, TRMT61A, BAG5, APOPT1, KLC1, XRCC3, ZFYVE21, PPP1R13B

^{1/}Genomic variance absorbed by 20-SNP moving windows obtained using single-step genomic-BLUP^{2/}Any genes with start and stop positions within the window were considered. Genes linked to SCS are in bold font while novel candidate genes are bold and underlined

บริเวณดีเอ็นเอบนจีโนมที่เกี่ยวข้อง และมีอิทธิพลสูงต่อ SCS ตั้งอยู่บน 3 โครโมโซม ได้แก่ BTA 16 (จาก 26,671,488 ถึง 27,692,042 bp) BTA 11 (จาก 45,736,640 ถึง 46,794,995 bp และ BTA 21 (จาก 68,751,694 ถึง 70,050,042 bp) ซึ่งพบว่ามี 52 ยีนที่เคยมีรายงาน และ 19 ยีนที่ยังไม่ได้กำหนดบทบาทหน้าที่ที่ชัดเจน (ยีนที่ขึ้นต้นด้วย LOC) (Figure 1 และ Table 2) จำนวนยีนที่พบในบริเวณดีเอ็นเอบนจีโนมเหล่านี้เมื่อเปรียบเทียบกับผลลัพธ์กับฐานข้อมูล (NCBI, Genecards และ UniProt) พบว่ามี 21 ยีนที่มีรายงานว่าเกี่ยวข้องกับ SCS

เซลล์โซมาติกในน้ำนมประกอบด้วยเซลล์เยื่อบุเต้านม และเซลล์ในระบบภูมิคุ้มกัน (Alhussien and Dang, 2018) โดยการศึกษาครั้งนี้ พบว่ามีการระบุยีนที่รู้จักว่าเกี่ยวข้องกับภูมิคุ้มกัน การอักเสบ หรือการเพิ่มจำนวนเซลล์ในบริเวณดีเอ็นเอจีโนมที่เกี่ยวข้องกับลักษณะ SCS ยีนเหล่านี้ ได้แก่ Melanoma Inhibitory Activity Family, Member 3 (**MIA3**) ซึ่งทำหน้าที่ควบคุมองค์ประกอบภายนอกเซลล์ที่แสดงออกในเนื้อเยื่อเต้านมภายหลังการติดเชื้อแบคทีเรีย (Bergstralh et al., 2007, Rossetti et al., 2016) ยีนในกลุ่ม Interleukin superfamily เช่น Interleukin 1, alpha (**IL1A**), interleukin 1, beta (**IL1B**), interleukin 1 family, member 10 (**IL1F10**), interleukin 36, alpha (**IL36A**), interleukin 36, beta (**IL36B**), Interleukin 36, gamma (**IL36G**) และ interleukin 37 (**IL37**) และตัวรับไซโตไคน์ IL1 เช่น interleukin 1 receptor antagonist (**IL1RN**) และ interleukin 36 receptor antagonist (**IL36RN**) ถูกตรวจ

พบบนโครโมโซม BTA 11 ยีนเหล่านี้มีบทบาทสำคัญในการตอบสนองของภูมิคุ้มกันชนิดเจาะจง และกระบวนการการอักเสบ (Benveniste, 2014; Queen et al., 2019) REST corepressor 1 (RCOR1) เกี่ยวข้องกับการทำหน้าที่ในการเปลี่ยนแปลงของเซลล์เม็ดเลือดขาวลิมโฟไซต์ชนิดบี และการตอบสนองต่อการอักเสบ (Yao et al., 2015; Xiong et al., 2020) TNF receptor-associated factor 3 (TRAF3) มีบทบาทสำคัญในกลไกทางชีววิทยาของการสร้างภูมิคุ้มกันในเซลล์เยื่อบุผิวของเต้านมของโค (Song et al., 2017) CDC42 binding protein kinase beta (CDC42BPB), exocyst complex component 3-like 4 (EXOC3L4) และ tumor necrosis factor, alpha-induced protein 2 (TNFAIP2) มีส่วนเกี่ยวข้องกับการขนส่งภายในเซลล์ และก่อนหน้านี้นี้มีการตรวจพบในบริเวณจีโนมที่เกี่ยวข้องกับระบบภูมิคุ้มกันแต่กำเนิดในโคมนโฮสต์โตของแคนาดา (Porat-Shliom et al., 2013; de Klerk et al., 2018) protein phosphatase 1, regulatory subunit 13B (PPP1R13B) เป็นตัวกระตุ้นการทำงานของ p53 ซึ่งเป็นตัวควบคุมในระบบภูมิคุ้มกัน และกระบวนการตายของเซลล์ (Mitchell et al., 2002; Liu et al., 2005) และ TATA box binding protein (TBP)-associated factor, RNA polymerase I, A (TAF1A) เป็นยีนที่ควบคุมการเพิ่มจำนวนของเซลล์ และมีส่วนเกี่ยวข้องกับการกระตุ้น (Bergstrahl et al., 2007, Rossetti et al., 2016)

นอกจากนี้ ในกลุ่มประชากรที่ศึกษาซึ่งตรวจพบยีนอื่น ๆ ที่มีความเชื่อมโยงกับ SCS แต่พบว่ามีความสัมพันธ์เกี่ยวข้องกับปริมาณและองค์ประกอบน้ำนมในการศึกษาอื่น ๆ ยกตัวอย่าง เช่น transfer RNA threonine (anticodon UGU) (TRNAT-UGU), transfer RNA cysteine (anticodon GCA) (TRNAC-GCA), HHIP-like 2 (HHIPL2) และ NCK adaptor protein 2 (NCK2) (Venturini et al., 2014; da Costa Barros et al., 2018; Johnston et al., 2018; Oliveira et al., 2018)

จากผลการวิจัยครั้งนี้พบว่ามียีนที่มีอิทธิพลต่อลักษณะสุขภาพเต้านมจำนวน 21 ยีน (ยีนตัวหนาใน Table 2) ซึ่งบ่งชี้ว่าลักษณะนี้เป็นลักษณะที่ควบคุมด้วยยีนหลายยีน โดยมียีนหลายยีนที่พบเหมือนกับรายงานอื่น ๆ ก่อนหน้านั้นในโคมนโฮสต์โตที่เกี่ยวข้องกับลักษณะ SCS ทั้งนี้เนื่องจากประชากรโคมนของไทยมีความเกี่ยวข้องกับประชากรโคมนโฮสต์โต อย่างไรก็ตามยีนบางตัวมีความแตกต่างจากโคมนสายพันธุ์อื่น ๆ ทั้งนี้เนื่องจากความแตกต่างของประชากรที่ศึกษาที่มีอิทธิพลต่อการแสดงออกของยีน (Oliveira et al., 2019) จากการศึกษาครั้งนี้มีข้อจำกัดเนื่องจาก Bovine UMD 3.1 มีการถอดรหัสพันธุกรรมจากโคเนื้อพันธุ์เฮิร์ฟฟอร์ดซึ่งมีความแตกต่างจากจีโนมของโคมนโฮสต์โต ยกตัวอย่างเช่น การแทรกและหายไป (insertion/deletion: INDEL) ของลำดับเบส มีความแตกต่างกันถึง 48,537,190 เบสจากจีโนมทั้งหมด (Koks et al., 2013) ซึ่งอาจส่งผลให้เกิดความคลาดเคลื่อนสำหรับการวิเคราะห์ข้อมูลจีโนมเพื่อหาตำแหน่ง ทำนายหน้าที่ของยีน และการระบุยีนที่เชื่อมโยงได้ ดังนั้นจึงควรหาตำแหน่ง และระบุยีนจากฐานข้อมูลจีโนมของประชากรในประเทศไทยเอง

การวิเคราะห์เครือข่ายและกลไกทางชีววิทยาของยีน

การวิเคราะห์เครือข่ายของยีน (network analysis) จากรายชื่อยีนที่ควบคุมลักษณะสุขภาพเต้านมโดยใช้โปรแกรม GeneMania ตาม Figure 2 พบว่ามีเครือข่ายยีนที่แสดงออกร่วมกันจำนวน 29 ยีน (26.97%) และยีนที่มีปฏิสัมพันธ์ร่วมกันจำนวน 141 รายการ ยีนเหล่านี้ส่วนใหญ่เกี่ยวข้องกับกลไกทางชีววิทยาระบบภูมิคุ้มกันและการแบ่งเซลล์ เช่น A) กลไกทางชีววิทยาที่เกี่ยวข้องกับคอมพลีเมนต์ C3 (C3 complement pathway) ซึ่งมีความเกี่ยวข้องกับการเก็บกินแบคทีเรีย B) โครงสร้างเชิงซ้อนที่เกี่ยวข้องกับการขนส่งสารผ่านเยื่อหุ้มเซลล์แบบ Exocytosis (Exocyst complex) C) กลไกทางชีววิทยา I-kappaB kinase/NF-kappaB signaling ที่มีความเกี่ยวข้องกับระบบภูมิคุ้มกันหลายอย่าง เช่น กระบวนการอักเสบ การอยู่รอดของเซลล์ การพัฒนาของระบบภูมิคุ้มกัน การแบ่งเซลล์ เป็นต้น กลไกทางชีววิทยาอื่นๆที่พบเช่น D) การแบ่งเซลล์ (Cell proliferation) และ E) การยับยั้งการเคลื่อนที่และการบุกรุกของเซลล์มะเร็ง (Inhibition of cell migration and invasion)

Table 3 แสดงกลไกทางชีววิทยาของกลุ่มยีนสำหรับ SCS โดยใช้ยีนที่ระบุได้ตามเกณฑ์การอธิบายความแปรปรวนทางพันธุกรรมทั้งหมดของแต่ละบริเวณดีเอ็นเอบนจีโนม (Table 2) กลไกทางชีววิทยาส่วนใหญ่ยีนย่นการค้นพบยีนที่เกี่ยวข้องกับระบบภูมิคุ้มกัน (immunological mechanism) กระบวนการอักเสบ (inflammation) และการติดเชื้อ (infection) เช่น กลไกการส่งสัญญาณที่ใช้ไซโตไคน์ที่เป็นสื่อกลาง (cytokine-mediated signaling pathway, GO: 0019221) การควบคุมเชิงบวกของการผลิตอินเตอร์ลิวคิน-6 (positive

regulation of interleukin-6, GO: 0032755) การทำให้เกิดไข้ (fever generation, GO: 0001660) การขนส่งสารออกนอกเซลล์ (exocytosis, GO: 0006887) การควบคุมเชิงลบของกลไกการส่งสัญญาณที่ใช้ไซโตไคน์ที่เป็นสื่อกลาง (negative regulation of cytokine-mediated signaling pathway, GO: 0001960) การตอบสนองของระบบภูมิคุ้มกัน (immune response, GO: 0006955) กลไกการตายของเซลล์แบบภายใน (intrinsic apoptotic signaling pathway, GO: 0097193) กลไกการส่งสัญญาณผ่านตัวรับ toll-like receptor (toll-like receptor signaling pathway, GO: 0002224) การควบคุมเชิงบวกของสัญญาณ I-kappaB kinase/NF-kappaB (positive regulation of I-kappaB kinase/NF-kappaB signaling, GO: 0043123) การจับตัวรับ interleukin-1 (interleukin-1 receptor binding, GO: 0005149) และกิจกรรมไซโตไคน์ (cytokine activity, GO: 0005125) กลไกทางชีววิทยาอื่น ๆ เกี่ยวข้องกับการแบ่งเซลล์ (positive regulation of cell division, GO: 0051781) การถอดรหัสพันธุกรรม (transcription from RNA polymerase I promoter, GO: 0006360) และกระบวนการของเซลล์ (cellular process)(extracellular space, GO: 0005615; exocyst, GO: 0000145; ARF guanyl-nucleotide exchange factor activity, GO: 0005086)

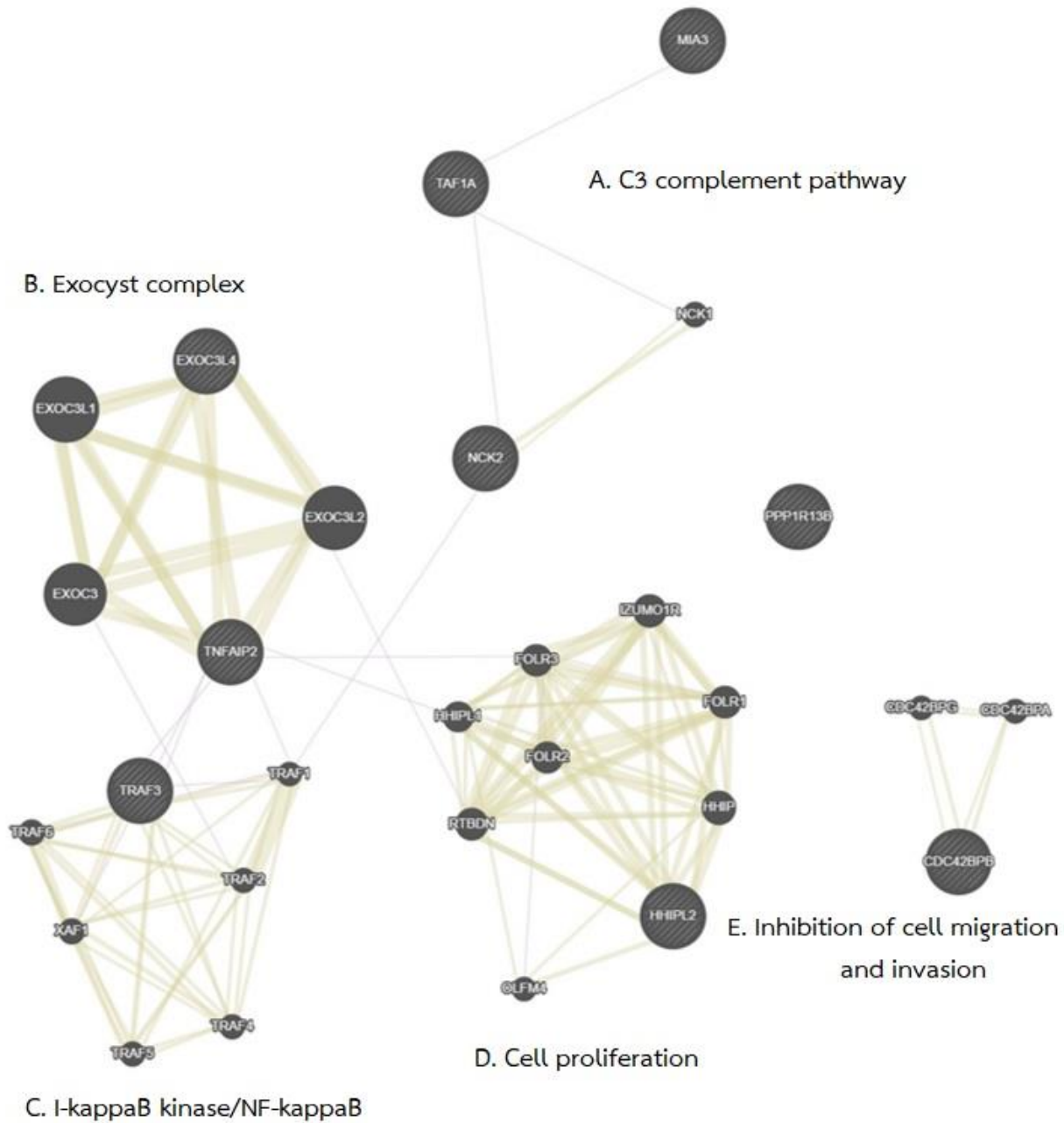


Figure 2 Gene network produced using GeneMANIA for udder health. The network consists of candidate genes and related genes (29 genes, circles) connected by 141 interactions (edges).

Table 3 Pathways identified by the Database for Annotation, Visualization and Integrated Discovery (DAVID version 6.7) in gene ontology (GO) term

Items	Gene count	P -value	GENES
GO 0019221: cytokine-mediated signaling pathway	9	5.30E-10	IL1A, IL1B, IL1F10, IL1RN, IL36RN, IL36A, IL36B, IL36G, IL37
GO 0032755: positive regulation of interleukin-6 production	3	5.10E-03	IL1A, IL36A, IL36B
GO 0001660: fever generation	2	5.30E-03	IL1A, IL1B
GO 0006887: exocytosis	3	6.70E-03	MIA3, TNFAIP2, EXOC3L4
GO 0001960: negative regulation of cytokine-mediated signaling pathway	2	1.60E-02	IL1RN, IL36RN
GO 0051601: exocyst localization	2	1.90E-02	TNFAIP2, EXOC3L4
GO 0006955: immune response	4	2.00E-02	IL1A, IL1B, IL36A, IL36B
GO 0097193: intrinsic apoptotic signaling pathway	2	3.70E-02	APOPT1, PPP1R13B
GO 0006360: transcription from RNA polymerase I promoter	2	3.70E-02	POLR1B, TAF1A
GO 0002224: toll-like receptor signaling pathway	2	3.70E-02	TRAF3, TLR5
GO 0043123: positive regulation of I-KappaB kinase/NF-kappaB signaling	3	4.50E-02	IL1A, IL1B, SLC20A1
GO 0051781: positive regulation of cell division	2	6.50E-02	IL1A, IL1B
GO 0005615: extracellular space	11	9.10E-05	AMN, CKB, IL1A, IL1B, IL1F10, IL1RN, IL36RN, IL36A, IL36B, IL36G, IL37
GO 0000145: exocyst	2	3.70E-02	TNFAIP2, EXOC3L4
GO 0005149: interleukin-1 receptor binding	9	5.40E-19	IL1A, IL1B, IL1F10, IL1RN, IL36RN, IL36A, IL36B, IL36G, IL37
GO 0005125: cytokine activity	9	1.30E-09	IL1A, IL1B, IL1F10, IL1RN, IL36RN, IL36A, IL36B, IL36G, IL37
GO 0005086: ARF guanyl-nucleotide exchange factor activity	2	5.50E-02	TNFAIP2, EXOC3L4
GO 0000149: SNARE binding	2	9.40E-02	TNFAIP2, EXOC3L4

สรุป

เมื่อทำการศึกษาความสัมพันธ์ทั่วทั้งจีโนมโดยใช้บันทึกวันทดสอบสำหรับลักษณะสุขภาพเต้านม (SCS) ในประชากรโคนมไทยโดยรวมแล้วบริเวณดีเอ็นเอบนจีโนมที่สำคัญ และเกี่ยวข้องกับ SCS พบอยู่บนโครโมโซม BTA 11, BTA 16 และ BTA 21 ซึ่งมีทั้งยีนที่เคยมีรายงาน และยีนที่ยังไม่ได้กำหนดบทบาทหน้าที่ที่ชัดเจน (ยีนที่ขึ้นต้นด้วย LOC) เมื่อเปรียบเทียบผลลัพธ์กับฐานข้อมูลพบว่า มี 21 ยีนที่มีรายงานเกี่ยวข้องกับภูมิคุ้มกัน การอักเสบ หรือการเพิ่มจำนวนเซลล์ ที่มีอิทธิพลต่อ SCS อย่างมีนัยสำคัญ เช่น ยีนในกลุ่ม Interleukin superfamily (IL1A, IL1B, IL1F10, IL36A, IL36B, IL36G, IL37) ตัวรับไซโตไคน์ IL1 (IL36RN, IL1RN) รวมทั้งยีนอื่น ๆ (MIA3, RCOR1, TRAF3, CDC42BPB, EXOC3L4, TNFAIP2, PPP1R13B, TAF1A, TRNAT-UGU, TRNAC-GCA, HHIPL2, NCK2 เครือข่ายและกลไกทางชีววิทยาของยีนสำหรับ SCS ส่วนใหญ่ยืนยันการค้นพบยีนที่เกี่ยวข้องกับระบบภูมิคุ้มกัน กระบวนการอักเสบ การติดเชื้อ และการแบ่งเซลล์ ยีนที่ระบุได้จากการศึกษานี้ไม่เพียงแต่จะช่วยยืนยันสำหรับการค้นพบที่มีมาก่อนหน้านี้ แต่ยังสำรวจพบยีนชุดใหม่ที่เกี่ยวข้องกับ SCS ซึ่งจะนำไปสู่ความรู้เกี่ยวกับข้อมูลพื้นฐานทางโครงสร้างพันธุกรรมของประชากรของลักษณะที่ศึกษา และสามารถใช้เป็นยีนเป้าหมายในการศึกษาการแสดงออกของยีน

อย่างไรก็ตามเพื่อตรวจสอบความถูกต้องของการค้นพบในครั้งนี้ จำเป็นจะต้องมีการวิจัยเพิ่มเติมโดยอาศัยข้อมูลที่มากขึ้นทั้งในส่วนของจำนวนสัตว์ ฟีนไทป์ และจีโนไทป์

เอกสารอ้างอิง

- สายัณห์ บัวบาน เกียรติศักดิ์ เหลืองหนูดำ และมนต์ชัย ดวงจินดา. 2563. การประมาณค่าพารามิเตอร์ทางพันธุกรรม และค่าแนวโน้มทางพันธุกรรมสำหรับลักษณะความต้านทานเต้านมอักเสบของแม่โคนมไทยโดยใช้ข้อมูลในวันทดสอบ. แท่นเกษตร. 48 (1): 23-34.
- Aguilar, I., I. Misztal, D. Johnson, A. Legarra, S. Tsuruta, and T. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal of Dairy Science*. 93: 743–752.
- Aguilar, I., I. Misztal, S. Tsuruta, A. Legarra, and H. Wang. 2014. PREGSF90-POSTGSF90: Computational tools for the implementation of single-step genomic selection and genome-wide association with ungenotyped individuals in BLUPF90 programs. Pages 1-3 in *Proceedings of the 10th World Congress on Genetics Applied to Livestock Production.*, Vancouver, BC, Canada. American Society of Animal Science, Champaign, IL.
- Alhussien, M. N., and A. K. Dang. 2018. Milk somatic cells, factors influencing their release, future prospects, and practical utility in dairy animals: An overview. *Veterinary World*. 11: 562-577.
- Ali, A.K. A., and G. E. Shook. 1980. An optimum transformation for somatic cell concentration in milk. *Journal of Dairy Science*. 63: 487-490.
- Benveniste, E. N. 2014. Cytokines. Pages 921-925 in *Encyclopedia of the Neurological Sciences*. 2nd ed. Academic Press, Oxford.
- Bergstrahl, D. T., B. J. Conti, C. B. Moore, W. J. Brickey, D. J. Taxman, and J. P. Y. Ting. 2007. Global functional analysis of nucleophosmin in Taxol response, cancer, chromatin regulation, and ribosomal DNA transcription. *Experimental Cell Research*. 313: 65-76.
- da Costa Barros, C., D. J. de Abreu Santos, R. R. Aspilcueta-Borquis, G. M. F. de Camargo, F. R. de Araújo Neto, and H. Tonhati. 2018. Use of single-step genome-wide association studies for prospecting genomic regions related to milk production and milk quality of buffalo. *Journal of Dairy Research*. 85: 402-406.

- Daniele, B. D. M., J. W. M. Marques, M. L. Bastiaansen, W. J. Broekhuijse, M. S. Lopes, E. F. Knol, B. Harlizius, S. E. F. Guimarães, F. F. Silva, and P. S. Lopes. 2018. Weighted single-step GWAS and gene network analysis reveal new candidate genes for semen traits in pigs. *Genetics Selection Evolution*. 50: 40.
- De Klerk, B., M. Emam, K. A. Thompson-Crispi, M. Sargolzaei, J. J. van der Poel, and B. A. Mallard. 2018. A genome-wide association study for natural antibodies measured in blood of Canadian Holstein cows. *BMC Genomics*. 19: 694.
- Falconer, D.S., and T.F.C. Mackay. 1996. *Introduction to quantitative genetics*. 4th ed. Addison Wesley Longman Ltd, Essex, England.
- Fan B., S.K. Onteru, Z.Q. Du, D.J. Garrick, K.J. Stalder, and M.F. Rothschild. 2011. Genome-wide association study identifies loci for body composition and structural soundness traits in pigs. *PloS one*. 6: e14726.
- Fragomeni, B. O., I. Misztal, D. L. Lourenco, I. Aguilar, R. Okimoto, and W. M. Muir. 2014. Changes in variance explained by top SNP windows generations for three traits in broiler chicken. *Frontiers in Genetics*. 5: 332.
- Garcia, A. L. S., B. Bosworth, G. Waldbieser, I. Misztal, S. Tsuruta, and D. A. L. Lourenco. 2018. Development of genomic predictions for harvest and carcass weight in channel catfish. *Genetics Selection Evolution*. 50: 66.
- Gengler, N., A. Tijani, G. R. Wiggins, and I. Misztal. 1999. Estimation of (co)variance function coefficients for test day yield with an expectation maximization restricted maximum likelihood algorithm. *Journal of Dairy Science*. 82: 1849-1871.
- Hayes, B. J., and M. Goddard. 2010. Genome-wide association and genomic selection in animal breeding. *Genome*. 53(11): 876-83.
- Hirschhorn J.N., and M.J. Daly. 2005. Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics*. 6: 95-108.
- Howard, J. T., S. Shihui Jiao, F. Tiezzi, Y. Huang, K. A. Gray, and C. Maltecca. 2015. Genome-wide association study on legendre random regression coefficients for the growth and feed intake trajectory on Duroc Boars. *BMC Genetic*. 16: 59.
- Huang, W., B. T. Sherman, and R. A. Lempicki. 2009a. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protocols*. 4: 44-57.
- Huang, W., B. T. Sherman, and R. A. Lempicki. 2009b. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*. 37: 1-13.
- Irano N., G. M. de Camargo, R. B. Costa, A. P. Terakado, A. F. Magalhaes, R. M. Silva, M. M. Dias , A. B. Bignardi, F. Baldi, R. Carneiro, H. N., and de L. G. Albuquerque. 2016. Genome-wide association study for indicator traits of sexual precocity in Nellore cattle. *PloS one*. 11: e0159502.
- Jamrozik, J., L. R. Schaeffer, and J. C. M. Dekkers. 1997. Genetic evaluation of dairy cattle using test day yields and random regression model. *Journal of Dairy Science*. 80: 1217-1226.
- Johansson A.M., M.E. Pettersson, P.B. Siege, and O. Carlborg. 2010. Genome-wide effects of long-term divergent selection. *PloS Genetics*. 6: e1001188.
- Johnston, D., I. Malo Estepa, H. A. Ebhardt, M. A. Crowe, and M. G. Diskin. 2018. Differences in the bovine milk whey proteome between early pregnancy and the estrous cycle. *Theriogenology*. 114: 301-307.

- Kang, H., L. Zhou, R. Mrode, Q. Zhang, and J. F. Liu. 2017. Incorporating the single-step strategy into a random regression model to enhance genomic prediction of longitudinal traits. *Heredity*. 119: 459-467.
- Koks, S., R. Lilleoja, E. Reimann, A. Salumets, P. Reemann, and Ü. Jaakma. 2013. Sequencing and annotated analysis of the Holstein cow genome. *Mammalian Genome*. 24: 309-321.
- Lee, S., C. Dang, Y. Choy, C. Do, K. Cho, J. Kim, Y. Kim, and J. Lee. 2019. Comparison of genome-wide association and genomic prediction methods for milk production traits in Korean Holstein cattle. *Asian-Australasian Journal of Animal Sciences*. 32: 913-921.
- Legarra A., C. Robert-Granié, E. Manfredi, and J.M. Elsen. 2008. Performance of genomic selection in mice. *Genetics*. 180: 611-618.
- Legarra, A., D.A. Lourenco, and Z. Vitezica. 2018. Bases for genomic prediction 2018. Available: <http://genow eb.toulo use.inra.fr/~alega rra/GSIP.pdf>. Accessed Oct. 10, 2018.
- Lemos, M. V. A., H. L. J. Chiaia, M. P. Berton, F. L. B. Feitosa, C. Aboujaoude, G. M. F. Camargo, A. S. C. Pereira, L. G. Albuquerque, A. M. Ferrinho, L. F. Mueller, M. R. Mazalli, J. J. M. Furlan, R. Carneiro, D. G. M. Gordo, R. Tonussi, R. Espigolan, R. M. O. Silva, H. N. Oliveira, S. Duckett, I. Aguilar, and F. Baldi. 2016. Genome-wide association between single nucleotide polymorphisms with beef fatty acid profile in Nellore cattle using the single step procedure. *BMC Genomics*. 17: 213.
- Liu, Z.J., X. Lu, and S. Zhong. 2005. ASPP-Apoptotic specific regulator of p53. *Biochimica et Biophysica Acta*. 1756: 77-80.
- Lourenco, D. A. L., B. O. Fragomeni, H. L. Bradford, I. R. Menezes, J. B. S. Ferraz, and I. Aguilar. 2017. Implications of SNP weighting on single-step genomic predictions for different reference population sizes. *Journal of Animal Breeding and Genetics*. 134: 463-71.
- Melo, T. P., G. M. F. de Camargo, L. G. de Albuquerque, and R. Carneiro. 2017. Genome-wide association study provides strong evidence of genes affecting the reproductive performance of Nellore beef cows. *PloS one*. 12: e0178551.
- Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 157: 1819-1829.
- Misztal, I., S. Tsuruta, D. A. L. Lourenco, Y. Masuda, I. Aguilar, A. Legarra, and Z. Vitezica. 2018. Manual for BLUPF90 family of programs. Vol. 2018 Animal and Dairy Science Department, Georgia. Available: http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all7.pdf. Accessed February 1, 2018
- Mitchell, R. A., H. Liao, J. Chesney, G. Fingerle-Rowson, J. Baugh, J. David, and R. Bucala. 2002. Macrophage migration inhibitory factor (MIF) sustains macrophage proinflammatory function by inhibiting p53: Regulatory role in the innate immune response. *Proceedings of the National Academy of Sciences of the United States of America*.
- Oliveira H. R., F. F. Silva, L. F. Brito, J. Jamrozik, D. A. L. Lourenco, and F. S. Schenkel. 2018. Genome-wide association study for milk, fat and protein yields in different lactation stages in Canadian Holstein and Jersey cattle. No. 11.601. *Proceedings of the 11th World Congress on Genetics Applied to Livestock Production.*, Auckland, New Zealand.

- Oliveira, H. R., D. A. L. Lourenco, Y. Masuda, I. Misztal, S. Tsuruta, J. Jamrozik, L. F. Brito, F. F. Silva, J. P. Cant, and F. S. Schenkel. 2019. Single-step genome-wide association for longitudinal traits of Canadian Ayrshire, Holstein, and Jersey dairy cattle. *Journal of Dairy Science*. 102: 9995-10011.
- Oliveira, H. R., F. F. Silva, M. V. G. B. Silva, O. H. G. B. D. Siqueira, M. A. Machado, J. C. C. Panetto, L. S. Gloria, and L. F. Brito. 2017. Bayesian models combining Legendre and B-spline polynomials for genetic analysis of multiple lactations in Gyr cattle. *Livestock Science*. 201: 78–84.
- Porat-Shliom, N., O. Milberg, A. Masedunskas, and R. Weigert. 2013. Multiple roles for the actin cytoskeleton during regulated exocytosis. *Cellular and Molecular Life Sciences*. 70: 2099-2121.
- Queen, D., C. Ediriweera, and L. Liu. 2019. Function and regulation of IL-36 signaling in inflammatory diseases and cancer development. *Frontiers in Cell and Developmental Biology*. 7: 317-317.
- Rossetti, S., A. J. Wierzbicki, and N. Sacchi. 2016. Mammary epithelial morphogenesis and early breast cancer. Evidence of involvement of basal components of the RNA Polymerase I transcription machinery. *Cell Cycle*. 15: 2515-2526.
- Rupp R., and D. Boichard. 2003. Genetics of resistance to mastitis in dairy cattle. *Veterinary Research*. 34: 671-688.
- Seegers, H., C. Fourichon, and F. Beaudeau. 2003. Production effects related to mastitis and mastitis economics in dairy cattle herds. *Veterinary Research*. 34: 475–491.
- Silva, M.V.B., D.J.A. dos Santos, S.A. Boison, A.T.H. Utsunomiya, A.S. Carmo, T.S. Sonstegard, J.B. Cole, and C.P. Van Tassell. 2014. The development of genomics applied to dairy breeding. *Livestock Science*. 166: 66–75.
- Silva, R. M. O., N. B. Stafuzza, B. O. Fragomeni, G. M. F. Camargo, T. M. Ceacero, J. N. S. G. Cyrillo, F. Baldi, A. A. Boligon, M. E. Z. Mercadante, D. L. Lourenco, I. Misztal, and L. G. Albuquerque. 2017. Genome-wide association study for carcass traits in an experimental Nelore cattle population. *PloS one*. 12: e0169860.
- Silva, R. P., M. P. Berton, L. Grigoletto, F. E. Carvalho, R. M. O. Silva, E. Peripolli, L. M. Castro, J. B. S. Ferraz, J. P. Eler, R. B. Lobo, and F. Baldi. 2019. Genomic regions and enrichment analyses associated with carcass composition indicator traits in Nellore cattle. *Journal of Animal Breeding and Genetics*. 136: 118-133.
- Snelling W. M., R. Cushman, J. W. Keele, C. Maltecca, M. G. Thomas, M. R. S. Fortes, and A. Reverter. 2013. Networks and pathways to guide genomic selection. *Journal of Animal Science*. 91: 537-552.
- Tiezzi, F., K. L. Parker-Gaddis, J. B. Cole, J. S. Clay, and C. Maltecca. 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*. 10: e0114919.
- Valente, T. S., F. Baldi, A. C. Sant'Anna, L. G. de Albuquerque, and M. J. Paranhos da Costa. 2016. Genome-wide association study between single nucleotide polymorphisms and fight speed in Nellore cattle. *PloS one*. 11: e0156956.
- VanRaden P. M. 2008. Efficient methods to compute genomic predictions. *Journal of Dairy Science*. 91: 4414-4423.
- Venturini, G., D. Cardoso, F. Baldi, A. Freitas, R. Aspilcueta-Borquis, D. Santos, G. Camargo, N. Stafuzza, L. Albuquerque, and H. Tonhati. 2014. Association between single-nucleotide polymorphisms and milk production traits in buffalo. *Genetics and Molecular Research*. 13: 10256-10268.

- Wang H., I. Misztal, I. Aguilar, A. Legarra, R.L. Fernando, Z. Vitezica, R. Okimoto R, T. Wing, R. Hawken, and W.M. Muir. 2014. Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS)for 6-week body weight in broiler chickens. *Frontiers in Genetics*. 5: 1-8.
- Wang, H., I. Misztal, I. Aguilar, A. Legarra, and W. M. Muir. 2012. Genome-wide association mapping including phenotypes from relatives without genotypes. *Genetics Research*. 94: 73-83.
- Warde-Farley D., S. L. Donaldson, O. Comes, K. Zuberi, R. Badrawi, P. Chao, Franz M., C. Grouios, F. Kazi, C. T. Lopes, A. Maitland, S. Mostafavi, J. Montojo, Q. Shao, G. Wright, G.D. Bader, and Q. Morris. 2010. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Research*. 38: 214-220.
- Williams, T. and C. Kelley. 2019. User manual gnuplot version 5.2: An Interactive Plotting Program. Available: http://www.gnuplot.info/docs_5.2/Gnuplot_5.2.pdf. Accessed February 1, 2018.
- Wu, P., Q. Yang, K. Wang, J. Zhou, J. Ma, Q. Tang, L. Jin, W. Xiao, A. Jiang, Y. Jiang, L. Zhu, X. Li, and G. Tang. 2018. Single step genome wide association studies based on genotyping by sequence data reveals novel loci for the litter traits of domestic pigs. *Genomics*. 110: 171-179.
- Xiong, Y., L. Wang, E. Di Giorgio, T. Akimova, U. H. Beier, R. Han, M. Trevisanut, J. H. Kalin, P. A. Cole, and W. W. Hancock. 2020. Inhibiting the coregulator CoREST impairs Foxp3+ Treg function and promotes antitumor immunity. *Journal of Clinical Investigation*. 130: 1830-1842.
- Yao, H., D. C. Goldman, G. Fan, G. Mandel, and W. H. Fleming. 2015. The corepressor Rcor1 is essential for normal myeloerythroid lineage differentiation. *Stem Cells*. 33: 3304-3314.
- Zhang, Q., B. Guldbrandtsen, J. R. Thomasen, M. S. Lund, and G. Sahana. 2016. Genome-wide association study for longevity with whole genome sequencing in 3 cattle breeds. *Journal of Dairy Science*. 99: 7289-7298.
- Zhang, Z., J. Liu, X. Ding, P. Bijma, D. J. de Koning, and Q. Zhang. 2010. Best linear unbiased prediction of genomic breeding values using a trait-specific marker-derived relationship matrix. *PLoS One*. 5: e12648.
- Zhou, C., C. Li, W. Cai, S. Liu, H. Yin, S. Shi, Q. Zhang, and S. Zhang. 2019. Genome-wide association study for milk protein composition traits in a Chinese Holstein population using a single-step approach. *Frontiers in Genetics*. 10: 72.