

STUDIO DI ASSOCIAZIONE GENICA PER IL CARATTERE CONTENUTO DI CELLULE SOMATICHE NEL LATTE DELLA RAZZA RENDENA

A cura di: Enrico Mancin¹, Beniamino Tuliozi¹, Nadia Guzzo² e Roberto Mantovani¹

¹Dipartimento di Agronomia, Animali, Alimenti, Risorse Naturali e Ambiente e ²Dipartimento di Biomedicina Comparata e Alimentazione – Università degli Studi di Padova

Premessa

La selezione genomica è oramai largamente praticata nel miglioramento genetico di molte specie e razze di interesse zootecnico grazie alla possibilità di utilizzare su larga scala la genotipizzazione degli animali (Misztal et al., 2020). La chiave di volta in questo processo è attribuibile all'utilizzo di marcatori molto numerosi presenti nel DNA di ciascun individuo e noti come SNP (single nucleotide polymorphism o polimorfismi a singolo nucleotide), uniformemente distribuiti sul genoma, facilmente identificabili con analisi di costo relativamente basso e completamente meccanizzabili e, soprattutto, per effetto della elevata numerosità, facilmente associabili a geni ad azione quantitativa o QTL, cioè geni che partecipano attivamente all'espressione di caratteri di importanza economico-zootecnica (latte, carne, fertilità, salute, longevità, etc.; Misztal et al., 2020). Con decine di migliaia di SNP, ben scelti per essere rappresentativi dell'intero genoma, ci si aspetta infatti sempre una vicinanza tra marcatore SNP e un particolare gene o frammento di DNA di interesse per la selezione. La selezione genomica basata sugli SNP (GAS/MAS) permette quindi la scelta degli animali potenzialmente miglioratori per uno o più caratteri tramite la misurazione del potenziale genetico individuale basato sulla quantificazione delle associazioni esistenti tra geni QTL e marcatori SNP (Misztal et al., 2020). Per determinare in maniera rapida ed efficiente il genotipo per migliaia di SNPs in molti individui si ricorre alla tecnologia del "DNA-chip", consistente in un supporto solido come vetro, plastica, o chip di silicio formanti un array (vettore) sulla cui superficie sono presenti da decine a centinaia di migliaia di sonde a DNA ciascuna delle quali corrisponde ad uno specifico marcatore SNP a cui si legherà il frammento di DNA dell'individuo. Nel corso del PSRN (Dualbreeding) è stato possibile realizzare numerosi campioni biologici e ottenere profili genomici di parecchi maschi e femmine della razza Rendena. Alla luce della disponibilità di questi numerosi profili genomici per la razza Rendena, è stato condotto uno studio finalizzato a valutare e identificare possibili associazioni tra marcatori SNP ed il carattere inerente alla quantità di cellule somatiche nel latte di bovine di razza Rendena espresso come punteggio SCS (Somatic Cells Score). Ciò nell'intento di identificare possibili QTL coinvolti nell'espressione di questo importante carattere di salute mammaria (e di

AZ. 7 – IOV 7.1: Produzione report con studio su GAS-MAS

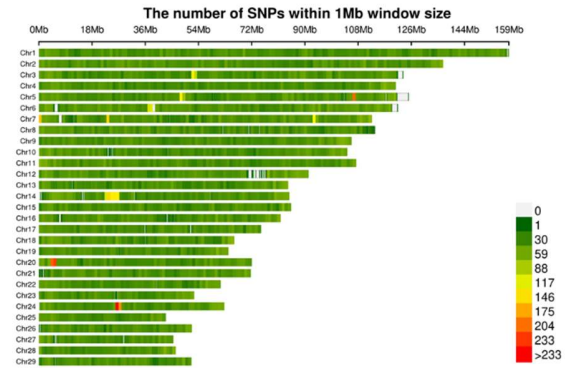
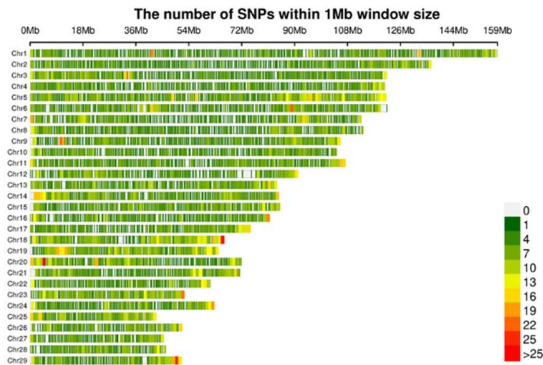
benessere) delle vacche di razza Rendena, con la finalità ultima di arrivare ad una possibile utilizzazione di questi QTL in processi di selezione di tipo GAS/MAS per la razza Rendena.

Dati e MetodiRilievi Genomici

I profili genomici iniziali comprendevano 1.253 vacche di razza Rendena genotipizzate a 33K (33.000 SNP circa; ottenuti con il chip a bassa densità GGP Bovine LD Array 33K di Illumina) e 720 tori genotipizzati in parte con chip a 33K (170 soggetti) e in gran parte (550 soggetti) con chip a più alta densità 150K (150.000 SNP circa; chip GGP Bovine HD Array 150K di Illumina). Inizialmente sono stati rimossi dal dataset gli animali con genotipizzazioni aventi un call-rate inferiore al 90% (inquinamento del campione biologico) e sono stati eliminati dai profili genomici tutti i marcatori SNP con call rate marcatore inferiore al 10% e con una Minor Allele Frequency (MAF) <0.05 . In entrambi i casi, infatti, la variante SNP identificata risulta poco presente o presente con bassa frequenza nel gruppo di animali analizzato, mettendo in luce, in definitiva, poche possibilità di poter avere stretti legami con geni ad azione quantitativa (QTL) legati al contenuto di cellule somatiche nel latte della razza Rendena. Quindi si è proceduto all'armonizzazione dei profili genomici ottenuti con differenti chip mediante imputazione dei marcatori SNPs per tutti i soggetti 33K utilizzando come riferimento il panel a 150K; questa operazione che è stata resa possibile dall'utilizzo delle relazioni di parentela tra tori e vacche sottoposti a genotipizzazione utilizzando allo scopo uno specifico programma (AlphaImpute2). Al termine del processo di imputazione sono stati infine rimossi tutti gli animali con bassa percentuale di imputazione, cioè con una soglia inferiore al 90% di call rate. Il dataset finale conteneva quindi 1.416 animali genotipizzati con una densità media di 113.280 SNP.

Figura 1: rappresentazione grafica del livello di copertura del genoma nei 29 cromosomi autosomici delle vacche genotipizzate: prima (a destra; 33K) e dopo (a sinistra; circa 113K) il processo di imputazione.

AZ. 7 – IOV 7.1: Produzione report con studio su GAS-MAS



Rilievi fenotipici

Il dataset inerente al contenuto di cellule somatiche nel latte consisteva in 13.402 record totali appartenenti a 862 a vacche aventi sia genotipo che fenotipo, mentre il pedigree conteneva 6.188 animali totali. Il contenuto di cellule somatiche/ml di latte è stato espresso canonicamente come SCS applicando la tradizionale formula:

$$3 + \log_2 (\text{cellule somatiche}/100)$$

Modello di associazione

Sono stati individuati tra tutti i marcatori SNP, quelli che presentavano un valore significativo di associazione fenotipica con il contenuto di cellule somatiche nel latte; questo valore è stato ottenuto mediante un valore soglia di probabilità-p-value ottenuto attraverso un test di Bonferroni prodotto dal metodo di analisi Single Step GBLUP (ssGBLUP; Aguilar et al., 2019).

Studi precedenti hanno dimostrato che ssGBLUP fornisce potenzialmente valori genomici (gEBV) più accurati e meno distorti rispetto agli alternativi metodi multistep, specialmente in presenza di piccole popolazioni e caratteri limitati dal sesso (Christensen e Lund, 2010).

Il modello matematico utilizzato in questo studio è stato il seguente:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Wp} + \mathbf{Za} + \mathbf{e}$$

dove \mathbf{y} è il vettore dei record SCS di una determinata vacca e \mathbf{X} , \mathbf{W} e \mathbf{Z} sono le matrici di incidenza degli effetti fissi, dell'effetto dell'ambiente permanente e gli effetti genetici, rispettivamente.

Il vettore \mathbf{b} include gli effetti fissi del giorno di controllo/allevamento per vacche dello stesso ordine di lattazione (massimo 3 lattazioni per bovina considerate), la classi di gestazione (in base ai giorni di gestazione al controllo), l'età al parto entro numero di lattazione ed il mese di parto entro lattazione. Entrambi questi due ultimi fattori hanno previsto l'uso di covariate basate sui coefficienti del polinomio

AZ. 7 – IOV 7.1: Produzione report con studio su GAS-MAS

di Legendre fino al terzo ordine. I vettori \mathbf{p} , \mathbf{a} , \mathbf{e} rappresentano i seguenti effetti casuali: l'ambiente permanente, la componente genetica additiva e l'errore residuo.

In ssGBLUP, gli effetti genetici additivi vengono campionati dalla distribuzione $\mathbf{a} \sim \mathbf{N}(\mathbf{0}, \mathbf{H}\sigma_a^2)$. La matrice \mathbf{H} contiene informazioni di parentela sia anagrafiche, sia genomiche. In ssGBLUP, la matrice inversa della struttura di (co)varianza dell'effetto additivo è rappresentato da \mathbf{H}^{-1} , descritta come:

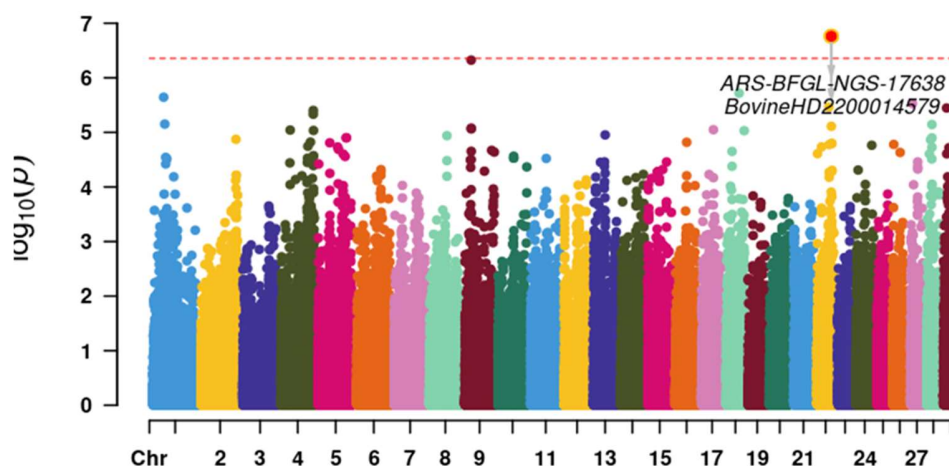
$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}$$

dove \mathbf{A}^{-1} e \mathbf{A}_{22}^{-1} sono gli inversi della matrice delle relazioni genealogiche rispettivamente per tutti gli animali e solo per gli animali genotipizzati. \mathbf{G}^{-1} è l'inversa della matrice di relazione genomica.

Risultati e conclusioni

Attraverso il metodo impiegato è stato possibile identificare associazioni statisticamente significative tra la conta delle cellule somatiche e SNP presenti nel cromosoma 20 e nel cromosoma 22 (UCSC Genome Browser, 2002; Figura 2).

Figura 2: Manhattan plot rappresentante l'associazione tra SNP e con SCS nella razza Rendena



Due SNP nel cromosoma 22 erano situati nella regione del gene PRKAR2A, che traduce per la proteina “cAMP-dependent protein kinase type II-alpha regulatory subunit”. Questa proteina è coinvolta nella regolazione della trasduzione del segnale mediante adenosin-monofosfato ciclico (cAMP; Uniprot, 2019; Gene Ontology, 2021), specialmente nella regolazione dell'insulina (Kyoto Encyclopedia of Genes and

AZ. 7 – IOV 7.1: Produzione report con studio su GAS-MAS

Genomes, 2000). Per nostra conoscenza, questa è la prima indicazione di una connessione tra SCS e, indirettamente, infiammazione mammaria, con il gene PRKAR2A (Wathes, 2012). Approcci proteomici hanno infatti associato PRKAR2A a diverse cause di stress infiammatorio, come la malattia respiratoria dei vitelli (Neupane et al. 2018) o dovute all'alta temperatura (Eslamizad et al., 2020), ma mai alla resistenza alla mastite, valore che indirettamente può essere legato al contenuto di cellule somatiche nel latte.

Questo risultato, peraltro interessante, dovrà essere ovviamente confortato in futuro, da una più ampia serie di dati fenotipici e genotipici, per ora ancora in uno stadio non particolarmente avanzato, ma questi aspetti potranno essere ulteriormente chiariti anche per le altre razze partecipanti al progetto Dualbreeding.

Bibliografia:

- Aguilar, I., Legarra, A., Cardoso, F., Masuda, Y., Lourenco, D., & Misztal I. Frequentist p - values for large - scale - single step genome - wide association, with an application to birth weight in American Angus cattle. *Genetics Selection Evolution* 51 (2019).
- Christensen, O.F., Lund, M.S. Genomic prediction when some animals are not genotyped. *Genetics Selection Evolution* 42 (2010).
- Eslamizad, M., Dirk A., and Björn, K. The effect of chronic, mild heat stress on metabolic changes of nutrition and adaptations in rumen papillae of lactating dairy cows. *Journal of Dairy Science* 103 (2020): 8601-8614.
- Gene Ontology. <http://geneontology.org>. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Research* 49 (2021): 325-334.
- Kyoto Encyclopedia of Genes and Genomes. <https://www.genome.jp/kegg/>. Kanehisa, M. and Goto, S.; KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* 28 (2000): 27-30.
- Misztal, I., Lourenco, D., Legarra A. Current status if genomic evaluation. *Journal of Animal Science* 98 (2020): 1-14.
- Neupane, M., et al. Gene set enrichment analysis of SNP data in dairy and beef cattle with bovine respiratory disease. *Animal genetics* 49 (2018): 527-538.
- UCSC Genome Browser. <http://genome.ucsc.edu>. Kent, W.J., et al. The human genome browser at UCSC. *Genome Research* 12 (2002): 996-1006.
- Uniprot. <https://www.uniprot.org/>. The UniProt Consortium, UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research* 47 (2019): 506-515.
- Wathes, D. C. Mechanisms linking metabolic status and disease with reproductive outcome in the dairy cow. *Reproduction in domestic animals* 47 (2012): 304-312.