



**UNIVERSITÀ DEGLI STUDI DI MESSINA**  
**DIPARTIMENTO DI SCIENZE VETERINARIE**  
**DOTTORATO DI RICERCA IN SCIENZE VETERINARIE**

Coordinatore: Prof.ssa A. Ferlazzo

*Curriculum: Sanità Pubblica Veterinaria e Sicurezza Alimentare*

---

**Caratterizzazione genetica del *suino Nero Siciliano* mediante tecniche di Next  
Generation Sequencing: Whole Genome sequencing, SNPs discovery ed  
identificazione di polimorfismi associati allo spessore del lardo dorsale**

**Tesi di:**  
**Dott.ssa Irene Sapienza**

**Docente Tutor:**  
**Prof. V. Chiofalo**

# Indice

<b>I</b>	<b>Introduzione</b>	<b>4</b>
<b>I</b>	<b>Il suino Nero Siciliano</b>	<b>4</b>
<b>II</b>	<b>Il Genoma Suino</b>	<b>8</b>
<b>III</b>	<b>Dal sequenziamento a terminazione di catena al sequenziamento di prossima generazione</b>	<b>10</b>
<b>IV</b>	<b>Illumina vs Ion Torrent</b>	<b>12</b>
<b>V</b>	<b>Sequenziamento di Genomi: stato dell'arte</b>	<b>15</b>
<b>VI</b>	<b>Scopo della Tesi</b>	<b>17</b>
<b>1</b>	<b>Parte Prima</b>	<b>19</b>
<b>1.1</b>	<b>Whole Genome SNPs discovery nel suino Nero Siciliano</b>	<b>19</b>
<b>1.2</b>	<b>Abstract</b>	<b>19</b>
<b>1.3</b>	<b>Introduzione</b>	<b>20</b>
<b>1.4</b>	<b>Materiali e Metodi</b>	<b>22</b>
1.4.1	Estrazione del DNA, preparazione della library e sequenziamento	22
1.4.2	Quality-check, variant calling e SNP's effects	23
<b>1.5</b>	<b>Risultati</b>	<b>25</b>
1.5.1	Whole Genome Sequencing	25
1.5.2	Whole genome variant calling	25
1.5.3	Variant calling in 21 geni correlati alla fitness e SNP's effects	27
<b>1.6</b>	<b>Discussioni</b>	<b>30</b>
<b>1.7</b>	<b>Conclusioni</b>	<b>33</b>
<b>1.8</b>	<b>Disponibilità di dati di supporto</b>	<b>33</b>
<b>2</b>	<b>Parte Seconda</b>	<b>35</b>
<b>2.1</b>	<b>Analisi comparativa <i>in silico</i> di geni candidati per la qualità della carne in genomi di suini <i>Nero Siciliano, Duroc, Large White e Landrace</i></b>	<b>35</b>
<b>2.2</b>	<b>Abstract</b>	<b>36</b>
<b>2.3</b>	<b>Introduzione</b>	<b>37</b>
<b>2.4</b>	<b>Materiali e Metodi</b>	<b>38</b>
2.4.1	Analisi comparativa <i>in silico</i>	38
<b>2.5</b>	<b>Risultati</b>	<b>40</b>
<b>2.6</b>	<b>Discussioni</b>	<b>44</b>
<b>2.7</b>	<b>Conclusioni</b>	<b>48</b>
<b>3</b>	<b>Parte Terza</b>	<b>50</b>
<b>3.1</b>	<b>SNPs discovery in RRLs da pool di DNA di soggetti aventi fenotipi estremi e divergenti per il carattere BFT</b>	<b>50</b>
<b>3.2</b>	<b>Abstract</b>	<b>50</b>
<b>3.3</b>	<b>Introduzione</b>	<b>52</b>
<b>3.4</b>	<b>Materiali e Metodi</b>	<b>56</b>
3.4.1	Campionamenti e rilievi fenotipici	56
3.4.2	Estrazione del DNA, check di qualità, preparazione dei pool equimolari	57
3.4.3	Preparazione delle RRLs (Reduced Representation Libraries) e delle fragment libraries	58
3.4.4	Size selection, quantifica delle libraries, preparazione template e sequenziamento	58
3.4.5	Analisi dei dati di sequenza ed identificazione SNPs	60
<b>3.5</b>	<b>Risultati</b>	<b>62</b>

3.5.1	Sequenziamento ed identificazione SNPs da pool	62
3.5.2	Differenze tra le due RRLs costruite da DNA pool di soggetti aventi fenotipi estremi per il carattere BFT	63
3.5.3	Associazione tra marker identificati nelle due RRLs e BFT	64
<b>3.6</b>	<b>Discussioni</b>	<b>65</b>
<b>3.7</b>	<b>Conclusioni</b>	<b>68</b>
<b>4</b>	<b>Considerazioni generali e prospettive future</b>	<b>87</b>
<b>5</b>	<b>Appendice</b>	<b>89</b>
5.1	Materiale supplementare 5.1: Accession number degli Assembly presenti in NCBI per <i>Sus scrofa</i>	89
5.2	Materiale supplementare 5.2: Illumina® TruSeq® DNA PCR-Free Library Prep	94
5.3	Materiale supplementare 5.3: Mangime impiegato per l'accrescimento/ingrasso nelle aziende in cui sono stati effettuati i campionamenti	99
5.4	Materiale supplementare 5.4: Preparazione Ion Xpress™ Plus gDNA Fragment Library	100
5.5	Materiale supplementare 5.5: quantifica delle libraries: Agilent 2100 Bioanalyzer, Qubit 2.0 Fluorometer e qPCR	101
5.5.1	Agilent® 2100 Bioanalyzer®	101
5.5.2	Qubit® 2.0 Fluorometer	104
5.5.3	qPCR	105
5.6	Materiale supplementare 5.6: preparazione e arricchimento delle template-positive Ion PGM™ Hi-Q™ Ion Sphere™, sequenziamento su Ion PGM™	108
5.7	Materiale supplementare 5.7: valutazione dell'efficienza di templaggio con il kit Ion Sphere™ Quality Control e il Qubit® 2.0 Fluorometer	111
5.8	Materiale supplementare 5.8: Genotyping mediante sequenziamento Sanger degli SNPs localizzati in geni annotati, che mostravano differenze significative nelle due RRLs	113
<b>6</b>	<b>Allegati</b>	<b>116</b>
6.1	Whole genome SNPs discovery and analysis of genetic diversity in Nero Siciliano pig	116
6.2	In <i>silico</i> comparison of meat quality candidate genes among <i>Nero Siciliano</i> , <i>Duroc</i> , <i>Large White</i> and <i>Landrace</i> genomes	118
<b>7</b>	<b>Bibliografia</b>	<b>120</b>
<b>8</b>	<b>Sitografia</b>	<b>144</b>
<b>9</b>	<b>Ringraziamenti</b>	<b>145</b>

## I. Il suino Nero Siciliano

Il suino Nero Siciliano è un'antica razza autoctona siciliana tradizionalmente allevata, allo stato semi-brado ed in *en plein air*, prevalentemente nell'area dei Monti Nebrodi e delle Madonie. Il ricorso a tali sistemi di allevamento, armonizzandosi con le peculiarità del territorio, vede l'estrinsecazione del massimo potenziale genetico della razza da secoli evolutasi congiuntamente all'ambiente che le ha conferito peculiarità uniche e rispetto al quale gode di resistenza anche in condizioni avverse (Simonella, 2011). La razza rappresenta dunque una preziosa risorsa genetica in quanto ben adattata all'ambiente in cui vive.

Secondo alcuni Autori (Giuliani, 1940) la razza deriverebbe da un ceppo indigeno, secondo altri (Faelli, 1928) non sarebbe estraneo alla sua costituzione la razza Napoletana, mentre per altri ancora (Montanaro) discenderebbe dal tipo iberico del "Sanson". La sua presenza sull'Isola è documentata già nel periodo greco e cartaginese (VII-VI secolo a.C.).

Dal 2001 figura nel Registro Anagrafico dei tipi genetici autoctoni e la sua consistenza si attesta sui 10132 capi, di cui 80 verri, 1112 scrofe e 8940 allievi (ANAS, 2018). L'Associazione Nazionale Allevatori Suini (ANAS), Ente Morale riconosciuto con DPR 23 Dicembre 1962 n. 2031, si propone di promuovere ed attuare iniziative che possano contribuire al miglioramento, alla valorizzazione tecnico-economica ed alla promozione dell'allevamento del suino e dei prodotti da esso derivati. ANAS, sotto la supervisione del Ministero delle politiche agricole alimentari, forestali e del turismo (MIPAAFT), è responsabile del Libro Genealogico Italiano, strumento per il miglioramento genetico dei suini di razza Large White Italiana, Landrace Italiana, Duroc Italiana e Pietrain, e del Registro Anagrafico, strumento per la conservazione delle razze suine non sottoposte ad un piano nazionale di selezione.

Si riportano di seguito i caratteri tipici che determinano l'aderenza dei soggetti di suino Nero Siciliano allo Standard di Razza, stabilito dall'ANAS nel Quadro C del Registro dei Tipi Genetici Autoctoni (D.M. 6485 del 30 maggio 2007):

Caratteri morfologici: colore nero ardesia della cute su cui si impiantano robuste setole nere che sul tratto cervicale, garrese, dorso, lombi raggiungono la lunghezza di 10 cm circa, assumendo la forma di una criniera, tanto da richiamare l'aspetto del cinghiale.

Altezza al garrese cm 60-65. Testa allungata con profilo fronto-nasale tendenzialmente diritto, grugno appuntito e robusto; orecchie mediamente sviluppate dirette obliquamente in alto con punte portate orizzontalmente in avanti. È tollerata, anche se non tipica della

razza, la presenza di tettole. Collo di medio sviluppo, tronco non eccessivamente lungo, depresso ai lati, addome leggermente avallato, linea dorso-lombare convessa, groppa inclinata, arti lunghi robusti, unghie forti da gran camminatore.

Caratteri fisiologici: animale molto rustico e resistente alle malattie. La maturità sessuale è raggiunta, in entrambi i sessi, a circa 6-8 mesi. Possono aversi due parti l'anno con 7-8 suinetti per parto che, grazie al notevole istinto materno della scrofa, arrivano tutti allo svezzamento. Il numero delle mammelle varia da 10 a 12. Parti concentrati, maggiormente, a fine inverno e inizio primavera ([www.ANAS.it](http://www.ANAS.it)).

I suinetti si alimentano con latte materno fino alla fine della lattazione della scrofa, a circa 60 giorni. L'accrescimento è lento, stentato, e a un anno di vita raggiungono pesi di 50-60 kg. Le masse muscolari sono modeste, mentre si hanno notevoli accumuli di grasso nelle regioni del collo, garrese, dorso, lombi. Il ciclo di accrescimento-ingrasso inizia con animali appena svezzati (peso medio 15 kg) e si prolunga per 4-5 mesi per ottenere soggetti da destinare al consumo diretto quando hanno raggiunto peso medio 60-70 kg. Solo per i soggetti da destinare alla trasformazione l'allevamento si protrae per altri 4-5 mesi, per l'ottenimento di soggetti del peso medio di 110 kg.

Nel passato, l'allevamento era condotto quasi esclusivamente allo stato brado o semi-brado con l'utilizzo della vegetazione spontanea del sottobosco del querceto. Oggi, invece, vi sono aziende gestite in maniera più razionale, con sistema di allevamento prevalentemente *en plein air*. La possibilità di introdurre, dal punto di vista tecnico-gestionale, tale sistema di allevamento nella realtà del Parco dei Nebrodi è stata valutata attraverso un progetto P.O.P. 1994/1999 (Programma Operativo Plurifondo), dall'Istituto di Zootecnia – Facoltà di Medicina Veterinaria dell'Università degli Studi di Messina.

Gli animali vengono alimentati con orzo e favino, o con mangime commerciale schiacciato o pellettato, che rappresenta un'integrazione rispetto a quanto il suino Nero Siciliano reperisce dal pascolo grufolando (ghianda, tuberi, frutti di bosco, radici, melo selvatico, corbezzoli, ecc.). Il *pabulum* a disposizione è, dunque, ampio e variegato e spesso non sotto stretto controllo dell'allevatore. Viene sfruttata la capacità del suino Nero Siciliano di valorizzare appieno il potenziale alimentare dei boschi, delle loro radure e fasce perimetrali, interessando livelli differenti di vegetazione. Questo si traduce in un'ampia e razionale utilizzazione dei territori marginali ed in pratiche eco-compatibili, che assicurano il benessere degli animali.

È noto come i sistemi di allevamento estensivi, tipicamente adottati nell'areale mediterraneo, abbiano positive influenze sulla qualità delle produzioni di razze autoctone (Bonneau & Lebret, 2010). Questa qualità superiore, molto apprezzata dal consumatore, è dovuta a un'interazione positiva genotipo x ambiente (Pugliese & Sirtori, 2011) e, in effetti, risulta compromessa quando gli animali vengono allevati in sistemi confinati piuttosto che estensivi, come riportato da diversi Autori per il suino Iberico (Cava *et al.*, 2000a, Ventanas *et al.*, 2007). Diversi studi, circa gli effetti del sistema di allevamento sulla composizione e sulle caratteristiche organolettiche delle carni e dei prodotti derivati, sono stati condotti anche per il suino Nero Siciliano (Liotta *et al.*, 2002; Chiofalo V., 2002; Zumbo *et al.*, 2002; Pugliese *et al.*, 2004a,b ; Madonia *et al.*, 2007a,b; Di Rosa *et al.*, 2009; Di Rosa *et al.*, 2013). Da questi è emerso un effetto significativo del regime di alimentazione (ghiana vs orzo) sui livelli di composti volatili in lardo fresco e stagionato (Zumbo *et al.*, 2007a) e sul profilo acidico del grasso intramuscolare (Zumbo *et al.*, 2007b).

Nei sistemi in cui razze locali sono allevate in condizioni estensive, il punto chiave nella strategia di alimentazione è, infatti, la presenza di elevate quantità di alimenti ricchi di amido, come ghiande o castagne, solitamente procacciate direttamente dagli animali durante il pascolamento. Il miglioramento qualitativo delle carni derivate da questi animali, è essenzialmente riconducibile alla presenza di composti aromatici peculiari contenuti negli alimenti naturali, e dalla crescita compensatoria durante la fase di ingrasso (Lebret, 2008). Inoltre, nei suini alimentati con ghiande, sono stati rilevati maggiori livelli di MUFA (principalmente acido oleico) e bassi livelli di SFA (principalmente acido palmitico e stearico) sia nel grasso intramuscolare (Andrés *et al.*, 2001; Cava *et al.*, 1999, 2000b) che in quello dorsale (Diaz *et al.*, 1996), in suino Iberico, in suino Nero Siciliano (Chiofalo B. *et al.*, 2007) e Cinta Senese (Pugliese *et al.*, 2005).

Il sistema di allevamento semibrado o *en plein air*, tipico dell'areale nebroideo, presenta tra i suoi elementi caratterizzanti proprio la possibilità per gli animali di accedere al pascolo e procacciarsi il cibo autonomamente. Il pascolamento e l'uso delle risorse del territorio influiscono positivamente sulla composizione qualitativa delle carni, caratterizzate da una sottile infiltrazione di grasso intramuscolare, e sulle caratteristiche bromatologiche, rendendole idonee per la produzione del prosciutto crudo, perché tale fattore si traduce in una maggior capacità di ritenzione idrica e una minor perdita di liquidi durante la fase di stagionatura.

Questo connubio tra sistema di allevamento, alimentazione con essenze tipiche dell'areale nebroideo e tradizionale tecnica di stagionatura impiegata per la produzione del prosciutto crudo, ha fatto sì che si instaurasse un forte legame tra la zona geografica del Parco dei Nebrodi ed il prosciutto crudo da secoli prodotto in questo areale, tanto da determinare l'avvio dell'iter per il riconoscimento di una DOP. Il prosciutto crudo di suino Nero dei Nebrodi rappresenta una produzione d'eccellenza del nostro territorio e la Facoltà di Medicina Veterinaria dell'Università degli studi di Messina, ha condotto diversi studi al fine di caratterizzare il prodotto ed evidenziarne la sua tipicità identitaria, con la *Sezione di Zootecnica e Nutrizione Animale* del Dipartimento MO.BI.FI.P.A (Morfologia, Biochimica, Fisiologia, Produzioni Animali), prima, (Chiofalo V. *et al.*, 2005; Chiofalo B. *et al.*, 2005) e l'*Unità di Produzione Animale* del Dipartimento di Scienze Veterinarie in tempi più recenti (Di Rosa *et al.*, 2015).

Inoltre, le attività condotte dal Consorzio di Ricerca Filiera Carni nell'ambito della ricerca scientifica, del supporto tecnico alle aziende, incluso la definizione del Disciplinare di produzione per il riconoscimento del marchio DOP del prosciutto crudo, rappresentano un ulteriore contributo alla valorizzazione della razza e dunque allo sviluppo dell'economia locale.

Nel mese di agosto 2018, l'*Associazione per la tutela del Prosciutto Crudo ed il Salame di suino Nero dei Nebrodi* ha inoltrato istanza di riconoscimento della DOP per il Prosciutto crudo, ai sensi del Reg. (UE) n. 1151/2012 e D.M. 14 ottobre 2013, alla Regione Siciliana, UO A5.02 - Riconoscimento e tutela prodotti di qualità. Il 3 settembre 2018, i competenti Uffici regionali hanno formulato le prime osservazioni e richiesto delle integrazioni al dossier presentato dall'Associazione.

## II. Il Genoma Suino

Lo Swine Genome Sequencing Consortium (SGSC) è stato costituito nel settembre del 2003 da rappresentanti accademici, governativi ed industriali, allo scopo di coordinare una strategia per il sequenziamento completo del genoma suino. L'Assembly è stato completato nel 2012 (Groenen *et al.*, 2012) e migliorato, nell'annotazione ed assembly, da Warr e collaboratori (Warr *et al.*, 2015). È stato reso disponibile, nella versione *Sscrofa 11.1*, sul database Ensembl (Zerbino *et al.*, 2017) nel gennaio 2017 e corrisponde al GenBank Assembly ID GCA\_000003025.6 ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)).

Il genoma suino è stato sequenziato seguendo un approccio integrato che ha rappresentato un miglioramento rispetto la strategia annunciata preliminarmente (Schook *et al.*, 2005). La sequenza rappresenta principalmente il genoma di una singola scrofa di razza Duroc (Duroc 2-14), mentre alcune regioni (chr Y) derivano dal sequenziamento di librerie di Bacterial Artificial Chromosome (BAC) generate da altri soggetti. La sequenza completa del genoma è stata ottenuta integrando le informazioni di sequenza generate mediante Clone based sequencing e Whole genome shotgun (WGS) sequencing (Archibald *et al.*, 2010).

Dopo la costruzione di una mappa fisica ad alta risoluzione (Humphray *et al.*, 2007) è stato identificato il minimum tiling path per procedere al sub clonaggio e sequenziamento shotgun dei cloni BAC.

La mappa dei contig è stata costruita a partire dai cloni di 4 librerie BAC: CHORI-242 generata da una singola scrofa Duroc (Duroc 2-14); PigEBAC generata da un singolo maschio Large White x Meishan F1 (Anderson *et al.*, 2000); RPCI-44 (costruita da pool di DNA provenienti da 4 soggetti maschi, 1/4 Meishan, 3/8 Yorkshire, and 3/8 Landrace) (Fahrenkrug *et al.*, 2001) e INRA BAC library (costruita da un singolo maschio Large White) (Rogel-Gaillard *et al.*, 1999). Questa fase del progetto (hierarchical shotgun sequencing) è stata intrapresa dal Wellcome Trust Sanger Institute (WTSI), con il supporto del National Institute of Agrobiological Sciences, Japan. Allo scopo di incrementare il coverage ottenuto dalla strategia di sequenziamento clone-based e risolvere alcuni gap, il DNA dello stesso soggetto (Duroc 2-14) è stato sottoposto a sequenziamento WGS.

L'Assembly *Sscrofa11.1*, di dimensioni pari a 2.478.444.698 bp, consta di 20 cromosomi (1-18, X e Y) e 583 scaffold non posizionati. La metrica N50, che rappresenta la lunghezza tale che il 50% del genoma assemblato si trova in blocchi della dimensione N50 o più, è pari a 48.231.277. A seguito della sua pubblicazione, diversi progetti di risequenziamento



sono stati intrapresi (Li *et al.*, 2013; Li *et al.*, 2014; Bosse *et al.*, 2012, 2014, 2015; Fang *et al.*, 2012; Molnár *et al.*, 2014; Esteve-Codina *et al.*, 2011, 2013; Choi *et al.*, 2015; Frantz *et al.*, 2015; Jeong *et al.*, 2015; Wang *et al.*, 2015), ma pochi di questi hanno riguardato razze locali (Molnár *et al.*, 2014; Herrero-Medrano *et al.*, 2014; Esteve-Codina *et al.*, 2011). In appendice al presente lavoro di tesi è riportato l'elenco delle razze risequenziate e per le quali è disponibile l'Assembly del genoma, con i rispettivi accession number (materiale supplementare 5.1).

Il suino Nero Siciliano non è mai stato incluso, finora, in progetti di *whole genome resequencing*, è annoverato però tra le 20 razze locali che verranno caratterizzate nell'ambito di TREASURE (<https://treasure.kis.si/>).

TREASURE è un progetto europeo, finanziato nell'ambito del programma quadro di Ricerca e Innovazione Horizon 2020, che ha come obiettivo principale la caratterizzazione genetica di razze suine locali europee, ricorrendo a strumenti genetici e genomici.

### **III. Dal sequenziamento a terminazione di catena al sequenziamento di prossima generazione**

Descritto per la prima volta da Frederick Sanger nel 1977 (Sanger *et al.*, 1977a), il sequenziamento del DNA è stata una delle innovazioni più influenti nella ricerca biologica, per la quale è stato insignito del suo secondo premio Nobel per la chimica nel 1980. Avvalendosi della chimica a terminazione di catena, che utilizza deossinucleotidi e di-deossinucleotidi tri fosfati marcati con fluorofori, il genoma del batteriofago  $\Phi$ X174 è stato il primo ad essere sequenziato completamente nel 1977 (Sanger *et al.*, 1977b).

La modificazione chimica del DNA, seguita dal taglio in un sito specifico, pubblicata anch'essa nel 1977 (Maxam & W. Gilbert, 1977), ebbe meno successo a causa delle sue complessità tecniche, della necessità di utilizzare sostanze chimiche tossiche e della difficoltà intrinseca di scalabilità.

Il primo strumento di sequenziamento automatico (AB370), che adottava l'elettroforesi capillare, è stato introdotto da Applied Biosystems nel 1987. L'AB370 poteva rilevare 96 sequenze alla volta, 500 K basi al giorno e la lunghezza di lettura poteva raggiungere le 600 basi.

Nell'ultimo ventennio si è assistito ad una vera e propria rivoluzione tecnologica e, accanto a queste tecniche di "prima generazione", le tecniche di "prossima generazione" (NGS, Next Generation Sequencing) stanno mutando radicalmente la natura della sperimentazione genetica, con un enorme impatto su tutte le scienze "omiche".

Gli sforzi metagenomici stanno fornendo nuovi strumenti nel campo dell'ingegneria genetica, consentendo in tal modo di immaginare un futuro in cui organismi sintetici possano essere impiegati per migliorare la nostra qualità della vita e ridurre l'impatto ambientale, produrre sostituti della benzina o fornire soluzioni per ridurre l'inquinamento atmosferico (Venter *et al.*, 2004; Yooseph *et al.*, 2007; Rusch *et al.*, 2007). Lo studio del microbioma chiarirà in che modo cambiamenti delle comunità microbiche che popolano il tratto gastro-intestinale siano correlate alla risposta immunitaria, all'obesità e al cancro (Turnbaugh *et al.*, 2007; Ansorge *et al.*, 2009). La possibilità di sequenziare, a costi sempre più contenuti, interi genomi permetterà di indagare i meccanismi molecolari che sottendono le variazioni fenotipiche, aumentando la nostra comprensione sulla suscettibilità alle malattie e rendendo sempre più applicabile la medicina personalizzata (Pettersson *et al.*, 2009). E ancora, l'analisi dell'espressione genica mediante sequenziamento dell'intero trascrittoma (Mortazavi *et al.*, 2008; Cloonan *et al.*, 2008;

Sultan *et al.*, 2008), la classificazione dei pattern di metilazione del DNA, ChIP-Seq per il rilevamento di modificazioni di istoni e cromatina e per l'identificazione di siti di legame per fattori di trascrizione (Robertson *et al.*, 2007), permetteranno di fare enormi passi in avanti nel campo delle Scienze della Vita.

La continua evoluzione ed implementazione di algoritmi computazionali appropriati, aumenterà in maniera esponenziale la nostra comprensione dei meccanismi alla base di fenomeni complessi. L'analisi del dato e la sua comprensione biologica rimane infatti, ancora oggi, la principale criticità.

Le principali differenze tra le tecniche di prima e seconda generazione sono essenzialmente riconducibili alla velocità nell'ottenimento del dato (throughput massivo ed in parallelo), data la possibilità di elaborare milioni di sequenze in parallelo anziché 96 alla volta come avviene in elettroforesi capillare, e nella modalità di produzione del dato. Le *reads*, ovvero l'output di un esperimento NGS, sono prodotte da "librerie" di frammenti di DNA opportunamente processati mediante ligazione di specifici adattatori ad entrambe le estremità di ciascun frammento.

A parte queste caratteristiche generalmente condivise, le varie piattaforme disponibili in commercio differiscono in modo significativo sia nella tecnologia adottata (Sequencing by Synthesis, Sequencing by Ligation, Nanopores: Branton *et al.*, 2008; Fuller *et al.*, 2009; Metzker *et al.*, 2010) che nella chimica di sequenza e nella modalità di rilevamento del segnale. Di seguito verranno descritte brevemente le due piattaforme impiegate nel presente lavoro di tesi, per le altre si rimanda alla letteratura di riferimento (**Roche 454**: Margulies *et al.*, 2005; Dressman *et al.*, 2003; Berka *et al.*, 2005; Huse *et al.*, 2007; **AB SOLiD** (Sequencing by Oligo Ligation Detection): Mardis, 2007; Liu *et al.*, 2012a; **PacBio**: Rhoads & Au, 2015; Au *et al.*, 2012). Per comparazioni tra le piattaforme si veda Kumar *et al.*, 2012; Glenn, 2011; Mardis, 2007; Pettersson *et al.*, 2009; Ansorge, 2009; Quail *et al.*, 2012; Liu *et al.*, 2012a; Goodwin *et al.*, 2016.

#### IV. Illumina vs Ion Torrent

Negli ultimi anni l'industria del sequenziamento è stata dominata da Illumina. La tecnologia, basata su un approccio di sequencing by-synthesis (SBS), utilizza nucleotidi a terminazione reversibile marcati con fluorocromi (Bentley *et al.*, 2008), su template di DNA amplificati clonalmente ed immobilizzati ad un rivestimento di acrilammide sulla superficie di una *flowcell* di vetro. La rilevazione della singola base avviene mentre questa viene incorporata sul filamento nascente di DNA.

A partire da una miscela di frammenti di DNA *single strand* ligati ad adattatori complementari ad oligo immobilizzati covalentemente sulla *flowcell*, avviene la fase di generazione dei *cluster* ed il sequenziamento vero e proprio.

Ciascuna cella di flusso è suddivisa in otto corsie separate e l'ibridazione dei frammenti di DNA agli oligo in essa immobilizzati avviene mediante una fase attiva di riscaldamento e raffreddamento, seguita da una successiva incubazione con una polimerasi che amplifica i frammenti in un'area discreta sulla superficie della cella di flusso, in un processo noto come *bridge amplification* (Mardis *et al.*, 2007). Una volta generati i *cluster*, vengono forniti una polimerasi e i 4 nucleotidi marcati con 4 diversi fluorocromi. I nucleotidi hanno il 3'-OH chimicamente inattivato per garantire che solo una singola base sia incorporata per ciclo, mediante legame ad una molecola con ingombro sterico (terminatori reversibili). Poiché durante ogni ciclo di sequenziamento sono presenti tutti e quattro i dNTPs legati ai terminatori reversibili, la competizione naturale riduce al minimo le distorsioni dovute all'incorporazione.

Ogni ciclo di incorporazione di una base è seguito da una fase di imaging per identificare il nucleotide incorporato in ciascun *cluster*, e da uno *step* per la rimozione chimica del gruppo fluorescente e del blocco all'estremità 3' del nucleotide, per il successivo ciclo di incorporazione.

Una limitazione della tecnica potrebbe essere riconducibile al necessario requisito della tecnologia di imaging e di nucleotidi modificati e reagenti specifici.

Per superare questi limiti e “*democratizzare ulteriormente la pratica del sequenziamento*” (Rothberg *et al.*, 2011), è stato perseguito un cambio di paradigma basato sul sequenziamento non ottico su circuiti integrati di nuova concezione. Il concetto di base era di eseguire un sequenziamento by synthesis, con il rilevamento elettrochimico delle reazioni, ciascuna delle quali accoppiata al proprio sensore, a sua volta integrato in un

array di sensori su un chip CMOS (complementary metal-oxide semiconductor) (Arden *et al.*, 2010; Rothberg *et al.*, 2011; Merriman & Rothberg, 2012).

L'ideale sarebbe stato quello di rilevare la sintesi direttamente, tramite un sensore basato su transistor, senza l'uso di nucleotidi marcati. I sottoprodotti chimici fondamentali della sintesi a seguito dell'incorporazione da parte della polimerasi, sono rappresentati da pirofosfato scisso dal nucleotide incorporato, che come è noto è il sistema su cui si basa il rilevamento nel pirosequenziamento, e il rilascio di uno ione idrogeno dal sito di incorporazione al 3' sul filamento nascente. Pertanto, un sensore in grado di rilevare o pirofosfato o  $H^+$  sarebbe sufficiente come mezzo per rilevare direttamente l'incorporazione. Il rilevamento basato su transistor dell' $H^+$  è una tecnologia più consolidata, ed i dispositivi che ne fanno uso sono noti come pHFET (pH-sensitive field effect transistor). È su questo dispositivo che si basa la tecnologia Ion Torrent.

Il nucleo centrale dell'intero sistema di sequenziamento è infatti l'Ion Chip. Esso è costituito da uno strato superficiale, in cui sono ricavati dei pozzetti appositamente strutturati per poter accogliere le sfere (Ion sphere particles) ed i reagenti per il sequenziamento. Ciascun pozzetto ha diametro di pochi micron, tale che in esso possa trovare spazio una sola sfera. Questo primo strato poggia su un secondo costituito da semiconduttori, che permette la trasmissione dei segnali allo strato ancora sottostante. Quest'ultimo è strutturato come una piastra di sensori, uno corrispondente a ciascun pozzetto, che hanno la capacità di registrare le piccolissime variazioni di pH che avvengono all'interno del pozzetto durante il sequenziamento, trasformandole in differenze di potenziale, ovvero, in dati digitali. Il processo di sequenziamento è caratterizzato da un susseguirsi ciclico sulla superficie del chip di lavaggi e flussi (*flows*) di soluzioni, contenenti ciascuno dei quattro diversi nucleotidi. In questo modo, quando viene dispensato il nucleotide (secondo un algoritmo denominato Samba) complementare alla prima base libera sul frammento stampo, esso è incorporato alla sequenza nascente. Poiché la sfera, su cui è avvenuta l'amplificazione clonale per emPCR (PCR in emulsione), presenta sulla propria superficie frammenti tutti uguali, il legame della nuova base si verifica contemporaneamente su migliaia di frammenti, amplificandone il segnale. La formazione di ogni nuovo legame fosfodiesterico determina l'idrolisi del gruppo trifosfato del nucleotide entrante, con la liberazione netta di un protone. Il rilascio dei protoni causa l'abbassamento del pH della soluzione presente all'interno del pozzetto (0,02 unità di pH per incorporazione di ogni singola base); questa variazione viene trasmessa e registrata dal sensore sottostante e digitalizzata. Durante ciascun *flow*, la variazione di pH registrata è

direttamente proporzionale al numero delle basi incorporate nella sequenza nascente, e, nel momento in cui la sequenza stampo presenta una regione omopolimerica, l'intensità della variazione di potenziale registrata è tanto superiore quante più sono le basi identiche consecutive. A differenza delle tecnologie che sfruttano i segnali di fluorescenza, il sistema Ion Torrent non prevede le reazioni chimiche necessarie per la eliminazione dei gruppi fluorofori legati ai nucleotidi e ciò favorisce la rapidità del processo di sequenziamento che caratterizza questa tecnologia. Un aspetto vantaggioso offerto dalla tecnologia Ion Torrent è la disponibilità di chip con capacità scalari che consente di modulare il sistema in base alle necessità sperimentali ([www.thermofisher.com](http://www.thermofisher.com)).

Ci sono molte differenze sostanziali tra le piattaforme Ion Torrent ed Illumina, che si traducono in differenze significative nella elaborazione del dato e nel suo processamento. Ion Torrent produce solo reads single-end (SE) che variano in lunghezza, mentre Illumina può produrre reads SE e paired-end (PE) con lunghezze fisse. Pertanto, queste due piattaforme impiegano diversi algoritmi di rimozione dei duplicati (Zhu *et al.*, 2014). Un'altra differenza è che Ion Torrent è incline agli errori di sequenziamento nelle regioni omopolimeriche (Yeo *et al.*, 2012; Zeng *et al.*, 2013; Albers *et al.*, 2011).

Nel presente lavoro di tesi, per tenere conto di questi aspetti, differenti filtri di qualità, algoritmi di chiamata delle varianti e pipeline sono stati dunque adottati nell'analisi, spesso ricorrendo a script Bash custom.

## V. Sequenziamento di Genomi: stato dell'arte

Aumentare la nostra conoscenza sulla biodiversità terrestre ed usufruire responsabilmente delle sue risorse, sono tra le sfide scientifiche e sociali più importanti del nuovo millennio (Lewin *et al.*, 2018). Per far fronte a queste sfide è necessario comprendere l'organizzazione, l'evoluzione, e le interazioni tra milioni di organismi che popolano il pianeta. A questo scopo il “*The Earth BioGenome Project*” (EBP) si propone di perseguire l'ambizioso obiettivo di sequenziare, catalogare e caratterizzare i genomi di tutta la biodiversità eucariotica della Terra, attraverso uno sforzo internazionale coordinato, entro i prossimi 10 anni (Pennisi, 2017).

I risultati dell'EBP forniranno risposte ad importanti questioni che riguardano l'umanità, quali l'impatto dei cambiamenti climatici sulla biodiversità, la conservazione delle specie e degli ecosistemi a rischio di estinzione e la conservazione e il miglioramento dei servizi ecosistemici.

Sebbene 10-15 milioni di specie eucariotiche e forse trilioni di specie batteriche adornino l'Albero della Vita (Lewin *et al.*, 2018), solo ~2.3 milioni sono noti (Hinchliff *et al.*, 2015), di cui meno di 15.000, per lo più microrganismi, hanno un genoma completato o parzialmente sequenziato. Da questa piccola frazione del bioma conosciuto della Terra, è emersa una parte significativa delle moderne conoscenze in biologia e scienze della vita, che hanno reso possibili enormi progressi nei campi dell'agricoltura, medicina e tecnologia e migliorato gli approcci per la conservazione delle specie in via di estinzione.

Nonostante questi grandi progressi, la biodiversità globale è in gran parte non caratterizzata ed è minacciata dal cambiamento climatico, dalla distruzione degli habitat, dallo sfruttamento incontrollato delle specie e da altre attività legate all'uomo (Wilson, 1999; World Wildlife Fund, 2016; International Union for Conservation of Nature, 2017). Questo si traduce non solo in una minaccia per le specie selvatiche ma anche per la disponibilità di alimenti, biomateriali, farmaci e bioenergie per la nostra specie (Biodiversity International, 2017; Sharma & Sarkar, 2013). Il sequenziamento e l'annotazione di genomi fino ad ora non caratterizzati, permetterà di scoprire nuovi geni utili, proteine e vie metaboliche, fino ad ora rimaste inesplorate, che potrebbero essere impiegate per far fronte alle esigenze future della popolazione mondiale. La comprensione dei processi evolutivi, dell'adattamento, della speciazione e le ragioni delle estinzioni così come il ruolo e l'importanza delle specie nel funzionamento degli

ecosistemi, renderebbe concreta la possibilità di ripristinare ecosistemi fondamentali per la sopravvivenza umana.

Nell'Ottobre 2017 erano solo 2.534 le specie eucariotiche uniche con genomi sequenziati nel database NCBI (National Center for Biotechnology Information), che rappresenta meno dello 0,2% degli eucarioti noti. Inoltre, solo 25 di queste specie soddisfano lo standard per il contig N50, scaffold N50 e altri parametri proposti per i genomi di riferimento dal Genome 10K organization (G10K) (<https://genome10k.soe.ucsc.edu/>).

A tutti i livelli di qualità dell'assembly, solo 25 specie, in gran parte funghi, sono assemblate con lo *status* di "genoma completo" definito dall'NCBI. Anche la distribuzione filogenetica degli assembly di riferimento esistenti è molto distorta, con solo sette phyla eucarioti rappresentati (Lewin *et al.*, 2018).

Emerge dunque chiaramente la necessità di ottenere genomi ben annotati, che permetteranno di trovare risposte ad importanti quesiti scientifici e di fornire una solida base per future scoperte ed innovazioni biotecnologiche.



## VI. Scopo della Tesi

L'avanzamento tecnologico e le tecniche di sequenziamento di nuova generazione oggi disponibili offrono la possibilità di sequenziare interi genomi e generare una grande quantità di dati. Lo scopo del presente lavoro di tesi era quello di studiare, con approcci di Next Generation Sequencing, il genoma di suino Nero Siciliano al fine di fornire un contributo al processo di caratterizzazione della razza.

Nella **Parte Prima** presentiamo il sequenziamento dell'intero genoma di un verro di suino Nero Siciliano e la sua comparazione con l'ultima versione dell'Assembly (*Sscrofa 11.1*), rilasciato dallo Swine Genome Sequencing Consortium nel gennaio 2017, con GenBank Assembly ID GCA\_000003025.6 ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). Avvalendoci della piattaforma Illumina HiSeq X e della tecnologia PCR free, abbiamo prodotto un totale di 346,8 milioni di raw reads paired-end che, dopo filtering e trimming di qualità, sono state mappate sul genoma di riferimento di *S. scrofa*, con un coverage medio di 39,5 X. Nello stesso studio abbiamo inoltre selezionato un set di 21 geni, candidati funzionali o posizionali, correlati all'adattamento all'ambiente e ai tratti riproduttivi. L'obiettivo era quello di identificare eventuali regioni genomiche che potessero riflettere l'adattamento allo specifico ambiente di allevamento del suino Nero Siciliano, da impiegare per successivi studi di associazione. Il sequenziamento della library è stato effettuato presso la Macrogen (Korea).

Nella **Parte Seconda** del presente lavoro di tesi abbiamo focalizzato l'attenzione su un set di 48 geni coinvolti in processi metabolici correlati al metabolismo lipidico. La deposizione dei grassi assume, infatti, un ruolo strategico non solo per i noti effetti tecnologici in fase di stagionatura, in particolare per la produzione del prosciutto crudo, ma anche per gli importanti riflessi sulle caratteristiche qualitative del prodotto. L'analisi comparativa *in silico* ivi condotta ha avuto lo scopo di fornire una panoramica generale delle varianti presenti nei genomi delle razze che, seppure con peculiarità tipiche ed esclusive dei circuiti DOP, insieme, rappresentano il 90 % della produzione italiana (Large White, Landrace, Duroc). Si è trattato dunque di uno screening preliminare che, non avendo la pretesa di essere esaustivo, ha rappresentato solo un primo *step* nel processo di caratterizzazione della razza.

Nella **Parte Terza** del presente lavoro di tesi è stato studiato il carattere *spessore del grasso dorsale*, con lo scopo di identificare SNPs che potessero essere associati a questo fenotipo nel suino Nero Siciliano. Come primo passo per valutare le putative associazioni tra polimorfismi distribuiti nell'intero genoma e lo spessore del grasso dorsale, è stato applicato un approccio di genotipizzazione selettiva, da pool di DNA, che studiava se le frequenze alleliche differivano tra gruppi divergenti di soggetti selezionati in base alle misurazioni fenotipiche effettuate per questo carattere. Solo gli SNPs che mostravano differenze significative tra i due gruppi, e che erano mappati su geni annotati, sono stati poi impiegati per la genotipizzazione individuale nell'intero campionamento originario, comprendente i soggetti che erano stati inclusi nella costruzione dei pool. Questo ha rappresentato il migliore compromesso tra il risparmio sui costi del pooling e la potenza statistica e la completezza di informazioni fornite dalla genotipizzazione individuale.

Nella Terza Parte del lavoro di tesi ci siamo avvalsi della piattaforma Ion PGM™ System (Thermo Fisher Scientific), in dotazione presso i laboratori PanLab siti presso il Dipartimento di Scienze Veterinarie dell'Università degli Studi di Messina.

# 1. Parte Prima

## 1.1. Whole Genome SNPs discovery nel suino Nero Siciliano

### 1.2. Abstract

#### Background

Local pig breeds represent an important genetic reserve to utilize mainly for the obtaining of typical products and for recover the primitive germplasm and biodiversity manteniment. Nero Siciliano pig is an autochthonous breed of Sicily (Italy). In contrast to modern livestock production systems, this breed is reared under semi-extensive conditions that ensure animal's welfare and sustainability of its productions. In order to explore its genetic variability, here we present for the first time the whole genome sequencing and SNPs discovery of a male domestic Nero Siciliano pig versus the last pig reference genome *Sus scrofa 11.1*. Additionally, we further evaluate the genetic variation occurring in 21 selected fitness-related genes.

#### Results

A total of 346, 8 million paired reads were generated by sequencing the whole-genome of a single, male, Nero Siciliano pig. After quality control, 99.03% of the reads were mapped to the existing *Sus scrofa* reference genome, and over 11 million variants were detected. Genome-wide variation analysis revealed that Nero Siciliano pig contains, on average, 1 variant every 222 bases resulting in ~36% missense, ~0.4% nonsense and over 63% silent mutations. Structural variant analysis showed 1,261,412 indels, 196,971 replacements, 2,383 tandem duplications and 1,029 inversions. In a subset of genes involved in biological processes related to environmental adaptation and reproductive traits 6,747 variants were identified, resulting in a rate of 1 variant every ~276 bases. Among these variants, 1,132 were novel to the dbSNP151 database.

#### Conclusions

Explore the genetic variability of this breed can help to discover its genetic uniqueness (genes pool and alleles combination) that can explain genetic bases of its behavioural, physiological or morphological traits.

**Keywords:** Nero Siciliano pig, Whole Genome Sequencing, Illumina sequencing, Variant calling, SNPs discovery, Fitness genes.

### 1.3. Introduzione

Negli ultimi decenni molte razze e popolazioni locali sono state soggette a erosione genetica e severa perdita di biodiversità, con il risultato dell'impoverimento di un prezioso pool genico, che ha interessato soprattutto le aree marginali e i sistemi di allevamento a ridotto input.

La FAO (Food and Agriculture Organization) ha espresso preoccupazione per la progressiva sostituzione di razze locali con razze cosmopolite migliorate, considerato che quest'ultime non possono competere con quelle autoctone in termini di rusticità, resistenza alle malattie e adattamento alle ridotte disponibilità alimentari (FAO; Muir *et al.*, 2008).

Molte razze autoctone presentano, infatti, caratteristiche uniche che possono contribuire ad affrontare le sfide legate al cambiamento climatico, la crescente richiesta di cibo connessa all'aumento della popolazione mondiale e alla sicurezza alimentare (DAD-IS FAO, <http://dad.fao.org/>).

La FAO stima che le razze animali a rischio di estinzione siano circa il 22% (FAO, 2012), mentre di molte specie non si conosce lo stato di rischio a causa di dati inadeguati circa l'effettiva dimensione delle loro popolazioni e la struttura genetica (Secondo Rapporto sullo Stato delle Risorse Genetiche Animali del mondo per l'Alimentazione e l'Agricoltura, 2015).

Per invertire questo trend negativo, diverse azioni sono state intraprese nell'ultimo decennio (Global Plan of Action for Animal Genetic Resources, 2007; Convention on Biological Diversity's Strategic Plan for Biodiversity 2011–2020), e la salvaguardia delle risorse genetiche animali allevate e addomesticate rappresenta uno degli obiettivi per lo sviluppo sostenibile previsti dall'ambiziosa Agenda Universale 2030 (Sustainable Development Goals, 2015), siglata da 193 Paesi membri dell'ONU nel Settembre 2015 (<http://www.un.org>). Tale obiettivo si inserisce in quello più ampio di porre fine alla fame, garantire la sicurezza alimentare e promuovere un modello di agricoltura sostenibile. In questo scenario le razze locali, rustiche e resilienti, potrebbero garantire la capacità di adattarsi ai mutevoli ambienti e di fornire prodotti e servizi per le comunità locali, in risposta al crescente interesse da parte dei consumatori nei riguardi di tematiche inerenti il benessere animale, la sostenibilità delle produzioni e la tipicità delle stesse.

Le razze locali sono ben adattate alle condizioni estensive o semi-estensive di allevamento e dunque capaci di valorizzare sottoprodotti o scarti delle lavorazioni, in grado di procacciarsi il cibo grazie anche, nel caso del suino Nero Siciliano, alla forte propensione

al pascolamento favorita dalla robustezza degli arti, la conformazione del grugno, la sua frugalità e il temperamento dinamico (Chiofalo L., 2007; Maiorano, 2009). Tali razze richiederebbero una dettagliata caratterizzazione morfologica e molecolare che tenga conto anche delle influenze dell'ambiente in cui si sono originate ed adattate, che ha potuto contribuire nel tempo a modellarne il genoma. Inoltre, la possibilità di identificare una connessione tra il sistema di allevamento tradizionale, la disponibilità di risorse pabulari naturali e le caratteristiche sensoriali e/o nutrizionali peculiari dei prodotti da esse derivate, permetterebbe il riconoscimento di un plus valore degli stessi, anche in termini di prezzi di mercato, come accade già per diverse produzioni tipiche (Kušec *et al.*, 2015; García-González *et al.*, 2013; Pugliese & Sirtori 2011).

La salvaguardia della biodiversità assume un ruolo prioritario non solo per la sua valenza etica, ma anche per far fronte ad eventuali esigenze future dal momento che la selezione ed il miglioramento genetico delle linee commerciali hanno determinato una severa riduzione della variabilità genetica. Diversi Autori (Muir *et al.*, 2008; Tapio *et al.*, 2006; Ollivier, 2009) hanno sottolineato l'importanza delle razze locali e popolazioni ancestrali, in quanto detentrici di combinazioni alleliche che sono andate perdute a seguito dei severi programmi selettivi che sono stati attuati nelle razze cosmopolite, e che potrebbero rappresentare germoplasma potenzialmente utile per introgressioni future e per l'identificazione di *signatures of selection* (Wilkinson *et al.*, 2013).

È noto come i genomi si evolvano in risposta alla selezione (naturale e uomo-mediata) e, a seguito della pubblicazione del genoma di riferimento di *Sus scrofa* (Groenen *et al.*, 2012), diversi progetti di risequenziamento sono stati intrapresi (Li *et al.*, 2013; Li *et al.*, 2014; Bosse *et al.*, 2012, 2014, 2015; Fang *et al.*, 2012; Molnár *et al.*, 2014; Esteve-Codina *et al.*, 2011, 2013; Choi *et al.*, 2015; Frantz *et al.*, 2015; Jeong *et al.*, 2015; Wang *et al.*, 2015), ma pochi di questi hanno riguardato razze locali (Molnár *et al.*, 2014; Herrero-Medrano *et al.*, 2014; Esteve-Codina *et al.*, 2011). Il suino Nero Siciliano non è mai stato incluso in progetti di *whole genome resequencing*.

Il suino Nero Siciliano è un'antica razza autoctona siciliana tradizionalmente allevata in sistemi estensivi e semi-estensivi sul territorio dell'Isola, anche se oggi si registrano prevalentemente allevamenti *en plein air*. In virtù del peculiare sistema di allevamento e delle particolari condizioni pedoclimatiche ed orografiche che caratterizzano l'areale dei Nebrodi, dove il suino Nero è principalmente allevato, questa razza potrebbe conservare ancora il pool genico originario. In realtà il meticciamiento incontrollato, avvenuto in passato per migliorare alcune caratteristiche produttive ed andare incontro alle esigenze di

mercato (incremento tagli magri), ha rischiato di "inquinare" il suo patrimonio genetico, anche a causa di inadeguati piani di salvaguardia e valorizzazione delle sue produzioni (Guastella *et al.*, 2010).

Diversi studi inerenti la caratterizzazione molecolare (Russo *et al.*, 2004; Davoli *et al.*, 2007; Guastella *et al.*, 2010; Fernández *et al.*, 2017), lo studio della struttura genetica (Chessa *et al.*, 2013) e l'analisi dei geni del colore del mantello (*MC1R* e *KIT*) per valutarne l'utilità ai fini della tracciabilità di razza (D'Alessandro *et al.*, 2007a,b; Fontanesi *et al.*, 2010a,e; Fontanesi *et al.*, 2011a; Fontanesi *et al.*, 2012e), ne hanno indagato la variabilità genetica.

L'avanzamento tecnologico e le tecniche di Next Generation Sequencing (NGS) oggi disponibili offrono la possibilità di sequenziare interi genomi e generare una grande quantità di dati. Nella Prima Parte di questo lavoro di tesi presentiamo, per la prima volta, il sequenziamento e l'analisi del genoma di un verro di suino Nero Siciliano e la sua comparazione con l'ultima versione del genoma suino di riferimento (*Sscrofa 11.1*).

Abbiamo inoltre selezionato un set di 21 geni (Tabella 1) che sono stati correlati, in studi precedenti, all'adattamento all'ambiente ed a tratti riproduttivi (geni candidati funzionali o posizionali). È noto come i geni coinvolti nella percezione sensoriale, immunità e difesa da ospite mostrino rapida evoluzione. Quest'ultima potrebbe probabilmente riflettere l'adattamento ai diversi agenti patogeni e avversità nei diversi ambienti (Groenen *et al.*, 2012; Groenen *et al.*, 2016).

L'obiettivo era quello di identificare eventuali regioni genomiche che potessero riflettere l'adattamento allo specifico ambiente di allevamento, da impiegare per successivi studi di associazione.

## **1.4. Materiali e Metodi**

### **1.4.1. Estrazione del DNA, preparazione della library e sequenziamento**

Un campione di sangue è stato prelevato da un verro Nero Siciliano (ANAS ID: 163347) e collezionato in K<sub>3</sub>EDTA (4°C) fino all'arrivo in laboratorio. La frazione leucocitaria è stata isolata per l'estrazione del DNA genomico totale (gDNA) utilizzando il Wizard® Genomic DNA Purification Kit (Promega Corporation, Italia), attenendosi al protocollo fornito dalla ditta produttrice.

La quantifica del DNA è stata effettuata in fluorimetria mediante Qubit 2.0 Fluorometer con il kit di dosaggio Qubit dsDNA HS (Thermo Fisher, Italia). La qualità del DNA è stata

valutata per via spettrofotometrica mediante Nanophotometer® P-330 (Implen GmbH) ed ispezione visiva dopo elettroforesi su gel di agarosio, per valutarne l'integrità (agarosio 1% in tampone TAE 1X).

La preparazione della library è stata eseguita con Illumina® TruSeq® DNA PCR-Free Library Prep kits (dimensione dell'inserito 350 bp), utilizzando 1 µg di gDNA, attenendosi al protocollo fornito dalla ditta produttrice (materiale supplementare 5.2). Il sequenziamento paired-end è stato eseguito sulla piattaforma HiSeqX (Illumina®), presso la Macrogen (Korea).

#### **1.4.2. Quality-check, variant calling e SNP's effects**

Le raw reads sono state ispezionate usando il programma FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) e processate con Trimmomatic v. 0.36 (Bolger *et al.*, 2014) al fine di rimuovere gli adattatori e le sequenze di bassa qualità (Phred score  $\leq 30$ ). Le rimanenti reads di buona qualità sono state mappate sul genoma di riferimento di *Sus scrofa* (versione 11.1; GenBank: GCA\_000003025.6) con BWA (versione 0.7.12-r1039) (Li & Durbin, 2009) e la qualità del mapping è stata valutata con Qualimap2 (Okonechnikov *et al.*, 2016).

Le analisi relative ai polimorfismi a singolo nucleotide (SNP), le brevi inserzioni e delezioni (INDEL) e le varianti strutturali (SV) sono state condotte con SUPERW (Sanseverino *et al.*, 2015) e PINDEL (Ye *et al.*, 2009).

Dopo rimozione di tutte le varianti che mostravano lo stesso genotipo del riferimento, le rimanenti sono state ulteriormente filtrate utilizzando i seguenti parametri: QUAL (punteggio di qualità della variante chiamata su scala Phred)  $\geq 30$ , DP (numero di basi di alta qualità per variante chiamata)  $\geq 10$ , AD (Allele Depth)  $\geq 10$ . I *quality scores*, in scala Phred, indicano la probabilità di una chiamata non corretta per ciascuna base, su scala logaritmica (Phred score=30 assume errore di tipo 1  $p = 0,001$ ; accuratezza della chiamata 99,9%). I putativi effetti funzionali degli SNPs identificati sono stati poi predetti *in silico* con il software SnpEff v4\_3m\_core (Cingolani *et al.*, 2012a).

Un set di 21 geni correlati all'adattamento all'ambiente ed alle caratteristiche vitali e riproduttive (Tabella 1) è stato analizzato ricorrendo a samtools (Li *et al.*, 2009a) e bcftools (Li, 2011).

Chr	Gene symbol	Gene function or putative association	Chromosome (GenBank Sequence)	Start	End
1	ESR1	total newborn, newborn alive	CM000812.5	14217032	14604906
1	VPS13A	maintenance of thermostatic status, blood circulation	CM000812.5	230069339	230331343
1	NR6A1	body size	CM000812.5	265320597	265570941
2	FSHB	total newborn, newborn alive	CM000813.5	30395769	30399282
3	EIF2AK3	gene overlaps with QTLs for osteochondrosis score and feet and leg conformation	CM000814.5	57423894	57506247
3	AZGP1	adaptation to environment	CM000814.5	7867521	7874857
4	PLAG1	body size	CM000815.5	75646585	75696718
6	IL12RB2	immune related gene	CM000817.5	145210251	145292399
6	FUT1	resistance to disease	CM000817.5	54077431	54080475
8	GNRHR	ovulation rate	CM000819.5	65470206	65488900
8	LCORL	body size	CM000819.5	12806878	12969370
9	AHR	litter size	CM000820.5	86511866	86555950
12	PPP1R1B	candidate genes affecting behaviour	CM000823.5	22681244	22690978
13	STAB1	immune related gene, defence against bacterial infection	CM000824.5	34630448	34659371
13	GPR149	potential effect on fertility, prolificacy	CM000824.5	94356371	94419917
13	CLDN1	potential effect on fertility	CM000824.5	127714857	127730628
14	RBP4	total newborn, newborn alive	CM000825.5	105037360	105044552
14	JMJD1C	potential effect on fertility	CM000825.5	66640845	66966911
15	DCAF17	maintenance of thermostatic status, hair growth	CM000826.5	77564629	77603913
16	PRLR	involved in several reproductive traits, including litter size	CM000827.5	20637568	20655881
18	TAS2R40	adaptation to specific dietary repertoires and environment	CM000829.5	7024418	7027197

**Tab.1- Geni correlati alla fitness indagati in questo studio**

Successivamente le mutazioni identificate nel cluster dei 21 geni, annotate da snpEff e classificate come aventi effetti ad alto impatto, sono state visivamente ispezionate con MEGA7 (Kumar *et al.*, 2016) utilizzando le CDS ed i relativi trascritti scaricati da GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>), al fine di valutarne il putativo effetto funzionale sulle sequenze proteiche.

Le varianti chiamate da SUPERW e PINDEL sono state confrontate con bedtools intersect (Quinlan *et al.*, 2010) ed i duplicati rimossi dall'output di PINDEL per evitare la ridondanza dei dati. Per identificare SNPs mai descritti e depositati prima in database, snpSift (Cingolani *et al.*, 2012b) è stato utilizzato contro dbSNP151 ([ftp.ncbi.nih.gov/snp/organisms/pig\\_9823/](ftp.ncbi.nih.gov/snp/organisms/pig_9823/)) e tutti gli SNPs *novel* risultanti sono stati esaminati e confermati manualmente.



## 1.5. Risultati

### 1.5.1. Whole Genome Sequencing

In questo studio, un totale di 346,8 milioni di raw reads paired-end sono state prodotte dal sequenziamento con l'Illumina HiSeq X. Dopo filtering e trimming di qualità, ~ 344,3 milioni (99,29%) di reads di alta qualità sono state mappate sul genoma di riferimento di *Sus scrofa*, con un coverage medio di 39,5 X. Le metriche relative al trimming e alla qualità del mapping sono elencate in Tabella 2.

<b>Trimmomatic</b>	<b>number of reads</b>	<b>%</b>
Input Read Pairs	346,832,803	
Both Surviving	344,372,450	99.29
Forward only surviving	2,179,929	0.63
Reverse only surviving	21,491	0.01
Discarded	258,933	0.07
<b>QUALIMAP2</b>		<b>%</b>
Reference size (bp)	2,501,895,775	
Number of reads	693,466,979	
Mapped reads	686,713,733	99.03%
Supplementary alignments	4,722,079	0.68%
Unmapped reads	6,753,246	0.97%
Mapped paired reads	686,713,733	99.03%
Mapped reads, first in pair	344,533,898	49.68%
Mapped reads, second in pair	342,179,835	49.34%
Mapped reads, both in pair	683,465,613	98.56%
Mapped reads, singletons	3,248,120	0.47%
Overlapping read pairs	23,189,648	6.69%
GC Percentage		42.52%
Coverage mean	39,5935	
Mean Mapping Quality	38,23	

Tab. 2 – Metriche relative al sequenziamento e mapping

### 1.5.2. Whole genome variant calling

Nell'intero genoma, un totale di 11253945 varianti genetiche sono state rilevate da SUPERW. Di queste, ~ 82% erano SNPs mentre ~ 12% e ~ 5% erano rispettivamente brevi inserzioni e delezioni. Del totale degli SNPs rilevati (~ 9000000) oltre il 58% erano in eterozigosi, mentre il rimanente 42% era in omozigosi alternativa rispetto la sequenza genomica di riferimento. La frequenza complessiva osservata era di 1 variante ogni 222 basi con una frequenza di mutazione di 1 SNP ogni 269 bp.

L'analisi condotta con SnpEff ha mostrato che la maggior parte delle varianti identificate erano localizzate in regioni non codificanti del genoma, come gli introni e le regioni intergeniche. La tabella 3 mostra nel dettaglio l'annotazione delle varianti sulla base della

loro localizzazione genomica, la tabella 4 la loro classificazione sulla base dell'impatto sulla funzionalità della proteina.

Migliaia di SNPs sono stati identificati anche all'interno delle regioni trascritte dei geni. Tali varianti, nelle regioni 5' e 3' UTR, *upstream* e *downstream*, potrebbero influenzare la trascrizione o la traduzione genica, ma le effettive conseguenze di tali mutazioni devono essere confermate caso per caso. In questo studio "*upstream*" è definito come 5 kilobasi (kb) a monte del sito di inizio della trascrizione mentre "*downstream*" è definito come 5 kb a valle del sito di aggiunta di poliA.

Circa il 36% delle mutazioni osservate erano missenso, lo 0,4% non senso (stop-gain SNP) e il 63,6% erano silenti (sinonime), risultando in un rapporto missenso/silente e Ts/Tv (transizione/transversione) di 0,5617 e 2,3956 rispettivamente.

Tra le varianti strutturali identificate da PINDEL, sono state osservate un totale di 808486 inserzioni, 452926 delezioni, 196971 sostituzioni, 2383 duplicazioni in tandem e 1029 inversioni. Di queste, 586686 erano in eterozigosi mentre 875109 erano in omozigosi alternativa.

Type	Count	Percent
GENE	997	0.002%
SPLICE_SITE_REGION	32992	0.053%
TRANSCRIPT	25043269	39.999%
INTERGENIC	5240828	8.371%
UPSTREAM	3012181	4.811%
UTR_5_PRIME	114247	0.182%
EXON	403120	0.644%
SPLICE_SITE_DONOR	771	0.001%
INTRON	25379343	40.536%
SPLICE_SITE_ACCEPTOR	888	0.001%
UTR_3_PRIME	292332	0.467%
DOWNSTREAM	3088483	4.933%

Tab. 3 – Annotazione delle varianti identificate nell'intero genoma. Molte varianti presentano annotazioni multiple

Type	Count	Percent
HIGH	9880	0.016%
LOW	169578	0.271%
MODERATE	72595	0.116%
MODIFIER	62357398	99.597%

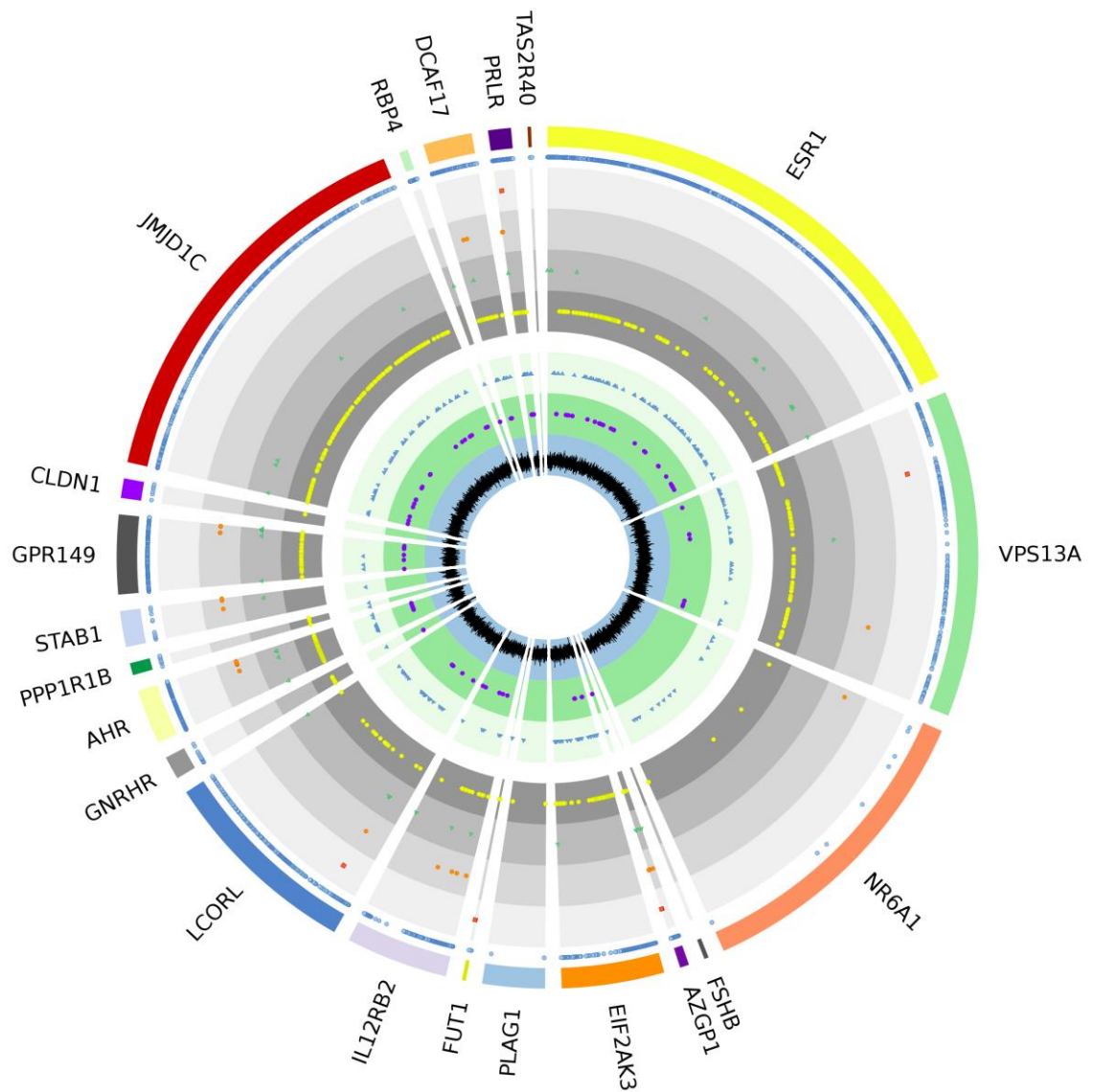
Tab. 4- Numero di effetti classificati per impatto sulla funzionalità proteica nell'intero genoma. Molte varianti presentano annotazioni multiple

### 1.5.3. Variant calling in 21 geni correlati alla fitness e SNP's effects

Nel pannello di geni correlati alla fitness selezionato in questo studio, sono stati identificati un totale di 6747 SNPs e short INDEL, di cui 7 classificati come aventi alto impatto sulla funzione proteica, 35 con moderato impatto, 54 come a basso impatto e 6651 come modificatrici (Tabella 5), risultanti in un tasso di mutazione di 1 ogni ~ 276 bp. Tra le varianti totali identificate 1132 erano novel, 476 in eterozigosi e 656 in omozigosi alternativa. SnpEff ha mostrato 88 mutazioni missenso e 206 mutazioni silenti, che si traduce in un rapporto missenso/silente di 0,4272 mentre il rapporto Ts/Tv è risultato 2,3302. Le 7 mutazioni ad alto impatto, tutte in omozigosi alternativa, hanno interessato 5 dei 21 geni esaminati: *VPS13A* (Vacuolar protein sorting 13 homolog A); *AZGP1* (Alpha-2-glycoprotein 1, zinc-binding); *LCORL* (Ligand-dependent nuclear receptor corepressor-like protein), *FUT1* (Fucosyltransferases 1); *PRLR* (Prolactin Receptor). Tra le varianti identificate 1 era uno SNP e 6 erano inserzioni nucleotidiche. 4 inserzioni erano mutazioni *gain of function* e ripristinavano il frame di lettura nei geni *VPS13A*, *AZGP1*, *FUT1* e *PRLR*, come evidenziato dall'analisi comparativa con il genoma di riferimento ed i rispettivi trascritti. Le rimanenti 2 inserzioni causavano un *premature stop codon* e un *lack stop codon* nei geni *AZGP1* e *LCORL* rispettivamente, mentre l'unico SNP rilevato era una mutazione missenso (ACG→GCG; Thr<sup>103</sup>→Ala<sup>103</sup>) che interessava il gene *FUT1*. 5 delle 7 mutazioni ad alto impatto non erano mai state riportate in dbSNP. Tra le varianti strutturali identificate da PINDEL, 101 erano sostituzioni (RPL), 132 inserzioni e 112 delezioni. Di queste, 203 erano in eterozigosi mentre 142 erano in omozigosi alternativa. La Figura 1 mostra la distribuzione di tutte le mutazioni identificate nel cluster dei 21 geni indagati ed il relativo coverage.

Gene symbol	Length	Variants (SNPs and short INDEL) classified by impact				Total	%Variants/length	Structural variants
		High	Low	Moderate	Modifier			
ESR1	387875	0	14	0	2891	2905	0,749	102
VPS13A	262005	1	1	1	370	373	0,142	36
NR6A1	250345	0	0	1	9	10	0,004	13
FSHB	3514	0	0	0	0	0	0,000	0
EIF2AK3	82354	0	1	0	410	411	0,499	24
AZGP1	7337	2	4	7	66	79	1,077	2
PLAG1	50134	0	0	0	3	3	0,006	1
IL12RB2	82149	0	4	4	318	326	0,397	11
FUT1	3045	2	0	0	3	5	0,164	1
GNRHR	18695	0	0	0	25	25	0,134	3
LCORL	162493	1	4	1	422	428	0,263	30
AHR	44085	0	8	9	494	511	1,159	20
PPP1R1B	9735	0	0	0	13	13	0,134	2
STAB1	28924	0	1	3	9	13	0,045	1
GPR149	63547	0	9	4	370	383	0,603	10
CLDN1	15772	0	0	0	8	8	0,051	0
RBP4	7193	0	1	0	56	57	0,792	2
JMJD1C	326067	0	5	0	762	767	0,235	62
DCAF17	39285	0	1	2	178	181	0,461	10
PRLR	18314	1	1	3	240	245	1,338	15
TAS2R40	2780	0	0	0	4	4	0,144	0
<b>TOTAL</b>	<b>1865648</b>	<b>7</b>	<b>54</b>	<b>35</b>	<b>6651</b>	<b>6747</b>	<b>0,362</b>	<b>345</b>

**Tab. 5- Varianti identificate nei 21 geni correlati alla Fitness**



**Fig. 1- Varianti rilevate in 21 geni correlati alla fitness.** Dall'esterno verso l'interno, gli anelli mostrano: il totale degli SNP e short INDEL (cerchi blu), mutazioni ad alto (quadrato rosso), moderato (cerchi arancioni), basso impatto sulla funzione proteica (triangoli verdi), nuovi SNPs e short INDEL (cerchi gialli), SV INDEL (triangoli blu), SV RPL (cerchi viola), coverage (istogrammi neri).

## 1.6. Discussioni

L'analisi di genome-wide SNPs discovery ha mostrato che meno dello 0,5% dell'intero genoma del suino Nero Siciliano sequenziato differisce dal genoma di riferimento *Sscrofa 11.1*. Come atteso, la maggior parte delle varianti genetiche erano localizzate in regioni non codificanti del genoma e circa il 58% delle mutazioni rilevate erano in forma eterozigote, mostrando un alto livello di eterozigosi rispetto altre razze locali (Molnár *et al.*, 2014).

Anche il numero delle varianti identificate (11253945), è risultato piuttosto elevato se comparato a quanto precedentemente riportato per altre razze locali (Molnár *et al.*, 2014), ed in particolare circa il doppio di quanto rilevato in Cinta Senese, Calabrese e Casertana in studi precedenti (Herrero-Medrano *et al.*, 2014).

Il rapporto Ts / Tv è risultato simile ad altri genomi suini (Fahrenkrug *et al.*, 2002; Kerstens *et al.*, 2009; Molnár *et al.*, 2014; Kang *et al.*, 2015; Bovo *et al.*, 2015) e a quanto riportato per il genoma umano (Wang *et al.*, 1998; DePristo *et al.*, 2011) e asinino (Bertolini *et al.*, 2015). È risultato inoltre paragonabile ai rapporti ottenuti in diversi altri studi di NGS, prodotti con la nostra stessa tecnologia di sequenziamento (Illumina), nei quali la frequenza delle transizioni era circa il doppio di quella delle transversioni (Bertolini *et al.*, 2015; Liu *et al.*, 2012b; Molnár *et al.*, 2014). Il tasso di mutazione degli SNPs osservato è risultato invece leggermente superiore rispetto quanto riportato in letteratura (Jungerius *et al.*, 2005).

L'analisi di SNPs discovery nei 21 geni correlati alla fitness ha mostrato un tasso di mutazione coerente rispetto ai dati dell'intero genoma. Nel presente lavoro di Tesi, abbiamo focalizzato l'attenzione sulle mutazioni ad alto impatto che potrebbero influenzare il prodotto genico.

Il gene *VPS13A* svolge un ruolo nel mantenimento delle condizioni termostatiche durante lo stress termico ed è coinvolto nella circolazione sanguigna (Groenen *et al.*, 2016). In questo studio è stata identificata l'inserzione di una G in posizione 230125827, mai riportata prima in dbSNP. Dall'analisi *in silico* è emerso che tale mutazione, causativa di un *frameshift*, ripristina il frame di lettura di *VPS13A*.

*AZGPI* è un gene candidato per l'adattamento all'ambiente (Herrero-Medrano *et al.*, 2014). Una mutazione in questo gene, localizzato in regioni in cui mappa un QTL per il numero di vertebre, grasso addominale e forma e dimensioni dell'orecchio è stata identificata in Mangalica, Cinta Senese e nel cinghiale europeo, ma non in suini commerciali (Herrero-Medrano *et al.*, 2014). La variabilità in questa regione potrebbe essere associata a diverse

firme di selezione tra razze suine, poiché la morfologia dell'orecchio è un carattere comunemente utilizzato per discriminare le razze. Inoltre, *AZGPI* è correlato alla mobilizzazione dei lipidi ed è considerato un gene candidato per la regolazione del peso corporeo e l'obesità nell'uomo (Mracek *et al.*, 2010). Abbiamo identificato due nuove mutazioni *frameshift* in questo gene, entrambe inserzioni di una C (localizzazione genomica: 7874326 e 7874521 del cromosoma 3), che potrebbero influenzare la funzione del gene *AZGPI*, con possibili effetti sulla deposizione dei grassi.

Il gene *LCORL* è localizzato in una regione genomica un cui è mappato un QTL coinvolto nelle modificazioni morfologiche occorse durante gli eventi di domesticazione che hanno riguardato l'allungamento della schiena e l'aumento del numero di vertebre (Rubin *et al.*, 2012). Questo gene è considerato un gene candidato per le dimensioni corporee. Una mutazione nota che determina un *frameshift* in *LCORL* (rs791023757; localizzazione genomica: 12829718) è stata rilevata in questo studio e risulta in un codone di stop mancante (*lacking stop codon*). Sfortunatamente nessun fenotipo è stato associato finora a questa mutazione, come evidenziato dalla mancanza di informazioni nel database dbSNP. Abbiamo identificato due varianti ad alto impatto anche nel gene *FUT1*. Questo gene codifica per una proteina di membrana coinvolta nella sintesi di un precursore dell'antigene H, necessario per la formazione degli antigeni A e B nei rispettivi gruppi sanguigni. Precedenti studi hanno dimostrato che i polimorfismi in questo gene sono associati all'adesione e alla capacità di colonizzazione di *Escherichia coli* fimbriato F18, alla mucosa intestinale (Meijerink *et al.*, 2000). Le tossine prodotte da questo microrganismo causano la diarrea post-svezzamento dei suinetti (Luo *et al.*, 2010). Abbiamo identificato una mutazione missenso in posizione 54079560 del cromosoma 6 (gene *FUT1*), che provoca una sostituzione amminoacidica in posizione 103 (Thr→Ala) della proteina. Questo SNP, già registrato in dbSNP (rs335979375), è risultato associato a genotipi *E. coli* F18-resistenti o suscettibili da Meijerink e collaboratori (Meijerink *et al.*, 2000). La seconda variante identificata era un inserzione di una G (posizione genomica: 54079637), ma ulteriori studi sono necessari per validare questi risultati e il ruolo di queste mutazioni in Nero Siciliano.

Il gene *PRLR* codifica per un recettore per la prolattina ed è considerato un gene candidato per vari tratti che influenzano direttamente (tasso di ovulazione) o indirettamente (peso ovarico, lunghezza uterina e numero di capezzoli) il numero dei suinetti della nidiata ed in generale le prestazioni riproduttive nel suino (Tomas *et al.*, 2006; Van Rens *et al.*, 2003; Vincent *et al.*, 1997). Anche nel gene *PRLR* abbiamo rilevato un inserzione di una di G in

posizione 20642378 (cromosoma 16), ma il suo contributo alla variazione fenotipica rimane da chiarire.

Il gene *TAS2R40* (Taste 2 receptor 40) ha mostrato invece, inaspettatamente, un ridotto numero di varianti. I recettori del gusto (TASR) sono coinvolti nell'identificazione e riconoscimento di vari composti chimici, e la famiglia *Tas2r* è deputata alla rilevazione del gusto amaro. Poiché il recettore T2R ha un importante ruolo nella capacità di discriminare ed evitare l'ingestione di sostanze tossiche (es. alcaloidi vegetali, tipicamente amari), mutazioni in T2R potrebbero riflettere diverse risposte di adattamento a diverse condizioni ambientali (Go *et al.*, 2005) e quindi delle preferenze e comportamenti alimentari, sebbene i suini mostrino una maggiore tolleranza nei confronti del sapore amaro rispetto l'uomo (Dong *et al.*, 2009; Fischer *et al.*, 2005). I polimorfismi nel gene *TAS2R40* potrebbero essere coinvolti nell'adattamento del suino Nero Siciliano a specifici repertori alimentari tipici delle aree in cui è tradizionalmente allevato (essenze boschive e prodotti del sottobosco), sebbene il processo di domesticazione sembra aver ridotto la variabilità nei geni *TAS2R* rispetto al cinghiale, suo progenitore selvatico (Ribani *et al.*, 2017). I nostri risultati hanno mostrato ridotto tasso di mutazione in questo gene nel soggetto analizzato. Solo 4 mutazioni sono state rilevate in *TAS2R40*, tutte classificate come modificatrici e tutte note. Di queste, due erano varianti geniche *downstream* (posizioni genomiche: 7025448 e 7025506), già presenti in dbSNP (rs326695329 e rs321250739 rispettivamente), mentre due erano varianti geniche *upstream* (posizioni genomiche: 7026930 e 7027039 associate rispettivamente a rs712711629 e rs334118595).

Per tutte le varianti *novel* e ad alto impatto identificate nel presente lavoro di tesi, nessuno studio di associazione è stato finora condotto. L'analisi *in silico* ha rilevato 7 mutazioni che potrebbero avere effetti sulla funzionalità della proteina. 4 di queste erano mutazioni *gain of function*. Poiché è possibile che le varianti derivino da errori di sequenza, o da errori nel genoma di riferimento, i polimorfismi dovrebbero essere validati ricorrendo al genotyping di più soggetti. Al fine di identificare varianti potenzialmente razza-specifiche, che possano con qualche probabilità riflettere lo specifico background genetico della razza, saranno dunque necessari ulteriori studi di genotyping mediante sequenziamento Sanger.



## 1.7. Conclusioni

Dalla pubblicazione del genoma di riferimento di *Sus scrofa* diversi progetti di risequenziamento sono stati intrapresi, ma pochi hanno riguardato razze locali. La caratterizzazione delle risorse genetiche è il primo passo per la loro gestione e valorizzazione. Molte razze autoctone presentano caratteristiche uniche che possono contribuire ad affrontare le sfide legate al cambiamento climatico, la crescente richiesta di cibo connessa all'aumento della popolazione mondiale e alla sicurezza alimentare, come evidenziato nel Domestic Animal Diversity Information System (DAD-IS) della FAO. In questo studio abbiamo sequenziato, per la prima volta, l'intero genoma di un verro di suino Nero Siciliano e condotto un'analisi di whole genome variant calling. Tuttavia, molto rimane ancora da fare dato che, al momento, è stato effettuato il solo mapping ed il variant calling. Prossimo *step* riguarderà l'Assembly del genoma.

Questo lavoro rappresenta un primo contributo alla caratterizzazione genetica della razza, volta all'identificazione di varianti che potrebbe spiegare le basi genetiche dei suoi peculiari tratti fisiologici e / o morfologici, e rappresentare la base di partenza per indagare la tipicità genetica del suino Nero Siciliano. È noto come le razze autoctone siano depositarie di combinazioni alleliche andate perdute a seguito dei severi programmi selettivi e possono dunque rappresentare germoplasma potenzialmente utile per introgressioni future, in quanto fonte di variabilità genetica. Il loro impiego, oltre a fornire prodotti di elevata qualità ed eticamente sostenibili in quanto rispettosi del benessere animale, potrebbe consentire anche una più ampia e razionale utilizzazione di territori marginali.

Ulteriori studi sono necessari per meglio comprendere le differenze genetiche che sottendono la diversità fenotipica del suino Nero Siciliano; studi che forniranno un contributo alla conservazione, valorizzazione e gestione della razza.

## 1.8. Disponibilità di dati di supporto

Le raw reads utilizzate per l'analisi dell'intero genoma sono state depositate nel Sequence Read Archive (SRA) dell'NCBI con il seguente numero di accesso: SRX3406507 (<https://www.ncbi.nlm.nih.gov/sra/SRX3406507/>).

I file a supporto del presente lavoro (SNPs e short INDELs rilevati da SUPERW nei geni correlati alla fitness e loro classificazione in categorie sulla base dell'impatto sulla

funzionalità proteica; varianti strutturali rilevate da PINDEL, SNPs *novel* rilevati in suino Nero Siciliano da snpSift vs dbSNP151), sono disponibili presso gli Autori.

Il presente studio è stato sottomesso alla rivista scientifica *Genetics and Molecular Biology* (<https://www.gmb.org.br/>) per la pubblicazione nel giugno 2018 (in revisione).

# 1. Parte Seconda

## 2.1. Analisi comparativa *in silico* di geni candidati per la qualità della carne in genomi di suini *Nero Siciliano, Duroc, Large White e Landrace*

### 2.2. Abstract

#### Background

In recent decades, many local breeds have been subjected to genetic erosion and loss of biodiversity resulting in the impoverishment of a precious gene pool that has mainly affected marginal areas and low input breeding systems. Nero Siciliano pig is a local breed reared mainly in the Nebrodi Mountains of Sicily (Italy). In 2003 was established a Consortium for the valorization of its productions and a request to label the fresh Nero Siciliano meat with the Protected Denomination of Origin (PDO) was issued in 2005. The request for the PDO has been started also for Nero Siciliano's cured ham in 2011.

In this study we report an *in silico* comparison of 48 candidate genes involved in meat quality traits, retrieved by using Nero Siciliano (NS), Large White (LW), Landrace (LAN) and Duroc (DU) genomes. The latter is the reference genome for pig (Sscrofa11.1).

We focused on genes related to muscle mass deposition and carcass fatness as these traits influence technological processes adopted for long matured pork meats products such as cured ham.

#### Results

More than twenty thousand variants were identified by comparing the gene set of each breed to the reference genome assembly. Of these ~22,000 were SNPs, ~3,000 short insertions and ~1,400 short deletions. Transitions / transversions ratio was 2.650 while missense/silent ratio resulting in 0.526. Furthermore, over 40% of intronic variants and ~45% of non coding transcript variants were also identified.

Among all variants detected in this study, more than 3,000 were shared among NS, LW and LAN while ~7,000 were unique for NS, ~2,000 for LW and ~6,000 for LAN. Among the unshared SNPs (5,659) of NS, 856 were in homozygosity while 4,803 in heterozygosity; 802 were novel while 4,857 were already present in dbSNP.

#### Conclusions

In recent decades, the FAO (Food and Agriculture Organization) has expressed concern about the progressive replacement of local breeds with improved cosmopolitan breeds, since the latter cannot compete with the autochthonous ones characterized by rusticity, resistance to disease and adaptation to reduced food availability. Many indigenous breeds'

present unique characteristics that can contribute to tackling the challenges linked to climate change, the growing demand for food connected to the increase in the world population and food security. In this study, we identified 7,293 unique variants in a subset of genes fatness-related, in Nero Siciliano pig. Unshared SNPs and fixed in the breed could be the starting point for a molecular detection of breed specific DNA markers useful for breed differentiation and the authentication of their products. Methods for the authentication of breed-specific products are key tools to defend the added economic value of these products that represent the strategy to obtain a sustainable conservation of local animal genetic resources.

**Keywords**

Meat quality traits, SRA, *in silico* comparison, shared variants, unique variants, fatness.

### 2.3. Introduzione

La deposizione del grasso è, in generale, uno dei processi metabolici chiave nella zootecnia da carne ed, in particolare, nel comparto suinicolo, con rilevanti implicazioni economiche e conseguenze sulla gestione aziendale, in quanto influisce sui parametri di efficienza alimentare e sulle caratteristiche qualitative delle carcasse. Come avvenuto in molti comparti, anche in quello suinicolo si è assistito alla graduale sostituzione di numerose razze e popolazioni autoctone con quelle migliorate di importazione ed in particolare con la Large White e la Landrace. Queste due razze nordeuropee, che coniugano un'elevata efficienza riproduttiva, buone performances di crescita, buona qualità delle carcasse ed ottima qualità delle carni, costituiscono i tipi genetici su cui si basa la produzione del suino pesante in Italia (<http://www.anas.it/>).

La FAO, come già riportato, ha espresso preoccupazione per la progressiva sostituzione di razze locali con razze cosmopolite migliorate, considerato che quest'ultime non possono competere con quelle autoctone in termini di rusticità, resistenza alle malattie e adattamento alle ridotte disponibilità alimentari (FAO; Muir *et al.*, 2008). Pertanto, la tutela e la valorizzazione del suino Nero Siciliano assumono una connotazione di natura tecnico-pratica, oltre che etica.

Tale razza risulta iscritta al Registro Anagrafico, strumento per la conservazione delle razze suine non sottoposte ad un piano nazionale di selezione. Negli ultimi decenni, al fine di salvaguardare la razza dal rischio di estinzione, molte azioni sono state intraprese. Nel 2003 è stato istituito il Consorzio di Tutela del suino Nero dei Nebrodi e, nel 2005, è stata presentata l'istanza di riconoscimento della DOP della carne e derivati di suino Nero dei Nebrodi, da parte dello stesso. In data 18.07.11 è stata presentata domanda di riconoscimento della DOP prosciutto crudo di suino Nero dei Nebrodi e, nel 2017, è stato siglato l'atto costitutivo dell'ATS "Associazione per la Tutela del prosciutto crudo ed il salame del suino Nero dei Nebrodi". Il 14.04.2018 il Disciplinare della DOP "*Prosciutto crudo di suino Nero dei Nebrodi*" e del Marchio collettivo di qualità sono stati presentati in occasione del Convegno "Promozione e valorizzazione del suino Nero dei Nebrodi", a S.Agata Militello. Nel corso del 2018, infine, come riportato nella Parte Prima del presente lavoro di tesi, è stato sequenziato l'intero genoma di un soggetto di suino Nero Siciliano. Questo ha rappresentato un notevole passo avanti nella caratterizzazione e valorizzazione della razza che acquisirà maggiore valenza dopo la pubblicazione dell'Assembly.

L'avanzamento tecnologico ed il progresso nel campo del sequenziamento dei genomi rendono infatti oggi possibile reperire informazioni utili ad interpretare i meccanismi biologici che sottendono alla manifestazione fenotipica dei tratti di interesse, anche se una maggiore comprensione deriverà dall'applicazione di altri approcci "omici" (Van Emon, 2016). È noto come la modulazione dell'espressione genica possa avvenire a diversi livelli e la trascrittomica e lo studio dei meccanismi epigenetici saranno in grado di fornire, nel prossimo futuro, risposte molto più accurate. Diversi studi hanno recentemente indagato, con approccio RNASeq, il profilo trascrizionale del tessuto del grasso dorsale in Italian Large White (Zambonelli *et al.*, 2016), al fine di identificare geni differenzialmente espressi in soggetti magri *vs* grassi; il profilo di espressione di small RNA (sRNA) (Davoli *et al.*, 2018) e quello di micro-RNAs (miRNAs) e miRNA-offset RNAs (moRNAs) (Gaffo *et al.*, 2014). Questi studi sottolineano quanto complesso sia il meccanismo bio-molecolare che regola la deposizione dei grassi nei suini, e quanto incomplete siano ancora le nostre conoscenze.

In questa Parte Seconda del lavoro di tesi abbiamo focalizzato l'attenzione su un set di 48 geni coinvolti in processi metabolici correlati al metabolismo lipidico. La deposizione dei grassi assume infatti un ruolo strategico non solo per i noti effetti tecnologici in fase di stagionatura (Bosi & Russo, 2004), in particolare per la produzione del prosciutto crudo (ANAS, 2016), ma anche per gli importanti riflessi sulle caratteristiche qualitative e nutrizionali del prodotto. L'analisi comparativa *in silico* ivi condotta ha avuto lo scopo di fornire una panoramica generale delle varianti presenti nei genomi delle razze che, seppure con peculiarità tipiche ed esclusive dei circuiti DOP, sono quelle maggiormente allevate in Italia. Si tratta dunque di uno screening preliminare che, non avendo la pretesa di essere esaustivo, rappresenta solo un ulteriore *step* nel processo di caratterizzazione della razza.

## **2.4. Materiali e Metodi**

### **2.4.1. Analisi comparativa *in silico***

Le raw reads di Large white (SRR3123346; SRR3123347), Landrace (ERX2292210) e Nero Siciliano (SRX3406507) sono state scaricate dal Sequence Read Archive (SRA) database di NCBI (Leinonen *et al.*, 2011). I dati relativi il suino Nero Siciliano sono stati prodotti in un nostro recente studio (dato non pubblicato, manoscritto in review), presentato nella Parte Prima del presente lavoro di tesi.

Le raw reads, dopo ispezione con FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), sono state processate con Trimmomatic v. 0.36 (version 0.30; opzioni usate: `-windowsSize 4 -requiredQuality 28 -MINLEN 35, LEADING: 30, TRAILING: 30`) (Bolger *et al.*, 2014) al fine di rimuovere gli adattatori e le sequenze di bassa qualità, e mappate sul genoma di riferimento *Sscrofa* 11.1 (GenBank: GCA\_000003025.6) con BWA (Burrows-Wheeler Aligner, version 0.7.5; opzioni usate: `mem: BWA-MEM algorithm`) (Li & Durbin, 2009). La qualità del mapping è stata valutata con Qualimap2 (Okonechnikov *et al.*, 2016).

Le analisi relative i polimorfismi a singolo nucleotide (SNPs) e le brevi (size < 10 bp) inserzioni e delezioni (INDEL) sono state condotte con SUPERW (Sanseverino *et al.*, 2015) che integra la funzione mpileup di SAMTOOLS e GATK (Genome Analysis Toolkit).

Dopo rimozione di tutte le varianti che mostravano lo stesso genotipo del riferimento, le rimanenti sono state ulteriormente filtrate utilizzando i seguenti parametri:  $QUAL \geq 30$ ,  $DP \geq 10$ ,  $AD \geq 10$ .

Le coordinate genomiche di 48 geni correlati al metabolismo lipidico (Tabella 6) sono state estratte dal VCF (Variant Calling Format) ricorrendo al linguaggio di scripting awk (Aho, Weinberger, Keringhan; 1979) ed i putativi effetti funzionali degli SNPs, identificati nel *cluster* di geni, sono stati poi predetti *in silico* con il software SnpEff v4\_3m\_core (Cingolani *et al.*, 2012a). Il tool VEP (Variant Effect Predictor) di Ensembl (Ensembl release 92 - April 2018) (McLaren *et al.*, 2016), che implementa l'algoritmo SIFT (Kumar *et al.*, 2009) per la predizione di varianti con probabili effetti deleteri sulla funzionalità proteica e APPRIS (Rodriguez *et al.*, 2013) per l'annotazione delle isoforme del trascritto principale dovute a diverse varianti di splicing del gene, è stato utilizzato per confermare le predizioni di SnpEff ed identificare le varianti *novel vs* già descritte. In particolare, per specie con un *Ensembl database of variants* come *Sus scrofa*, VEP può annotare le varianti sulla base delle evidenze sperimentali presenti nei seguenti database: dbSNP (Version 150), Animal\_QTLdb (version 21/12/2017), DGVa (Database of Genomic Variants Archive, version 03/2017), OMIA (Online Mendelian Inheritance in Animals, version 05/02/2018).

I geni candidati per i loro coinvolgimento in processi correlati al metabolismo lipidico e alla deposizione dei grassi, sono stati identificati sulla base di un sistematico studio della letteratura di riferimento (**IGF2**: Fontanesi *et al.*, 2010b; Jeon *et al.*, 1999; Jungerius *et al.*, 2004; Nezer *et al.*, 1999; Oczkowicz *et al.*, 2009; Wan Laere *et al.*, 2003; **LEP**; **LEPR**:

Jiang *et al.*, 1999; Kennes *et al.*, 2001; Amills *et al.*, 2008; de Oliveira Peixoto *et al.*, 2006; Barb *et al.*, 2006; Chen *et al.*, 2004; **MC4R, LEP, H-FABP**: Chao *et al.*, 2012; Galve *et al.*, 2012; **LEP, LEPR, MC4R, PIK3C3, VRTN**: Hirose *et al.*, 2014; Muñoz *et al.*, 2011; Óvilo *et al.*, 2005; **SCD**: Ren *et al.*, 2004; Ros-Freixedes *et al.*, 2016; **PIK3C3**: Hirose *et al.*, 2011; Kim *et al.*, 2005a,b; **CTSB, CTSD, CSTB**: Russo *et al.*, 2002; Russo *et al.*, 2008; **TRIB3**: Prudente *et al.*, 2009; Fontanesi *et al.*, 2010c; **PCSK1**: Fontanesi *et al.*, 2012a; **PDE1C, CRISPI, STAT4, STAB1**: Fontanesi *et al.*, 2012b; Benzinou *et al.*, 2008; **MUC4**: Fontanesi *et al.*, 2012c; **FTO**: Fontanesi *et al.*, 2008a; Fontanesi *et al.*, 2010d; Dvořáková *et al.*, 2012; Fan *et al.*, 2009; **TBC1D**: Chadt *et al.*, 2008; Fontanesi *et al.*, 2011; Brandl *et al.*, 2013; **GPR120**: Fontanesi *et al.*, 2015; Waguri *et al.*, 2013; **PLIN1, PLIN2**: Davoli *et al.*, 2011; Ruiz *et al.*, 2011; Vykoukalová *et al.*, 2009; **ATPIA2, CA3, DECRI**: Davoli *et al.*, 2006; Pena *et al.*, 2013; **ATPIB1, APOE, CTSK, FABP4, GNRHR, IGF-1, LIF, MTHFR, PPARGCIA, PPARG, RETN, TNNI1, TNNI2, VDBP, PPYRI**: Switonski *et al.*, 2010; **SCD, FASN, LIPE**: Zappaterra *et al.*, 2016) e dalla consultazione del database di pathway metabolici KEGG (<http://www.genome.jp/kegg/>).

## 2.5. Risultati

L'analisi *in silico* sui 48 geni correlati al metabolismo lipidico ha rilevato la presenza di 26649 varianti. In tabella 6 è riportato il numero di SNPs e short INDEL per ciascun gene, che è variato da un minimo di 3 al locus *FABP4* ad un massimo di 6423 al locus *PPARGCIA*. Il tasso di variazione, normalizzato sulla lunghezza del gene, è risultato compreso tra lo 0,03 di *VDBP* e l'1,78 di *SCPEP1*.

L'annotazione funzionale delle varianti identificate, ne ha permesso la classificazione in base alla tipologia (*silente/missense*) ed alla localizzazione genomica. La maggior parte delle varianti identificate erano localizzate nelle regioni introniche (46,943%) e in regioni ncRNA (SO:0001619; 47,583%), lo 0,213% nel 5'-UTR e lo 0,63% nel 3'-UTR.

Tra il totale delle varianti identificate ~ 22000 erano SNPs, con un rapporto Transizioni / Trasversioni risultato in media pari a 2,650, mentre le inserzioni (~ 3000) sono risultate più rappresentate delle delezioni (~ 1400).



Gene	N° Variants (SNPs/short INDEL)		Seq_name	Chr	Start	End	Length	%variants/length
	SNPs	short INDEL						
IGF2	95		CM000813.5	2	1469183	1496417	27235	0,349
LEP	222		CM000829.5	18	20106867	20124071	17205	1,290
LEPR	1028		CM000817.5	6	146802297	146896152	93856	1,095
MC4R	1		CM000812.5	1	160772013	160774124	2112	0,047
PIK3C3	1448		CM000817.5	6	125890598	126038753	148156	0,977
CTSB	221		CM000825.5	14	15014139	15035081	20943	1,055
CTSD	75		CM000813.5	2	1188548	1197642	9095	0,825
TRIB3	95		CM000828.5	17	34808273	34819592	11320	0,839
PCSK1	152		CM000813.5	2	103004794	103050707	45914	0,331
MUC4	135		CM000824.5	13	134192412	134248435	56024	0,241
FTO	2369		CM000817.5	6	31177112	31564674	387563	0,611
TBC1D1	2986		CM000819.5	8	29370344	29590574	220231	1,356
PDE1C	3536		CM000829.5	18	40818316	41410555	592240	0,597
CRISP1	416		CM000818.5	7	43883536	43911877	28342	1,468
STAT4	375		CM000826.5	15	95656206	95764099	107894	0,348
GPR120	72		CM000825.5	14	105011408	105033633	22226	0,324
CSTB	11		CM000824.5	13	206706060	206710598	4539	0,242
SCPEP1	558		CM000823.5	12	33265509	33296805	31297	1,783
SCD	66		CM000825.5	14	111461560	111478033	16474	0,401
FASN	55		CM000823.5	12	922405	937560	15156	0,363
LIPE	31		CM000817.5	6	49543671	49560126	16456	0,188
PLIN1	35		CM000818.5	7	55237517	55250797	13281	0,264
PLIN2	139		CM000812.5	1	203683867	203709277	25411	0,547
CA3	7		CM000815.5	4	51227485	51237861	10377	0,067
DECR1	179		CM000815.5	4	46731359	46767427	36069	0,496
STAB1	84		CM000824.5	13	34630448	34659371	28924	0,290
IDH3B	27		CM000828.5	17	32955061	32960110	5050	0,535
FYB	1900		CM000827.5	16	24332539	24507768	175230	1,084
FABP3	33		CM000817.5	6	87943003	87950708	7706	0,428
MTTP	492		CM000819.5	8	120820660	120871468	50809	0,968
PGAM2	316		CM000829.5	18	48693832	48712787	18956	1,667
CAST	1389		CM000813.5	2	103255738	103378623	122886	1,130
LEPROT	166		CM000817.5	6	146974641	146987261	12621	1,315
ATP1B1	39		CM000815.5	4	81912543	81937015	24473	0,159
APOE	5		CM000817.5	6	51373113	51375333	2221	0,225
CTSK	44		CM000815.5	4	98389459	98404969	15511	0,284
FABP4	3		CM000815.5	4	55096400	55101209	4810	0,062
GNRHR	26		CM000819.5	8	65470206	65488900	18695	0,139
IGF-1	563		CM000816.5	5	81762027	81909253	147227	0,382
LIF	58		CM000825.5	14	47221540	47239513	17974	0,323
MTHFR	262		CM000817.5	6	71863637	71882118	18482	1,418
PPARGC1A	6423		CM000819.5	8	17841844	18527953	686110	0,936
PPARD	328		CM000818.5	7	31222533	31295221	72689	0,451
RETN	5		CM000813.5	2	71484164	71485577	1414	0,354
TNNI1	98		CM000821.5	10	23788279	23800554	12276	0,798
TNNI2	17		CM000813.5	2	1251566	1253883	2318	0,733
VDBP	13		CM000819.5	8	68326287	68364937	38651	0,034
PPYR1	51		CM000825.5	14	88276340	88283720	7381	0,691

Tab. 6- Varianti identificate nei 48 geni correlati alla Fatness in tutti i genomi analizzati in questo studio

L'obiettivo di annotare le varianti è quello di fornire una previsione di quali siano funzionalmente rilevanti. In questo studio, solo 5 varianti sono state classificate come aventi alto impatto sulla funzionalità della proteina. In particolare 2 varianti sono state predette come causative di frameshift (SO\_0001589, ai loci *PLIN1* e *CTSB*), 1 come causativa di perdita di start codon (al locus *TBC1D1*) e 2 come modificatrici dei siti di splicing (accettore al locus *PPYR1*, SO\_0001574 e donatore al locus *PGAM2*, SO\_0001575). Tutti gli SNPs identificati in questo studio e classificati come aventi alto impatto, erano già stati descritti e depositati in dbSNP con i seguenti rs code: rs339837205 (*TBC1D1*), rs335709223 (*PPYR1*), rs330054959 (*PGAM2*), come rilevato dal tool VEP di Ensembl, ma per nessuno di essi erano disponibili dati fenotipici in associazione. Tali SNPs erano presenti in omozigosi alternativa solo in LAN (*TBC1D1*), in eterozigosi in NS e LAN (*PPYR1*) ed in eterozigosi solo in LAN al locus *PGAM2*.

Le short INDEL (loci *PLIN1* e *CTSB*), identificate da SnpEff e classificate come frameshift variants, erano *novel*.

L'elenco delle varianti aventi effetto ad alto impatto sulla funzionalità proteica è mostrato in tabella 7.

Il rapporto missenso / silente è risultato pari a 0,526, mostrando come la maggior parte delle varianti identificate fossero sinonime, non determinando una variazione nella sequenza amminoacidica.

Chr	Position	Gene	Type	Effect/Impact	Ref	Alt
CM000818.5	55237669	PLIN1	INDEL	TCCCCC frameshift_variant HIGH PLIN1 transcript	TCCCCC	TCCCCC
CM000819.5	29485163	TBC1D1	SNP	G start_lost HIGH TBC1D1 transcript	A	G
CM000825.5	15034647	CTSB	INDEL	CGGGGGGGG frameshift_variant HIGH CTSB transcript	CGGGGGGGG	CGGGGGGGG
CM000825.5	88276518	PPYR1	SNP	C splice_acceptor_variant&intron_variant HIGH PPYR1	T	C
CM000829.5	48709077	PGAM2	SNP	C splice_donor_variant&intron_variant HIGH PGAM2	T	C

**Tab.7- Varianti aventi effetto ad alto impatto sulla funzionalità proteica identificate da SnpEff.** È riportata la localizzazione genomica (cromosoma e posizione), il tipo di variante, il gene in cui è localizzata e la descrizione completa dell'effetto sulla proteina corrispondente in base alla predizione SnpEff.

Tra tutte le varianti rilevate in questo studio, più di 3000 erano condivise tra NS, LW e LAN mentre ~ 7000 erano uniche per NS, ~ 2000 per LW e ~ 6000 per LAN (Figura 2), mostrando un buon grado di variabilità genetica tra razze studiate.



**Fig. 2- Diagramma di Venn.** Il diagramma mostra il totale delle varianti condivise (interne) ed uniche (esterne) tra i genomi, identificate in questo studio.

Per meglio caratterizzare le varianti risultate esclusive di suino Nero Siciliano nel presente lavoro, le 7293 INDEL e i SNPs identificati in NS ma non nei genomi di LW, LAN e Duroc, sono stati filtrati ed i soli SNPs sottoposti a screening con il tool VEP. Delle 7293 varianti non condivise, 1634 erano INDEL mentre 5659 erano SNPs. Di questi, 856 erano presenti in omozigosi alternativa, 4803 allo stato eterozigote (Tabella 8). ~ l'86 % delle varianti in omozigosi erano già note, mentre il 13 % erano novel. La maggior parte degli SNPs erano localizzati nelle regioni introniche (Figura 3) mentre pochi in quelle *upstream* e *downstream* che, secondo la terminologia di Sequence Ontology (SO) utilizzata da Ensembl, comprende le varianti localizzate entro 5 kb dal trascritto. Nessuna delle varianti in omozigosi è stato predetto avere effetti ad alto impatto sulla funzione biologica della proteina, ad eccezione della mutazione missenso g. 160773437 G> A (rs81219178) al *locus MC4R* (ENSSSCG00000004904), che determina una sostituzione amminoacidica in posizione 298 (Asp→Asn) della proteina. Questa mutazione è l'unica classificata come "deleteria" (score 0,01) da SIFT (sorts intolerant from tolerant).

Anche le varianti in eterozigosi hanno mostrato la stessa proporzione tra note e *novel*.

Category	Count	
	Homozygous	Heterozygous
Variants processed	856	4803
Novel variants	113 (13.2)	689 (14.3)
Existing variants	743 (86.8)	4114 (85.7)

Tab. 8- Statistiche relative gli SNPs identificati in NS ma non in LAN, LW e Duroc (Sscrofa 11.1 Assembly).

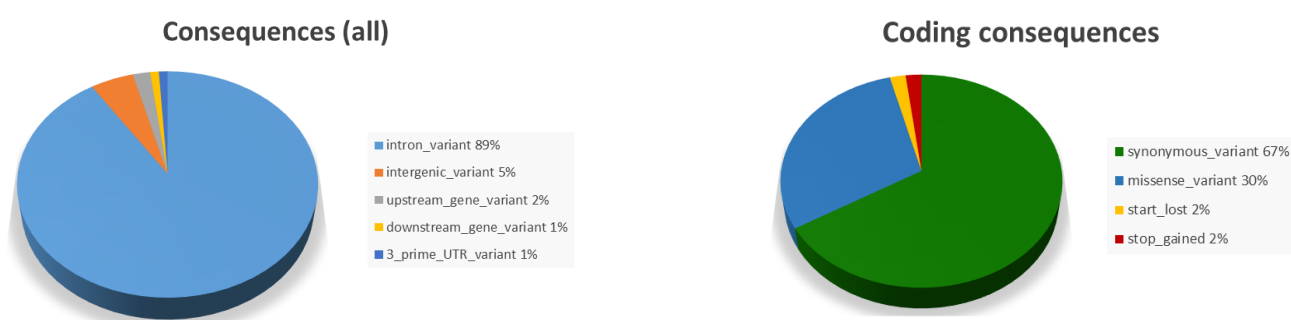


Fig. 3- VEP output per gli SNPs identificati in NS ma non in LAN, LW e Duroc (Sscrofa 11.1 Assembly). I pie charts mostrano la classificazione delle varianti e le relative conseguenze.

## 2.6. Discussioni

L'analisi di SNPs discovery nei 48 geni correlati al metabolismo lipidico ha mostrato che più di 3000 varianti erano condivise tra NS, LW e LAN mentre ~ 7000 erano uniche per NS, ~ 2000 per LW e ~ 6000 per LAN. Di tutte le varianti identificate, solo 5 sono state classificate come aventi alto impatto sulla funzionalità proteica, ai loci *PLIN1*, *TBC1D1*, *CTSB*, *PGAM2* e *PPYR1*.

Le perilipine (*PLIN*) appartengono ad una famiglia di proteine strutturali con funzioni regolatrici sul deposito e la mobilizzazione dei lipidi intracellulari. Sono localizzate sulla superficie delle gocce lipidiche contenute negli adipociti. Esistono cinque diverse famiglie di geni *PLIN* (del tipo 1, 2, 3 4 e 5) che codificano per cinque diversi tipi proteici. In particolare, la proteina perilipina-1 (*PLIN1*) è maggiormente rappresentata negli adipociti maturi mentre la perilipina-2 (*PLIN2*) è la proteina più preminente negli adipociti immaturi. Mutazioni nei geni *PLIN* sono state associate alla massa grassa corporea in topo

(Saha *et al.*, 2004) e uomo (Qi *et al.*, 2004; Corella *et al.*, 2005; Ruiz *et al.*, 2011), mentre nel suino pochi lavori hanno indagato l'associazione di polimorfismi ai loci *PLIN1* e *PLIN2* con tratti di produzione. Vykoukalová *et al.* (2009) ha identificato due polimorfismi negli esoni 3 e 6 di *PLIN1* associati con ADG e BFT in Italian Large White, mentre Davoli *et al.* (2011) ha mostrato l'associazione di un polimorfismo nella regione 3'-UTR del gene *PLIN2* (g.98G> A) con l'accrescimento ponderale giornaliero, l'indice di conversione alimentare e la percentuale di tagli magri in Duroc, con effetto favorevole dell'allele A. Tale associazione è stata anche confermata da Gol *et al.*, (2015), che non ha invece trovato correlazione tra mutazioni al locus *PLIN1* con tratti di crescita e deposizione grassi in Duroc.

Nel presente lavoro di tesi una short INDEL, in posizione 55237669 del gene *PLIN1* è stata identificata, condivisa tra NS e LAN. Tale mutazione non è stata prima descritta in nessuna razza e non sono disponibili dati di associazione.

Il dominio *TBC1* (*TBC1 domain family member 1*) della famiglia di proteine 1, proteina attivante la GTPasi-Rab coinvolta nella traslocazione del trasportatore 4 del glucosio (GLUT4) a livello della membrana cellulare (Brandl *et al.*, 2013), è considerato un candidato per la regolazione del metabolismo del glucosio.

Due studi indipendenti sull'uomo hanno dimostrato un'associazione tra la variante R125W del gene *TBC1D1* (rs35859249) ed il maggiore rischio di obesità (Stone *et al.*, 2006; Meyre *et al.*, 2008) e, nel 2007, due diversi studi di Genome-wide association analysis hanno rilevato l'associazione di SNPs in *TBC1D1* con il diabete umano di tipo 2 (Saxena *et al.* 2007; Zeggini *et al.* 2007). In topi knock-out per questo gene è stato inoltre riscontrato un peso corporeo minore ed una ridotta captazione di glucosio a livello del muscolo scheletrico (Chadt *et al.*, 2008).

Sulla base di tali evidenze sperimentali in umano e nel topo e considerando le funzioni biologiche ed i processi metabolici in cui è coinvolta la proteina *TBC1D1*, il gene codificante la proteina è stato studiato da Fontanesi *et al.* (2011b) ai fini di rilevare polimorfismi associati ai tratti correlati alla deposizione di grassi nei suini. Cinque polimorfismi sono stati identificati e uno di questi (GenBank FN677935: g.219G >A) è risultato essere associato al peso del prosciutto, allo spessore del grasso dorsale ed alla percentuale di tagli magri in Italian Large White, al grasso intermuscolare visibile in Italian Duroc. Lo stesso Fontanesi nel 2012 (Fontanesi *et al.*, 2012d) ha identificato ulteriori tre SNPs al locus *TBC1D1*, due dei quali (g.40A> G e g.151C> T) in *linkage disequilibrium* con g.219G>A, associati allo spessore del grasso dorsale in Italian Large White.

Nel presente lavoro di tesi una mutazione missenso (rs339837205) è stata identificata in posizione 29485163. Questo SNP causa perdita di start codon (ATG→GTG, posizione 1331 del mRNA) con conseguente variazione di un residuo amminoacidico M [Met] ⇒ V [Val]. Pur trattandosi di una mutazione nota, nessun dato fenotipico in associazione è riportato per questo polimorfismo.

Il gene *CTSB* (*cathepsin B*) codifica per proteasi lisosomiali a localizzazione prettamente muscolare. La catepsina B è un membro della famiglia degli enzimi proteolitici ed è espressa nei lisosomi citoplasmatici di diversi tipi cellulari, in tessuti normali e neoplastici. Si tratta di una cistein-proteasi e, come molte catepsine, è coinvolta a nel metabolismo cellulare come proteina di degradazione.

L'elevata attività catepsinica del muscolo scheletrico suino è stata correlata al difetto dell'eccessiva morbidezza della carne nei prosciutti crudi, colore scuro, retrogusti astringenti o metallici, depositi di cristalli di tirosina, e la formazione di pellicole bianche sulla superficie di taglio (Virgili *et al.*, 1995; Virgili *et al.*, 1998), che rappresentano dall'1-3% circa delle motivazioni per cui le cosce sono scartate nella produzione di prosciutti DOP o di alta qualità in paesi come Italia, Spagna, Francia e Slovenia (Russo & Nanni Costa, 1995; Virgili & Schivazappa, 2002; Bosi & Russo, 2004).

Considerando gli effetti dell'attività della catepsina nel determinare le caratteristiche sensoriali e la consistenza di prosciutti stagionati, i geni che codificano per questi enzimi lisosomiali e loro inibitori possono essere considerati candidati per le caratteristiche qualitative delle carni. Nel presente lavoro di tesi una INDEL, in posizione 15034647, è stata identificata al locus *CTSB* ma anche in questo caso non è possibile avanzare alcuna ipotesi sui possibili effetti funzionali sulla proteina e dunque sui tratti correlati alla qualità dei trasformati.

*PGAM2* (*phosphoglycerate mutase 2*) codifica per enzimi coinvolti nei pathway metabolici implicati nella via glicolitica, prima della trasformazione anaerobica del piruvato in lattato. Nel suino è localizzato sul cromosoma 18 in una regione dove diversi QTLs per drip loss, colore della carne, deposizione di grasso, contenuto di carne magra, diametro della fibra muscolare e qualità della carcassa sono stati mappati (Bidanel *et al.*, 2001; Malek *et al.*, 2001; Nezer, Moreau, Wagenaar, & Georges, 2002; Ovilò *et al.*, 2002; Rohrer & Keele, 1998a,b). Un polimorfismo al locus *PGAM2* (g.77 C/A) è stato associato al peso del prosciutto in Large White (Fontanesi *et al.*, 2008b), mentre una mutazione identificata in sequenza parziale di cDNA isolato da muscolo scheletrico (accession number Z98802, Davoli *et al.*, 2002), è risultata significativamente associata drip loss in Large White,

Landrace, Duroc, Belgian Landrace, Hampshire, Piétrain e Meishan (Fontanesi *et al.*, 2003).

Nel presente lavoro di tesi una mutazione al *locus* *PGAM2* è stata identificata, esclusivamente in LAN, in posizione 48709077 (g.48709077T>C). Tale mutazione, classificata come donatrice di sito di splicing, risulta già depositata in dbSNP (release 150) con ID: rs330054959. Non sono noti dati fenotipici in associazione.

*PPYR1* (*Pancreatic Polypeptide Receptor 1*) è un gene con un ruolo chiave nella regolazione dell'omeostasi energetica e nell'assunzione di nutrienti. Studi su uomo (Sha *et al.*, 2009) hanno suggerito che copy number variation (CNV) in questa regione potrebbero essere associati a tratti correlati all'obesità, in particolare all'indice di massa corporea (IBM), in popolazioni cinesi. Basandosi su un approccio di genomica comparativa per la selezione di candidati volti all'identificazione di polimorfismi associati a tratti di adiposità della carcassa in suino (Switonski *et al.*, 2010; Stachowiak *et al.*, 2016), tale gene è stato incluso nel *cluster*. Nel presente lavoro di tesi uno SNP (rs335709223), in posizione 88276518 del gene *PPYR1*, è stato identificato in omozigosi alternativa in LAN. Tale mutazione, che determina la sostituzione di una timina con una citosina, causa la formazione di un sito accettore di splicing. Questo potrebbe determinare la sintesi di isoforme proteiche con diversa influenza sulla modulazione del fenotipo.

Tra le varianti in omozigosi esclusiva del soggetto NS sequenziato solo la mutazione missenso g. 160773437 G> A (rs81219178) al *locus* *MC4R* è stata classificata come deleteria da SIFT. Il gene *MC4R* (*Melanocortin-4 receptor*) è coinvolto nel controllo del bilancio energetico (Seeley *et al.*, 1997) e diversi studi sono stati intrapresi al fine di chiarirne il ruolo e valutarne l'associazione con i tratti di adiposità delle carcasse. La mutazione missenso identificata in questo studio, già nota in letteratura, è responsabile della sostituzione amminoacidica Asp298Asn (NP\_999338.1) nella settima regione transmembrana del recettore melanocortina-4 (Kim *et al.*, 2000). Tale mutazione ha mostrato avere significativa associazione con ingrassamento, tasso di crescita e assunzione di cibo in numerose razze (Bruun *et al.*, 2006; D'Andrea *et al.*, 2008; Jokubka *et al.*, 2006; Hernandez-Sanchez *et al.*, 2003; Huston *et al.*, 2004; Meidtner *et al.*, 2006; Van den Maagdenberg *et al.*, 2007). Più recentemente l'associazione del polimorfismo con tratti di adiposità delle carcasse è stata valutata da Davoli *et al.* (2012), mostrando come questo influenzi l'indice di conversione alimentare, l'accrescimento giornaliero ed il peso dei prosciutti in Italian Duroc ed Italian Large White, la percentuale di tagli magri in Italian Duroc (con animali portatori del genotipo AA che presentavano i migliori EBV rispetto

agli animali GG) e lo spessore del grasso dorsale in Italian Large White (con animali portatori del genotipo AA che presentavano maggiore spessore del grasso dorsale e maggior tasso di crescita di quelli GG). I risultati di associazione per l'Italian Large White sono stati confermati anche da Fontanesi *et al.* (2013).

Nel sopracitato studio, Davoli e collaboratori sottolineano come l'effetto allelico abbia però mostrato marcate differenze per il carattere "percentuale di tagli magri" e come l'analisi della varianza non abbia evidenziato nessuna associazione tra lo SNP analizzato e i tagli magri in ILW e con lo spessore del grasso dorsale in IDU. Come riportato da Bruun *et al.* (2006), questo potrebbe essere dovuto al fatto che c.1426G> A non sia la mutazione causativa ma sia con essa in *linkage disequilibrium* (Ibeagha-Awemu *et al.*, 2008) e che possa esserci interazione epistatica di altri geni con effetto sull'utilizzo dei nutrienti, indirizzando la crescita verso la deposizione di tessuto adiposo o muscolare in funzione dei diversi background genetici e metabolici delle diverse razze.

Alla luce di queste considerazioni, e di quanto riportato in letteratura, sarebbe auspicabile condurre uno studio di associazione nella razza suino Nero Siciliano al fine di comprendere il ruolo del gene *MC4R* nei processi che regolano l'assunzione di cibo e la deposizione del grasso.

## **2.7. Conclusioni**

Nel corso degli ultimi anni, gli studi genetici sulla qualità della carne suina sono diventati oggetto di un interesse sempre crescente. In questo senso, l'aumento dell'attenzione da parte dei consumatori nei riguardi della qualità degli alimenti, ha indotto il sistema allevatorio a porre particolare attenzione nel predisporre adeguati programmi di selezione anche in funzione dell'ottenimento di prodotti che rispettino specifici requisiti di qualità.

Questo ha riguardato soprattutto le razze migliorate, iscritte ai Libri Genealogici, che sono sottoposte a severi programmi selettivi. Per il suino Nero Siciliano, non soggetto a piani di selezione, la priorità ha riguardato la necessità di incrementare la consistenza della popolazione, contenendo al contempo il tasso di consanguineità, per salvaguardarne la variabilità genetica. La tutela della razza, tuttavia, risulta strettamente connessa alla valorizzazione delle sue produzioni, quindi al miglioramento della qualità delle stesse.

La qualità della carne è un carattere complesso, difficile da definire in modo semplice ed univoco, influenzato da numerosi fattori tra cui quelli ambientali, le condizioni pre e post-mortem e la genetica dell'animale. Nonostante l'ereditabilità dei caratteri che influenzano la qualità della carne sia bassa, la conoscenza del genoma rappresenta uno strumento



importante per introdurre rilevanti innovazioni nel settore suinicolo, con vantaggi sul piano dell'efficienza della selezione.

In questo studio, abbiamo identificato 7293 varianti non condivise tra i genomi oggetto di studio ed esclusive di Nero Siciliano, in un sottogruppo di geni correlati al grasso. Gli SNPs non condivisi e fissati nella razza, potrebbero essere il punto di partenza per l'identificazione di marcatori molecolari razza-specifici utili per la differenziazione della stessa e l'autenticazione dei relativi prodotti. La possibilità di sviluppare metodiche molecolari per la tracciabilità e l'autenticazione di prodotti razza-specifici, rappresenta uno strumento chiave per la loro valorizzazione ed una strategia per la conservazione sostenibile delle risorse genetiche animali locali (Fontanesi *et al.*, 2016). Allo stesso tempo, l'identificazione di varianti alleliche associate a tratti favorevoli in suino Nero Siciliano permetterebbe di impostare piani di accoppiamento che tengano conto anche di aspetti correlati alla qualità delle produzioni, in modo da orientare la scelta dei riproduttori anche in funzione dell'ottenimento di prodotti che rispettino specifici requisiti. Numerosi Autori concordano nell'affermare che i pool genici coinvolti nell'accumulo di tessuto adiposo differiscono tra le razze in funzione del loro background (Szabo *et al.*, 2010; Lebret *et al.*, 2014; Nakajima *et al.*, 2011), l'identificazione di associazioni tra polimorfismi e fenotipi di interesse in suino Nero Siciliano rappresenterebbe dunque uno *step* importante nel processo di caratterizzazione della razza.

Il risultati preliminari di questo studio comparativo *in silico* sono stati presentati al XII° Convegno Nazionale sulla Biodiversità, Università degli Studi di Teramo, nei giorni 13-14 e 15 Giugno 2018.

## **2. Parte Terza**

### **3.1. SNPs discovery in RRLs da pool di DNA di soggetti aventi fenotipi estremi e divergenti per il carattere BFT**

#### **3.2. Abstract**

##### **Background**

Fat deposition is a key biological process that has implications in pig economic management as it affects carcass quality and aptitude to production of cured ham. The association of single nucleotide polymorphisms (SNPs) with carcass traits in pig has been confirmed in several studies but the association depended from genetic backgrounds of different breeds. Here we present a study carried out in order to detect single nucleotide polymorphisms (SNPs) that could be associated with back fat thickness (BFT) in Nero Siciliano pigs. Genomic DNA from two groups of Nero Siciliano pigs with divergent phenotype for BFT were pooled and digested with BsuRI (HaeIII) restriction enzyme for preparation of reduced representation libraries (RRLs). Then, sequencing of the two libraries was carried out on Personal Genome Machine (Life technologies).

##### **Results**

The two RRLs produced 4124595 (BFT+) and 4052107 (BFT-) sequenced reads, which after cleaning (filtering and trimming) were mapped on *Sus scrofa* reference genome (Sscrofa 11.1 assembly). Only reads with mapping quality score > 20 were retained for subsequent analysis. SNP calling was performed using SNAPE, a software that implement a Bayesian approach for SNP calling in pooled samples. 47,791 putative SNPs were called by SNAPE, of these 32,235 (67.4%) were polymorphic while 15,556 (32.5%) were monomorphic. Sanger sequencing was carried out to confirm NGS data on a few regions showed alternative read count between the two libraries, in individual pigs included in each pool. Fischer's exact test was performed in order to evaluate differences in allele frequency estimates as determined by alternative read count between the two libraries. Of all SNPs detected in this study, 22 showed enriched alleles in one or in the other RRLs. These SNPs, some of these localised in genes involved in fat metabolism, might be potential markers associated with BFT in Nero Siciliano pig.

## **Conclusions**

Fatness-related traits, in particular the backfat thickness, are very important in pig production since they influence meat quality and technological processes adopted for long matured products such as cured ham. For valorisation of Nero Siciliano's productions, a request to label the fresh meat with the Protected Denomination of Origin (PDO) was issued in 2005 while for Nero Siciliano's cured ham the request for the PDO has been started in 2011. In this study, we identified SNPs potentially associated with BFT that might be utilized for applications in breeding programs. Attitude to high fat deposition (in particular in neck, withers and back) for the Nero Siciliano pig is known and our results could contribute to explain the biology of fat metabolism in this breed.

## **Keywords**

Personal Genome Machine, backfat thickness, RRLs, pool, selective genotyping approach, SNPs

### 3.3. Introduzione

La deposizione di grasso è un processo biologico di fondamentale importanza per i noti risvolti sulle caratteristiche qualitative delle carcasse. Le peculiarità del sistema produttivo italiano, orientato essenzialmente alla produzione del suino pesante, hanno indirizzato la selezione ANAS nell'ultimo ventennio, plasmando razze con caratteristiche nettamente distinte da quelle estere e che riflettono le specifiche dei disciplinari delle produzioni DOP ([www.ANAS.it](http://www.ANAS.it)).

Un'adeguata ed uniforme copertura di grasso della carcassa è, infatti, necessaria al fine di consentire la più idonea concentrazione di sale nel muscolo nei prosciutti stagionati DOP e ridurre il più possibile le perdite di stagionatura. Inoltre, l'elevata correlazione negativa tra spessore del grasso dorsale e contenuto di acidi grassi insaturi fa sì che un'eccessiva riduzione dello spessore del lardo determini un aumento dei grassi insaturi che possono causare problemi di irrancidimento durante la stagionatura. È dunque uno dei principali obiettivi del miglioramento genetico nel nostro Paese il mantenimento dell'idonea copertura del grasso, misurata come spessore del grasso dorsale.

I tipi genetici utilizzati per la produzione di suini pesanti in Italia derivano dalle tre razze attualmente riconosciute dai disciplinari di produzione dei prosciutti DOP (Parma e San Daniele) ovvero la razza Landrace italiano, Large White e Duroc, mentre quelli utilizzati per la produzione di suini leggeri da macelleria, sono rappresentati dalle razze Pietrain e Landrace belga.

Come riportato precedentemente, per il suino Nero Siciliano è in corso il riconoscimento della DOP per il prosciutto crudo di suino Nero dei Nebrodi, ai sensi del Reg. (UE) n. 1151/2012 e D.M. 14 ottobre 2013.

Per questa razza è nota l'attitudine all'accumulo del grasso nella regione del collo, schiena e garrese (Chiofalo L., 2007) ed in letteratura è documentata la notevole capacità adipogenetica delle razze non migliorate (Serra *et al.*, 1998; Labroue *et al.*, 2000; Fortina *et al.*, 2005; Maiorano *et al.*, 2009). La loro gestione deve dunque far fronte ai problemi di efficienza produttiva connessi ai processi metabolici legati alla conversione alimentare, alla deposizione del grasso e all'accrescimento lento. I caratteri produttivi e le caratteristiche della carcassa sono caratteri complessi, ma è noto come questi siano in parte geneticamente determinati. Nonostante l'ereditabilità di tali caratteri sia bassa, la conoscenza del genoma rappresenta uno strumento importante per introdurre rilevanti

innovazioni nella gestione degli allevamenti e, nel caso di razze non ancora ben caratterizzate come quella del suino Nero Siciliano, per la quale nessun piano di selezione e miglioramento genetico è operato dall' ANAS, per meglio conoscerne la struttura genetica.

Gli SNP (*Single Nucleotide Polymorphism*) rappresentano la base funzionale primaria della variabilità genetica che è riflessa nelle differenze fenotipiche tra ed entro razze. Questi marcatori possono essere applicati in modo efficiente per identificare le varianti genetiche più favorevoli per il miglioramento dei caratteri quantitativi di interesse zootecnico o ai fini dell'identificazione e tracciabilità dei prodotti, qualora i polimorfismi identificati rappresentino delle impronte selettive esclusive della razza allo studio. La caratterizzazione genetica è dunque un prerequisito fondamentale per la gestione delle risorse genetiche e può essere sfruttata per la definizione di protocolli di autenticazione molecolare o per l'implementazione di piani di selezione orientati alla qualità delle produzioni.

Nel presente lavoro di tesi è stato preso in considerazione il carattere *spessore del grasso dorsale*, data la grande importanza che riveste la deposizione di grasso ai fini della produzione di prosciutti e per i noti effetti tecnologici in fase di stagionatura (Bosi & Russo, 2004; ANAS, 2016), con lo scopo di identificare SNPs che potrebbero essere associati a questo fenotipo nel suino Nero Siciliano. Una valutazione precoce, e con metodi innovativi, della attitudine alla trasformazione della carne in prodotti stagionati potrebbe essere utile per migliorare le produzioni di suino Nero, anche nell'ottica del possibile ottenimento di un riconoscimento a Denominazione d'Origine.

L'avvento delle tecniche di *Next Generation Sequencing* (NGS) ha rivoluzionato gli approcci genomici e trascrittomici alla biologia. Questi nuovi strumenti di sequenziamento hanno anche avuto impatto sulla identificazione e validazione di marcatori genetici nelle popolazioni. Nonostante il processo di genotyping si sia notevolmente velocizzato grazie all'avvento di chip di marcatori SNPs ad alta densità, c'è qualche preoccupazione riguardante la possibile distorsione insita nell'utilizzo stesso del chip (Albrechtsen *et al.*, 2010) soprattutto per le popolazioni locali, selvatiche o molto divergenti dal momento che, per la specie suina, quest'ultime non sono state incluse nella costruzione del Porcine SNP60 Beadchip (Ramos *et al.*, 2009). La genotipizzazione di nuove popolazioni sarebbe dunque influenzata dagli alleli presenti nel saggio originale (Herrero-Medrano *et al.*, 2014). Ad ogni modo le nuove tecniche in rapida e continua evoluzione, a costi sempre meno proibitivi, soppianteranno i metodi fin qui utilizzati a vantaggio di progetti di whole-genome resequencing (Huang *et al.*, 2009). Sequenziare l'intero genoma di molti individui

in una popolazione è però ad oggi costoso e spesso non necessario, dal momento che la risposta a molte domande biologiche può derivare dall'identificazione di polimorfismi derivanti da screening in un sottoinsieme di regioni genomiche definite da enzimi di restrizione, che evitano il sequenziamento di regioni altamente ripetute (Altshuler *et al.*, 2000; Davey *et al.*, 2010). Tra i diversi metodi di genome-wide marker discovery e genotyping che impiegano la digestione del genoma target con enzimi di restrizione al fine di ridurre la complessità, sono annoverati il sequenziamento a rappresentazione ridotta che prevede la costruzione di reduced-representation libraries (RRLs) (Altshuler *et al.*, 2000), complexity reduction of polymorphic sequences (CRoPS) (van Orsouw *et al.*, 2007), restriction-site-associated DNA sequencing (RAD-seq) (Baird *et al.*, 2008; Hohenlohe *et al.*, 2010; Etter *et al.*, 2011) e genotipizzazione a bassa copertura (low coverage genotyping) che comprende multiplexed shotgun genotyping MSG (Xie *et al.*, 2010; Andolfatto *et al.*, 2011) e genotyping by sequencing (GBS) (Elshire *et al.*, 2011).

Tutti i metodi sopracitati implicano i seguenti passaggi chiave: la digestione di più campioni di DNA genomico (da individui o popolazioni) con uno o più enzimi di restrizione; una selezione dei frammenti di restrizione risultanti; sequenziamento con piattaforme NGS del pool finale di frammenti di dimensioni compatibili con la tecnologia di sequenziamento adottata (Davey *et al.*, 2010).

Allo scopo della presente tesi è stato adottato il metodo delle RRLs, sequenziate poi su piattaforma PGM (Thermo Fisher Scientific). Si rimanda alla letteratura di riferimento per gli approfondimenti relativi alle altre tecniche citate (**RADSeq**: Miller *et al.*, 2007; Baird *et al.*, 2008; Hohenlohe *et al.*, 2010; Emerson *et al.*, 2010; Davey *et al.*, 2010; Rowe *et al.*, 2011; Hohenlohe *et al.*, 2011; Chutimanitsakun *et al.*, 2011; Pfender *et al.*, 2011; Baxter *et al.*, 2011; **MSG**: Xie *et al.*, 2010; Andolfatto *et al.*, 2011; **GBS**: Elshire *et al.*, 2011; Sonah *et al.*, 2013).

Le RRLs sono state utilizzate per la prima volta per generare una mappa SNPs del genoma umano (Altshuler *et al.*, 2000) e rappresenta una strategia già mostratasi efficiente per l'identificazione di SNPs in diverse specie (Van Tassell *et al.*, 2008; Wiedmann *et al.*, 2008; Ramos *et al.*, 2009; Amaral *et al.*, 2009; Kerstens *et al.*, 2009; Gore *et al.*, 2009; Sánchez *et al.*, 2009; Amaral *et al.*, 2011; Esteve-Codina *et al.*, 2011; Aslam *et al.*, 2012; Bertolini *et al.*, 2014; Bovo *et al.*, 2015). Il DNA genomico proveniente da più individui, viene digerito con un enzima di restrizione con elevata frequenza di taglio e raggruppato in pool. I frammenti di restrizione risultanti sono selezionati in funzione della size e quindi sequenziati, producendo una copertura parziale, ma a livello di genome-wide, ad una

frazione del costo del sequenziamento dell'intero genoma (whole-genome). Quando è disponibile un genoma di riferimento di alta qualità, come nel caso di *Sus scrofa*, le *reads* del sequenziamento a rappresentazione ridotta possono essere mappate al genoma di riferimento e gli SNPs possono essere “chiamati” come per i progetti di whole-genome resequencing (Li *et al.*, 2009b; Nielsen *et al.*, 2011). Un aspetto fondamentale nella costruzione di RRLs è rappresentato dalla selezione dell'enzima da impiegare per la digestione del genoma target, con l'obiettivo di ridurre al minimo il contenuto di elementi ripetuti (LINEs, SINEs, LTR), e che deve essere appropriato per le caratteristiche della specie allo studio (dimensione genoma, suo contenuto in GC, ploidia, grado di metilazione del genoma, elementi ripetuti).

Per la costruzione delle libraries si è inoltre scelto di raggruppare in due soli pool il DNA di più soggetti: solo gli individui con fenotipi estremi sono stati genotipizzati per lo studio di associazione, in accordo con la strategia di selective genotyping. Diversi Autori hanno dimostrato che questo approccio rappresenta un metodo potente per rilevare l'associazione per un tratto quantitativo (Van Gestel *et al.*, 2000; Zhang *et al.*, 2006; Paternoster *et al.*, 2011) ed è stato già ampiamente utilizzato in studi di associazione genome wide per BFT in suini pesanti italiani (Fontanesi *et al.*, 2012b; Bovo *et al.*, 2015; Fontanesi *et al.*, 2015; Fontanesi *et al.*, 2012d).

L'approccio di genotipizzazione selettiva, congiunta alla strategia di DNA-Pooling, può dunque rappresentare una soluzione per la riduzione dei costi che deriverebbero invece dalla genotipizzazione individuale di tutti i campioni. Gli studi di associazione presentano infatti l'inconveniente di avere dei costi molto elevati che li rende pertanto poco accessibili. Con tale approccio invece, identiche aliquote di DNA appartenente a diversi individui vengono riunite a costituire un unico campione rappresentativo del gruppo in esame, che viene impiegato per l'analisi delle varianti genetiche.

Presupposto imprescindibile è rappresentato dal fatto che i DNA usati per la costruzione dei pool debbano essere accuratamente quantizzati singolarmente mediante analisi fluorimetrica e combinati alle identiche concentrazioni molari, in questo modo la frequenza delle *reads* sarà funzione delle frequenze alleliche nel pool.

Nel presente lavoro di tesi è stato adottato un disegno sperimentale in due stadi, in cui i marcatori che mostravano un'associazione positiva nello studio preliminare di pooling sono stati impiegati per la genotipizzazione individuale nella popolazione originale, in modo da validare la chiamata delle varianti e confermare l'associazione. Questo ha rappresentato il

migliore compromesso tra il risparmio sui costi del pooling e la potenza statistica e la completezza di informazioni fornite dalla genotipizzazione individuale.

## **3.4. Materiali e Metodi**

### **3.4.1. Campionamenti e rilievi fenotipici**

192 soggetti appartenenti alla razza di suino Nero Siciliano rispettanti lo standard richiesto dall'ANAS per l'iscrizione al relativo registro, sono stati campionati all'atto della macellazione. Gli animali, selezionati tra soggetti non parenti almeno in seconda generazione, sono stati macellati al raggiungimento del peso vivo medio di  $92 \pm 5$  kg, le misurazioni ed i rilievi fenotipici sono stati effettuati entro le 3 h *post mortem*.

Nella rilevazione dei dati è stata utilizzata una scheda contenente tutte le indicazioni relative l'animale e l'azienda di provenienza. Sono stati riportati i dati dell'azienda, marca auricolare del soggetto, età dell'animale alla macellazione, peso vivo alla macellazione, sesso, peso della carcassa, peso mezzena sinistra, spessore del grasso alla sommità del muscolo *gluteus medius*, spessore del grasso a livello di 1<sup>a</sup> e ultima vertebra toracica. Dalle misurazioni sono state ricavate le medie del fenotipo “back fat thickness” (BFT) poi impiegate per la costituzione dei gruppi estremi e divergenti per il carattere in oggetto, in accordo con il metodo di Selective Genotyping. Brevemente, dal totale dei soggetti campionati, sono stati selezionati 50 soggetti appartenenti alla coda positiva della distribuzione dei fenotipi per il carattere in oggetto e 50 soggetti aventi valori estremi per la coda negativa della distribuzione, avendo cura di rendere omogenei i gruppi in funzione della tipologia di alimentazione (le aziende erano condotte *en plein air* con somministrazione di mangime commerciale per ingrasso ad integrazione, materiale supplementare 5.3) e di bilanciarli in termini di rapporto tra i sessi. Il DNA estratto da tali soggetti è stato poi riunito in due pool a concentrazione equimolare ed utilizzato per la costruzione delle libraries. La media e la deviazione standard dei fenotipi dei soggetti nella coda negativa e positiva della distribuzione erano  $25,70 \pm 2,97$  mm e  $34,40 \pm 4,65$  mm, rispettivamente.

Le misurazioni sulla carcassa sono state effettuate manualmente con l'ausilio di un calibro certificato (INECO). Tale metodo di misurazione, basato sulla misurazione manuale dello spessore in millimetri del lardo di copertura e dello spessore del muscolo rilevati sulla mezzena sinistra, è stato anche autorizzato, con Decisione n. 38 del 24 gennaio 2014, pubblicata sulla Gazzetta Ufficiale dell'Unione Europea n. L23 del 28 gennaio 2014 (la cui



rettifica è stata pubblicata sulla GUE n. L54 del 22 febbraio 2014), per la classificazione ufficiale delle carcasse ai fini della stima della percentuale di tagli magri delle stesse.

I campionamenti sono stati eseguiti presso lo stabilimento di macellazione- Rocca di Capri Leone (ME), nei mesi di gennaio e febbraio 2017.

Gli animali provenivano da 6 aziende di 3 comuni della provincia di Messina (Sant'Angelo di Brolo, Mirto, Sinagra), ricadenti nell'area geografica designata per la DOP.

I campioni di sangue, prelevati in Vacutainer contenenti K<sub>3</sub>EDTA come anticoagulante, sono stati mantenuti a 4° C, sino all'arrivo in laboratorio.

#### **3.4.2. Estrazione del DNA, check di qualità, preparazione dei pool equimolari**

I campioni di sangue, collezionati in K<sub>3</sub>EDTA, sono stati centrifugati a 3000 rpm a 4°C per 20 minuti in modo da isolare la frazione leucocitaria, impiegata per l'estrazione del DNA genomico (gDNA) mediante il Wizard® Genomic DNA Purification Kit (Promega Corporation, Italia), attenendosi al protocollo fornito dalla ditta produttrice.

La valutazione della concentrazione del DNA è stata effettuata per via spettrofotometrica (NanoPhotometer® P330 – Implen), a partire da 1 µl di campione, misurando l'Assorbanza a 260 nm. Il rapporto fra la lettura a 260 nm e quella a 280 nm (D.O. 260/D.O. 280) ha fornito una stima della purezza dell'acido nucleico. Una ulteriore lettura a 230 nm ha fornito indicazioni circa la presenza di contaminanti. La quantifica del DNA è stata effettuata anche mediante saggio fluorimetrico (Qubit 2.0 Fluorometer, Thermo Fisher Scientific), con l'ausilio del Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific) come raccomandato dai protocolli Thermo Fisher Scientific per la costruzione delle libraries.

Tutti i campioni che hanno superato entrambi i check di qualità, sono stati assegnati al rispettivo gruppo sulla base delle misurazioni fenotipiche (valori estremi e divergenti per il carattere BFT) e riuniti a concentrazioni equimolari per la costruzione dei due pool (pool 1: BFT+; pool 2: BFT-), la cui concentrazione attesa è stata poi verificata in fluorimetria (Qubit 2.0 Fluorometer, Thermo Fisher Scientific). Tutte le fasi inerenti la costituzione dei pool sono avvenute in DNA LoBind® Tubes (Eppendorf) al fine di massimizzare il recupero del DNA, riducendo in modo significativo il legame dei campioni con la superficie.

I pool così ottenuti sono stati impiegati per la costruzione delle libraries.

### **3.4.3. Preparazione delle RRLs (Reduced Representation Libraries) e delle fragment libraries**

10 µg di DNA proveniente da ciascuno dei due pool è stato digerito overnight con 50 U di *BsuRI* (HaeIII) ed i prodotti di digestione visualizzati su gel di agarosio (1% in TBE). Tale enzima, 4 base-cutter che riconosce il sito di restrizione GG<sup>A</sup>CC, è stato scelto in quanto non produceva pattern visibili attribuibili ad elementi ripetuti nell'intervallo di 500-700 bp, come già riportato in precedenti studi (Wiedmann *et al.*, 2008; Bertolini *et al.*, 2014; Bovo *et al.*, 2015). Le bande di DNA nel range di 500-700 bp sono state asportate dal gel e purificate con il kit Wizard® SV Gel and PCR Clean-Up System (Promega) seguendo le istruzioni del produttore.

Le LRRs così ottenute dai 2 pool di DNA enzimaticamente digeriti, costruite a partire dai soggetti con fenotipi estremi e divergenti per il carattere BFT, sono state impiegate per la preparazione delle fragment libraries secondo il protocollo Ion Xpress™ Plus gDNA Fragment Library Preparation (Thermo Fisher Scientific), in allegato alla presente tesi (materiali supplementari 5.4). Questo *step* si è reso necessario per ottenere dei frammenti di dimensione compatibile con la chimica di sequenza della tecnologia Ion torrent (200 bp).

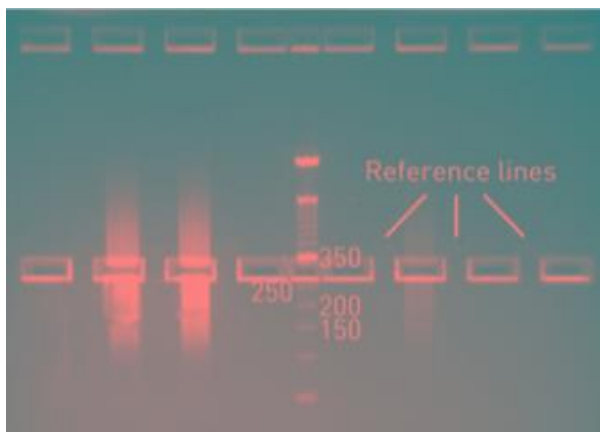
Brevemente, per ciascuna library, 200 ng di DNA purificato da gel sono stati enzimaticamente frammentati mediante Ion Shear™ Plus 10X Reaction Buffer e Ion Shear™ Plus Enzyme Mix II (Thermo Fisher Scientific) per un periodo di incubazione di 5 minuti a 37° C e sottoposti ad uno *step* di purificazione con Agencourt® AMPure® XP Reagent (1.8X del volume del campione) (Beckman Coulter). È seguita la ligazione degli adattatori, il nick-repair per rendere la ligazione covalente su entrambi gli strands (Ion Xpress™ Plus Fragment Library Kit, Thermo Fisher Scientific) ed un ulteriore *step* di purificazione del DNA ligato.

### **3.4.4. Size selection, quantifica delle libraries, preparazione template e sequenziamento**

Alla purificazione è seguita una fase di size selection tramite E-Gel® SizeSelect™ Agarose Gel (Thermo Fisher Scientific) ed i frammenti corrispondenti a 200 bp di inserto sono stati collezionati per il successivo processamento (figura 4). Le librerie così ottenute sono state amplificate su un 2720 Thermal Cycler (Applied Biosystems™) secondo quanto riportato in tabella 9 e 10, al fine di ottenere materiale sufficiente per una accurata determinazione. All'amplificazione è seguito un secondo passaggio di purificazione in due

*step*, mediante aggiunta dell'Agencourt® AMPure® XP Reagent a 0.5X e 1.2X volumi rispettivamente nei due *step* consecutivi.

La quantifica è avvenuta in fluorimetria (Qubit 2.0 Fluorometer, Thermo Fisher Scientific), in qPCR (StepOne Plus™, Applied Biosystems) e con l'ausilio del Bioanalyzer 2100 (Agilent). Per ulteriori dettagli circa la quantifica delle libraries, si veda in appendice i materiali supplementari 5.5.



**Fig. 4- Size selection tramite E-Gel® SizeSelect™ Agarose Gel (Thermo Fisher Scientific)**

<b>Component</b>	<b>Volume</b>
Platinum® PCR SuperMix High Fidelity	200 µL
Library Amplification Primer Mix	10 µL
Unamplified library	50 µL
<b>Total</b>	<b>260 µL</b>

**Tab. 9- Reagenti di amplificazione delle fragment libraries e relativo volume**

<b>Stage</b>	<b>Step</b>	<b>Temperature</b>	<b>Time</b>
Holding	Denature	95° C	5 min
	Denature	95° C	15 sec
Cycling (8)	Anneal	58° C	15 sec
	Extend	70° C	1 min
	-	4° C	Hold

**Tab. 10- Protocollo termico per la reazione di amplificazione delle fragment libraries**

Le librerie sono poi state normalizzate alla concentrazione di 100 pM e ulteriormente diluite a 26 pM, come da protocollo, clonalmente amplificate in emPCR ed arricchite su

Ion OneTouch™ 2 System (Thermo Fisher Scientific) con l'ausilio del kit Ion PGM™ Hi-Q™ OT2 (Thermo Fisher Scientific). Ciascuna libreria è stata caricata su un Ion 318™ Chip V2 BC, previa valutazione dell'efficienza di templaggio con il kit Ion Sphere™ Quality Control e il Qubit® 2.0 Fluorometer, e sequenziata su Ion PGM™ System con Ion PGM™ Hi-Q™ Sequencing Kit (parametro di corsa 500 *flow*).

Per ulteriori dettagli circa la preparazione e arricchimento delle template-positive Ion PGM™ Hi-Q™ Ion Sphere™ e la valutazione dell'efficienza di templaggio, si veda in appendice il materiale supplementare 5.6 e 5.7.

Solo alcuni degli SNPs che mostravano differenze significative nelle due RRLs e che erano localizzati in geni annotati sono stati infine analizzati, mediante sequenziamento Sanger, nella popolazione originaria (intero campionamento), compresi gli individui utilizzati per la costruzione dei pool in modo da confermarne l'eventuale associazione con il carattere di interesse. La media e la deviazione standard dei fenotipi dell'intero campione erano 31,05 ± 5,60 mm. Per ulteriori dettagli circa il disegno dei primer, i protocolli di sequenziamento e l'allestimento delle reazioni si veda in appendice il materiale supplementare 5.8.

### **3.4.5. Analisi dei dati di sequenza ed identificazione SNPs**

Le *reads* ottenute dal sequenziamento delle due RRLs sono state ispezionate con FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) e processate utilizzando il software Ion Torrent Suite v3.6 (Life Technologies). Sono state filtrate preliminarmente le sequenze policlonali, quelle di bassa qualità ( $Q < 20$ ) ed è stato effettuato il trimming degli adattatori. Script Bash sono stati usati per filtrare solo le reads processate di lunghezza  $\geq 30$  bp. Le reads di buona qualità sono state infine mappate sul genoma di riferimento (Sscrofa 11.1; GCA\_000003025.6) con TMAP 3.6 ALIGNER (<https://github.com/iontorrent/TS/tree/master/Analysis/TMAP>) che utilizza tre differenti algoritmi di allineamento: BWA-short (Li & Durbin 2009), BWA-long (Li & Durbin 2010) e SSAHA (Ning *et al.*, 2001). Il file bam generato da TMAP è stato processato per la rimozione dei duplicati eseguendo OTG –snpcaller (Zhu *et al.*, 2014), pipeline ottimizzata per dati Ion Torrent, che implementa il metodo di calcolo RDAST (Remove Duplicates According to Alignment Score Tag) e quello di filtering denominato Alignment Optimize Structure (AOS) per ovviare alle criticità nel rilevamento di regioni omopolimeriche note per questa piattaforma (Bragg *et al.*, 2013; Quail *et al.*, 2012) e migliorare dunque

l'accuratezza, la sensibilità e la specificità del rilevamento di SNPs nella successiva fase. Opzioni usate AS score 160, AOS gap ratio 0.2.

Dopo questa fase di pre-processamento, il file bam è stato convertito in formato PILEUP con Samtools (Li *et al.*, 2009a) ed utilizzato per il variant calling con SNAPE (code.google.com/p/snape-pooled/), opzioni: divergence 0.01, prior nucleotide diversity 0.001, folded spectrum; settando la *probabilità a posteriori* di segregazione  $> 0.90$  (Raineri *et al.*, 2012), considerando solo i siti polimorfi con coverage minimo di 3x. SNAPE utilizza l'approccio bayesiano per il calcolo della minor allele frequency (MAF) e lo SNP calling in dati da pool. Poiché la tecnologia è soggetta a errori in caso di regioni omopolimeriche (Bragg *et al.*, 2013) le INDEL non sono state considerate in questo studio. Il tool VEP (Variant Effect Predictor) di Ensembl (McLaren *et al.*, 2016) è stato utilizzato per interrogare l'Ensembl Variation database (Ensembl release 92, Aprile 2018) e verificare quali degli SNPs identificati da SNAPE non fossero già riportati in dbSNP (Version 150). Il tool VEP è stato impiegato anche per mappare ciascuna mutazione e predirne gli effetti. Tutte le mutazioni *missense* identificate da VEP sono poi state analizzate con SIFT (Kumar *et al.*, 2009) per valutare se queste potessero avere conseguenze deleterie sulle proteine tradotte.

Al fine di valutare le differenze nelle frequenze alleliche derivate dalla conta del numero di reads alternative tra le due RRLs, il test esatto di Fisher è stato calcolato per ogni posizione genomica alternativa, coperta da una profondità minima di 3x. Tutte le posizioni con  $P_{Fisher} \leq 0.05$  sono state anche ispezionate visivamente con il software IGV (Integrative Genomics Viewer) (Thorvaldsdóttir *et al.*, 2013). L'IGV fornisce un'immagine visiva dell'allineamento di sequenza con il riferimento, facilitando così il rilevamento di false varianti come risultato di possibili errori di sequenziamento o di strand bias.

I dati della genotipizzazione individuale degli SNPs le cui frequenze alleliche differivano significativamente tra le due RRLs (*loci MC4R, FASN, FFAR4, SCD, LEP, LEPR*) sono stati impiegati per testarne la possibile associazione con il carattere di interesse. L'equilibrio di Hardy–Weinberg (HWE) è stato testato per ciascun *locus* utilizzando il test esatto di Guo & Thompson (1992), implementato nel software GENEPOP 1.2 (Raymond & Rousset, 1995). L'analisi di associazione, eseguita utilizzando la procedura MIXED del pacchetto statistico SAS<sup>®</sup> (v 9.1; SAS<sup>®</sup> Inst.Inc., Cary, NC), ha incluso nel modello gli effetti fissi del genotipo ad ogni *locus* considerato e del sesso; mentre tra gli effetti random l'età e il peso alla macellazione. Sono state considerate significative le associazioni con  $p \leq 0,05$ .

### 3.5. Risultati

#### 3.5.1. Sequenziamento ed identificazione SNPs da pool

Un totale di 4124595 (BFT+) e 4052107 (BFT-) reads sono state rispettivamente ottenute dal sequenziamento delle due RRLs, costruite a partire dai pool di DNA di soggetti aventi fenotipi estremi per il carattere BFT (Tabella 11). Previo filtering e trimming, 2592637 (BFT+) e 2411235 (BFT-) reads di buona qualità ( $Q_m \geq 20$ ) sono state mappate sul genoma di riferimento di *Sus scrofa* (GenBank: GCA\_000003025.6). Dopo rimozione dei duplicati, 47791 SNPs putativi sono stati identificati da SNAPE. Di questi, 32235 (67,4%) sono risultati polimorfi mentre 15556 (32,5%) erano monomorfi, con un rapporto Transizioni/Trasversioni risultato pari a 2,240.

L'analisi condotta con VEP ha mostrato che la maggior parte delle varianti identificate (~74%) era già presente in dbSNPs (release 151), mentre ~ 26% erano *novel*.

Metrics	BFT+	BFT-
Sequenced Reads	4124595	4052107
Surviving after pre-processing	3924742	3859965
Removed duplicates	774187	658741
Mapped reads ( $Q_m \geq 20$ )	2592637	2411235
SNPs called by SNAPE ( $Q_m \geq 20$ ; $RD \geq 3$ )	23784	24007
Sequenced bases retained by SNAPE in common between the two RRLs	452288	

**Tab. 11-** Principali metriche relative al sequenziamento e SNP calling delle due RRLs di DNA pool, costruiti a partire da soggetti aventi fenotipi estremi per il carattere BFT.  $Q_m$ = mapping quality;  $RD$ = read depth.

Come atteso, la maggior parte delle varianti identificate erano localizzate in regioni non codificanti del genoma, introni e regioni intergeniche (tabella 12). In appendice alla presente sezione, in tabella 13, sono riportati gli SNPs identificati nelle sole regioni trascritte. Tra quelli localizzati nelle regioni codificanti del genoma, 96 erano mutazioni sinonime, 277 missenso, 10 stop-gained (es. nei geni *ADH6*, *PGM2*, *TMA16*, *MRPL1*), 9 stop lost (es. in *WDR25*, *HECTD1*) e 1 stop\_retained (in ENSSSCG00000036380). Tra le mutazioni missenso predette, 38 sono state classificate come deleterie da SIFT. Tali mutazioni sono state annotate in geni che rivestono un ruolo in processi correlati al metabolismo lipidico (*MIGAI*, *CUBN*) ed in svariati meccanismi cellulari (attivazione risposta immunitaria innata: *TLR10*, migrazione neuronale: *SLIT2*, segnale apoptotico: *DTHD1*).

Gene position or SNP effect	Count
SPLICE_SITE_REGION	16
UPSTREAM	1271
UTR_5_PRIME	7
SPLICE_SITE_DONOR	7
INTRON	17649
SPLICE_SITE_ACCEPTOR	15
UTR_3_PRIME	195
DOWNSTREAM	1241
INTERGENIC	26997
MISSENSE VARIANT	277
STOP GAINED	10
STOP LOST	9
STOP_RETAINED_VARIANT	1
SYNONYMOUS VARIANT	96

Tab. 12 – Annotazione condotta da VEP degli SNPs identificati da SNAPE.

### 3.5.2. Differenze tra le due RRLs costruite da DNA pool di soggetti aventi fenotipi estremi per il carattere BFT

SNAPE è stato utilizzato per stimare, dalla conta delle *reads* alternative, la frequenza allelica di ogni posizione genomica polimorfica, tra le 452288 in comune tra le due RRLs. Dopo correzione di Bonferroni, 22 posizioni genomiche mostravano  $P_{\text{Fisher}} \leq 0,05$  confrontando le *reads* alternative osservate nelle due RRLs costruite da DNA di soggetti aventi fenotipi estremi e divergenti per il carattere BFT (tabella 14), nessuno dei quali con  $P_{\text{Fisher}} \leq 0,01$ . Gli SNPs identificati erano localizzati in cromosomi autosomici, nei quali sono stati mappati geni candidati e QTLs (Quantitative Trait Loci) per BFT (chr 1,2,6,7,13,14).

Delle 22 varianti identificate (10 delle quali già note in dbSNP), ~31% erano varianti intergeniche, ~31 % erano localizzate nelle regioni introniche, il 4 % erano varianti *downstream* e il 3% *upstream*. Quest'ultime (*downstream* e *upstream*) erano localizzate in geni annotati con funzioni note: Melanocortin-4 receptor (*MC4R*), Fatty acid synthase (*FASN*), Phosphodiesterase 6C (*FFAR4*), Stearoyl-CoA desaturase (*SCD*) mentre quelle introniche nei geni Leptin (*LEP*) e Leptin Receptor (*LEPR*).

Chr	Genomic position	Reference	Alt	BFT+		BFT-		$P_{Fischer}$
				N_Ref	N_Alt	N_Ref	N_Alt	
1	4553	A	GA	0	6	8	2	2,86E-02
1	1082	G	TG	0	7	7	1	4,82E-02
1	284340	T	TA	29	4	17	7	4,05E-02
1	124584196	C	CG	17	2	6	5	1,28E-02
1	160774173	A	AG	8	1	0	3	3,37E-02
2	123750235	G	GA	13	0	1	4	1,70E-02
2	123751016	T	TC	1	8	13	3	4,65E-02
6	956438	G	GA	9	0	4	6	3,35E-02
6	145607675	C	CA	11	1	3	7	3,55E-02
6	145631685	G	AG	13	1	4	9	1,27E-02
6	145654695	C	CT	15	2	2	4	3,63E-02
6	146893537	A	GA	8	0	4	7	1,87E-02
6	67619980	G	TG	27	13	17	32	1,03E-02
7	75888372	T	TA	11	0	3	7	1,25E-02
12	921860	T	TG	32	3	8	14	2,63E-02
13	31684656	G	GA	29	0	11	2	3,69E-02
14	105011727	C	CT	8	13	10	1	4,37E-02
14	111461097	C	TC	1	15	7	2	1,29E-02
15	155198719	C	CG	0	6	7	1	4,96E-02
15	155253410	C	CA	0	7	8	0	2,95E-02
16	36782695	T	GT	8	43	12	21	3,27E-02
18	20123273	C	CT	5	49	57	13	1,29E-02

**Tab. 14-** Elenco dei 22 SNPs che hanno mostrato differenze significative ( $P \leq 0.05$ ) tra le due RRLs (Positive Library: BFT+, Negative Library: BFT-). Chr = Cromosoma; Pos = posizione nucleotidica sul genoma di riferimento Sscrofa 11.1; Reference = Reference nucleotide su Sscrofa11.1 reference genome; SNP\_NGS = variante identificata dal sequenziamento; N\_Ref = Numero di reads che supportano il reference nucleotide; N\_Alt = Numero di reads alternative al reference nucleotide;  $p_{Fischer}$  = p-value del test di Fisher. La qualità media delle reads che supportano il dato è  $\geq 40$  per ciascuna posizione.

### 3.5.3. Associazione tra marker identificati nelle due RRLs e BFT

Dei 192 soggetti costituenti la popolazione originaria, per i quali disponevamo dei dati fenotipici registrati all'atto della macellazione, solo 180 hanno prodotto elettroferogrammi di buona qualità per tutti i *loci* allo studio.

I dati della genotipizzazione individuale sono stati impiegati per il calcolo delle frequenze alleliche e genotipiche a ciascun *locus*, nessuno dei quali si discostava dall'equilibrio di Hardy-Weinberg. L'analisi di associazione è stata condotta separatamente per ciascun *locus* (un *locus* per run) e l'effetto fisso del sesso, inizialmente incluso nel modello, è stato escluso dall'analisi in quanto risultato non significativo per nessuno dei marker analizzati ( $p=0,57-0,83$ ). Solo i genotipi ai *loci* *LEP* e *FASN* sono risultati significativamente associati al carattere spessore del lardo dorsale. Le frequenze alleliche al locus *LEP* sono risultate 0,496 e 0,504 per allele C e T rispettivamente mentre per *FASN* le frequenze dell'allele G e T sono risultate 0,412 e 0,588. I risultati dell'analisi di associazione per



questi marker sono mostrati in tabella 15. *LEP* è risultato significativamente associato con lo spessore del grasso dorsale ( $p= 0,047- 0,030$ ), con il genotipo TT che sembra contribuire all' incremento dello spessore del carattere.

Lo SNP sul gene *FASN* ha mostrato significativa associazione con il carattere in esame ( $p= 0,032-0,037$ ), con il genotipo GG che sembra invece decrementare il carattere.

Locus	Genotype 1/1	Genotype 1/2	Genotype 2/2	P-value
	LSM (SE)	LSM (SE)	LSM (SE)	
<b>LEP_20123273</b>	<b>CC</b>	<b>CT</b>	<b>TT</b>	
GM	28,173 (0,422)	29,293 (0,331)	33,754 (0,529)	0,047
FT	32,180 (0,160)	33,784 (0,251)	36,832 (0,324)	0,041
LT	23,215 (0,180)	24,751 (0,162)	26,269 (0,479)	0,030
<b>FASN_921860</b>	<b>TT</b>	<b>TG</b>	<b>GG</b>	
GM	33,479 (0,727)	30,513 (1,499)	27,764 (1,431)	0,032
FT	35,601 (0,477)	33,924 (0,076)	31,178 (0,157)	0,049
LT	27,125 (0,268)	25,443 (1,128)	23,961 (1,117)	0,037

**Tab. 15- Analisi di associazione tra SNPs ai loci *LEP* e *FASN* e spessore del lardo dorsale** (Back-fat thickness: misurato al livello del muscolo *Gluteus medius* (GM), misurato al livello della prima vertebra toracica (FT), misurato al livello dell'ultima vertebra toracica (LT); le misurazioni sono espresse in mm). Per ogni classe genotipica a ciascun locus è riportato il valore di *least square means* e relativo *standard error* (tra parentesi).

### 3.6. Discussioni

Come primo passo, per valutare le putative associazioni tra polimorfismi distribuiti nell'intero genoma ed il carattere spessore del grasso dorsale, è stato applicato un approccio di genotipizzazione selettiva che studiava se le frequenze alleliche differivano tra i gruppi divergenti di soggetti, selezionati in base alle misurazioni fenotipiche effettuate per questo carattere. Solo 6 delle 22 posizioni genomiche che mostravano differenze significative tra i due gruppi, erano mappate su geni annotati. Tali SPNs sono stati caratterizzati, nella successiva analisi, nell'intero campionamento originario comprendente anche i soggetti che erano stati inclusi nella costruzione dei pool. Dei 6 marker analizzati solo 2 hanno mostrato significativa associazione con i tratti oggetto di studio, il polimorfismo 20123273 C>T al locus *LEP* e 921860 T>G al locus *FASN*. In letteratura non si dispone di dati riguardanti questi due polimorfismi in altre razze.

Il gene *LEP* (*Leptin*) codifica per una proteina di 146 amminoacidi, espressa principalmente nel tessuto adiposo (D'Andrea *et al.*, 2008). Questo ormone agisce sul recettore ipotalamico modulando il senso di sazietà e quindi regolando il peso corporeo ed

il bilancio energetico. Nei suini, il gene *LEP* è stato mappato sul cromosoma 18q13-q21 (Neuenschwander *et al.*, 1996) ed è composto da tre esoni e due introni. Diverse mutazioni a questo *locus* sono state associate a tratti correlati ad incremento ponderale medio giornaliero, grasso intramuscolare e spessore del grasso sottocutaneo (Chao *et al.*, 2012; Hirose *et al.*, 2014; de Oliveira Peixoto *et al.*, 2006) nelle razze Casertana, Duroc, Large White e Meishan, anche se in letteratura sono presenti risultati controversi che sottolineano, come è noto, come gli effetti delle mutazioni dipendano dal background genetico della razza indagata e da diverse strutture di *linkage disequilibrium* nelle popolazioni studiate (Hirose *et al.*, 2014; Kennes *et al.* 2001; Pérez Montarelo *et al.*, 2012; Chen *et al.* 2004; Amills *et al.* 2008; Switonski *et al.* 2010).

Il polimorfismo identificato in suino Nero Siciliano, risultato associato al fenotipo oggetto del presente lavoro di tesi, non è stato indagato in precedenza in altre razze. Il risultato della genotipizzazione individuale è coerente con quanto emerso dal sequenziamento delle 2 RRLs, in cui le reads che presentavano l'allele T sono state mappate in maniera preponderante in libreria BFT+. Dato il ruolo funzionale del gene *LEP*, gli effetti di tale mutazione su tratti correlati all'adiposità della carcassa potrebbero essere spiegati da una sua influenza nella modulazione dell'ingestione alimentare.

Il gene *FASN* codifica per una proteina avente un ruolo essenziale nella sintesi degli acidi grassi a catena lunga, a partire da acetil CoA e utilizzando malonil-CoA come donatore di carbonio 2 (Menendez *et al.*, 2009; Wakil, 1989). Tale gene è stato mappato sul cromosoma 12 (Muñoz *et al.*, 2003), in una regione che comprende numerosi QTLs per tratti correlati all'adiposità della carcassa (<http://aaa.animalgenome.org/cgi-bin/gbrowse/pig/?name=Chr.12>).

Anche in questo caso il risultato dello screening individuale al locus *FASN*, che sembra suggerire come l'allele G decrementi il carattere, è in accordo con i ritrovamenti nella genotipizzazione selettiva dei pool che ha mostrato una significativa presenza dell'allele T (ref) nella libreria BFT+ e dell'allele G (alt) in BFT-.

Uno dei principali vantaggi degli studi di DNA pooling è rappresentato dal fatto che è ridotta la quantità di genotipizzazione necessaria per stimare le frequenze alleliche in un campione. La maggiore criticità è invece rappresentata dalla possibilità di errori sperimentali, specifici per gli studi di DNA pooling, che possono potenzialmente determinare falsi positivi. Tali errori sperimentali possono derivare sia da una quantità ineguale di DNA apportata dagli individui che costituiscono il pool, sia da imprecisioni nelle reazioni di PCR e nella stima delle frequenze alleliche (Le Hellard *et al.*, 2002;

Barratt *et al.*, 2002; Bader *et al.*, 2001). Per questi motivi abbiamo dato particolare importanza alla quantifica del gDNA e delle librerie in fase di costituzione delle stesse e abbiamo utilizzato SNAPE, pacchetto software che implementa un approccio bayesiano per l'identificazione di SNP, ottimizzato per pool sequenziati. Inoltre il pooling determina necessariamente la perdita di informazioni che potrebbero essere ottenute dalla genotipizzazione individuale sia perché, come discusso sopra, il raggruppamento comporta errori sperimentali che non si verificano nella genotipizzazione individuale, sia perché per tratti quantitativi, il raggruppamento consente di esaminare le differenze tra i pool ma non quelle all'interno del pool.

Per minimizzare la perdita di informazioni e ottenere stessa significatività e potenza di uno studio basato sulla genotipizzazione individuale, è possibile agire, in un progetto a due pool, mediante la scelta ottimale dei criteri di selezione per i due pool (Sham *et al.*, 2002). In assenza di errori sperimentali, il criterio di selezione ottimale per riunire un campione casuale di individui non imparentati, è selezionare individui dalle code estreme della distribuzione di un tratto quantitativo, ciascuno comprendente il 27% del campione (Bader *et al.*, 2001; Jawaid *et al.*, 2002; Bader *et al.*, 2002). Questa frazione di pool è ampiamente indipendente dalla frequenza del marker e dalla modalità di ereditarietà del carattere (Sham *et al.*, 2002) ed è stata rispettata nel presente lavoro di tesi.

Una criticità del presente lavoro ha riguardato la bassa profondità di copertura (3x) che è stata settata come filtro di SNAPE, che può non essere sufficiente per una stima affidabile delle frequenze alleliche da dati NGS.

In una prima fase dell'analisi con SNAPE era stato impostato un filtro per estrarre dai pool solo gli SNPs con  $RD \geq 10$ , ma questo avrebbe comportato un ingente perdita di dati. Abbiamo dunque mantenuto il coverage di 3x dato che la validazione delle putative associazioni sarebbe avvenuta poi mediante genotipizzazione individuale, utilizzando il pooling come metodo per la selezione di un sottoinsieme di marcatori da studiare più nel dettaglio. È stato però dimostrato che impostando una profondità di copertura uguale o superiore al doppio della profondità minima di copertura utilizzata in questo studio ( $\geq 6$ ), la correlazione tra le frequenze alleliche stimate da dati Ion Torrent e quelle ottenute dalla genotipizzazione con Illumina PorcineSNP60 BeadChip in suini Large White, è aumentata di circa 3 volte (Bovo *et al.*, 2015).

In questo studio abbiamo costruito RRLs come strategia per ridurre la complessità del genoma ed estrarre informazioni da una frazione di esso, campionata dopo digestione con enzimi di restrizione. Abbiamo identificato ~50k SNPs, un numero piuttosto elevato

considerando il throughput limitato della tecnologia Ion Torrent rispetto le piattaforme Illumina (Quail *et al.*, 2012) e considerando il fatto che abbiamo escluso dall'analisi le INDEL, dato che la tecnologia è soggetta ad errori in caso di regioni omopolimeriche (Bragg *et al.*, 2013).

Abbiamo identificato 277 mutazioni missenso, 38 delle quali classificate come deleterie da SIFT (tabella 13). Queste mutazioni, con putativi effetti sulla funzione della proteina codificata, potrebbero essere caratterizzati in futuri studi di associazione.

### **3.7. Conclusioni**

L'elevato numero di geni coinvolti nella determinazione di tratti connessi alla deposizione dei grassi, ed i loro effetti relativamente piccoli, così come le interazioni epistatiche tra i geni, rendono complesso lo studio dei meccanismi molecolari che ne regolano la manifestazione fenotipica, così come l'identificazione delle eventuali mutazioni causative (Switonski *et al.*, 2010). Anche se le associazioni tra i 2 polimorfismi e il carattere allo studio sono risultate significative nel presente lavoro di tesi, l'effetto sui fenotipi potrebbe non essere stato indotto dalle mutazioni nei geni indagati ma da altre in *linkage disequilibrium* con essi.

Questo studio è il primo che applica un approccio di selective genotyping congiunto alla strategia di DNA pooling, utilizzando RRLs, per studi di associazione in suino Nero Siciliano, ed il primo che impiega la tecnologia Ion Torrent per l'identificazione di marcatori da impiegare in successivi studi nella razza. L'obiettivo era quello di impostare una strategia che potesse ridurre il più possibile il costo di sequenziamento, e che potesse produrre dati utili per l'identificazione di marcatori per il carattere target nel suino Nero Siciliano. Lo spessore del grasso dorsale è, infatti, un carattere importante per la produzione di prosciutti crudi di alta qualità, per i quali la copertura di grasso è un fattore chiave durante le fasi di lavorazione e stagionatura. Per il suino Nero Siciliano non si riscontrano in letteratura studi di associazione tra polimorfismi e tratti produttivi e questo lavoro, a nostro parere, può rappresentare un importante contributo nel processo di caratterizzazione della razza. I risultati ottenuti potrebbero rappresentare un primo *step* per il possibile sviluppo di strumenti per la tipizzazione high-throughput della razza, anche ai fini selettivi, nonché per la formulazione di schemi di gestione riproduttiva che, congiuntamente alla necessità di preservarne la variabilità, potrebbero promuovere una selezione coerente con la qualità dei prodotti.

**Tab.13- SNPs identificati nelle sole regioni trascritte.** Tra quelli localizzati nelle regioni codificanti del genoma, 96 erano mutazioni sinonime, 277 missenso, 10 stop-gained, 9 stop lost e 1 stop\_retained. Tra le mutazioni missenso predette, 38 sono state classificate come deleterie da SIFT.

Location	Allele	Consequence	Symbol	Gene	Exon	cDNA position	CDS position	Protein position	Amino Acids	Existing variant	SIFT
1:23541-23541	T	5_prime_UTR_variant	PSMB1	ENSSSCG00000027257	1/6	174	-	-	-	rs1113871278	-
6:131728030-131728030	G	3_prime_UTR_variant	ADGRL2	ENSSSCG00000003761	23/23	5342	-	-	-	rs196950207	-
6:131728292-131728292	G	3_prime_UTR_variant	ADGRL2	ENSSSCG00000003761	23/23	5380	-	-	-	rs328537502	-
6:131728301-131728301	T	3_prime_UTR_variant	ADGRL2	ENSSSCG00000003761	21/21	4558	-	-	-	rs338304955	-
6:133597183-133597183	G	missense_variant	-	ENSSSCG00000027421	4/4	778	778	260	W/R	rs696437272	1
6:134426181-134426181	G	missense_variant	ADGRL4	ENSSSCG00000003762	8/14	1544	971	324	N/S	rs196956938	0,71
6:134426202-134426202	A	missense_variant	ADGRL4	ENSSSCG00000003762	8/14	1565	992	331	R/K	rs196949043	0,94
6:134461369-134461369	A	3_prime_UTR_variant	ADGRL4	ENSSSCG00000003762	15/15	3208	-	-	-	rs323824066	-
6:134461390-134461390	T	3_prime_UTR_variant	ADGRL4	ENSSSCG00000003762	15/15	3229	-	-	-	rs196955760	-
6:134463885-134463885	C	3_prime_UTR_variant	ADGRL4	ENSSSCG00000003762	16/16	5352	-	-	-	rs81391749	-
6:134464111-134464111	C	3_prime_UTR_variant	ADGRL4	ENSSSCG00000003762	16/16	5578	-	-	-	rs340363097	-
6:134464920-134464920	T	3_prime_UTR_variant	ADGRL4	ENSSSCG00000003762	15/15	6759	-	-	-	rs319309130	-
6:134465292-134465292	T	3_prime_UTR_variant	ADGRL4	ENSSSCG00000003762	16/16	6759	-	-	-	rs344380210	-
6:134465361-134465361	A	3_prime_UTR_variant	ADGRL4	ENSSSCG00000003762	15/15	7200	-	-	-	rs329499082	-
6:134649534-134649534	G	missense_variant	IFI44	ENSSSCG00000003763	2/8	1027	133	45	E/Q	rs709073642	0,56
6:134649699-134649699	T	missense_variant	IFI44	ENSSSCG00000003763	1/7	1113	190	64	V/I	rs691317383	1
6:134660352-134660352	A	3_prime_UTR_variant	IFI44L	ENSSSCG00000027660	8/8	1551	-	-	-	rs691218797	-
6:134660427-134660427	G	3_prime_UTR_variant	IFI44L	ENSSSCG00000027660	9/10	2538	-	-	-	rs1113714155	-
6:134660565-134660565	C	3_prime_UTR_variant	IFI44L	ENSSSCG00000027660	8/8	1338	-	-	-	rs691071864	-
6:134660599-134660599	T	3_prime_UTR_variant	IFI44L	ENSSSCG00000027660	9/9	1497	-	-	-	rs700795419	-
6:134665287-134665287	A	missense_variant	IFI44L	ENSSSCG00000027660	7/10	1095	905	302	S/I	rs81213480	0,01
6:134667395-134667395	T	missense_variant	IFI44L	ENSSSCG00000027660	5/8	941	826	276	V/M	rs81213477	1
6:134727476-134727476	T	3_prime_UTR_variant	PTGFR	ENSSSCG00000039548	3/3	4861	-	-	-	rs327513536	-
6:134727540-134727540	T	3_prime_UTR_variant	PTGFR	ENSSSCG00000039548	3/3	4797	-	-	-	rs318740062	-
6:134728506-134728506	C	3_prime_UTR_variant	PTGFR	ENSSSCG00000039548	3/3	3831	-	-	-	rs322833457	-

6:134728589-134728589	G	3_prime_UTR_variant	PTGFR	ENSSSCG00000039548	3/3	3748	-	-	-	rs334439102	-
6:134728888-134728888	A	3_prime_UTR_variant	PTGFR	ENSSSCG00000039548	3/3	3449	-	-	-	rs321770356	-
6:134729191-134729191	A	3_prime_UTR_variant	PTGFR	ENSSSCG00000039548	3/3	3146	-	-	-	rs81299115	-
6:134729553-134729553	G	3_prime_UTR_variant	PTGFR	ENSSSCG00000039548	3/3	2784	-	-	-	rs342951556	-
6:134729722-134729722	G	3_prime_UTR_variant	PTGFR	ENSSSCG00000039548	3/3	2615	-	-	-	rs318752531	-
6:134729780-134729780	A	3_prime_UTR_variant	PTGFR	ENSSSCG00000039548	3/3	2557	-	-	-	rs337556876	-
6:134729837-134729837	A	3_prime_UTR_variant	PTGFR	ENSSSCG00000039548	3/3	2500	-	-	-	rs320429148	-
6:134730788-134730788	G	3_prime_UTR_variant	PTGFR	ENSSSCG00000039548	3/3	1549	-	-	-	rs343834353	-
6:134730793-134730793	C	3_prime_UTR_variant	PTGFR	ENSSSCG00000039548	3/3	1544	-	-	-	rs324375315	-
6:135133466-135133466	T	3_prime_UTR_variant	GIPC2	ENSSSCG00000035582	6/6	1711	-	-	-	rs332883763	-
6:135133475-135133475	A	3_prime_UTR_variant	GIPC2	ENSSSCG00000035582	6/6	1702	-	-	-	rs345958940	-
6:135133493-135133493	A	3_prime_UTR_variant	GIPC2	ENSSSCG00000035582	6/6	1684	-	-	-	rs336517340	-
6:135133516-135133516	G	3_prime_UTR_variant	GIPC2	ENSSSCG00000035582	6/6	1661	-	-	-	rs345563992	-
6:135133643-135133643	T	3_prime_UTR_variant	GIPC2	ENSSSCG00000035582	6/6	1534	-	-	-	rs326714497	-
6:135133654-135133654	C	3_prime_UTR_variant	GIPC2	ENSSSCG00000035582	6/6	1523	-	-	-	rs337741248	-
6:135133655-135133655	A	3_prime_UTR_variant	GIPC2	ENSSSCG00000035582	6/6	1522	-	-	-	rs321241331	-
6:135133666-135133666	T	3_prime_UTR_variant	GIPC2	ENSSSCG00000035582	6/6	1511	-	-	-	rs335549975	-
6:135133678-135133678	G	3_prime_UTR_variant	GIPC2	ENSSSCG00000035582	6/6	1499	-	-	-	rs331568639	-
6:135133746-135133746	C	3_prime_UTR_variant	GIPC2	ENSSSCG00000035582	6/6	1431	-	-	-	rs329692247	-
6:135133748-135133748	G	3_prime_UTR_variant	GIPC2	ENSSSCG00000035582	6/6	1429	-	-	-	rs700762236	-
6:135133755-135133755	G	3_prime_UTR_variant	GIPC2	ENSSSCG00000035582	6/6	1422	-	-	-	rs710557133	-
6:135133762-135133762	C	3_prime_UTR_variant	GIPC2	ENSSSCG00000035582	6/6	1415	-	-	-	rs340563248	-
6:135133864-135133864	C	3_prime_UTR_variant	GIPC2	ENSSSCG00000035582	6/6	1313	-	-	-	rs324452958	-
6:135134079-135134079	T	3_prime_UTR_variant	GIPC2	ENSSSCG00000035582	6/6	1098	-	-	-	rs342580671	-
6:135244135-135244135	T	3_prime_UTR_variant	DNAJB4	ENSSSCG0000003766	4/4	2287	-	-	-	rs196960515	-
6:135247649-135247649	A	missense_variant	DNAJB4	ENSSSCG0000003766	3/4	1161	758	253	T/I	rs196960057	0,1
6:135314663-135314663	T	3_prime_UTR_variant	FUBP1	ENSSSCG0000003771	18/21	4640	-	-	-	rs323640415	-
6:135317642-135317642	C	3_prime_UTR_variant	FUBP1	ENSSSCG0000003771	21/21	5777	-	-	-	rs321865369	-
6:135446316-135446316	C	missense_variant	-	ENSSSCG00000035986	3/3	261	261	87	H/Q	rs196959795	-
6:135446318-135446318	C	missense_variant	-	ENSSSCG00000035986	3/3	259	259	87	H/D	rs196951013	-

6:135493202-135493202	C	missense_variant	MIGA1	ENSSSCG00000003770	3/16	289	177	59	H/Q	rs344218046	0,01
6:135719984-135719984	G	missense_variant	-	ENSSSCG00000003772	6/15	2190	1640	547	K/R	rs322696669	0,16
6:135806542-135806542	C	missense_variant	AK5	ENSSSCG00000003773	9/14	1465	1068	356	D/E	rs346239879	0,46
6:135806547-135806547	A	missense_variant	AK5	ENSSSCG00000003773	7/12	1006	895	299	D/Y	rs325394905	0
6:135863418-135863418	C	missense_variant	AK5	ENSSSCG00000003773	8/14	1415	1018	340	T/A	rs322178994	0,56
6:136105095-136105095	T	missense_variant	PIGK	ENSSSCG00000038329	9/10	995	941	314	T/I	-	0,04
6:137993677-137993677	G	missense_variant	SLC44A5	ENSSSCG00000003777	16/21	1590	1516	506	I/V	rs341825404	1
6:138455941-138455941	A	missense_variant	CRYZ	ENSSSCG00000003780	6/9	653	578	193	G/E	rs330593039	0
6:138583942-138583942	A	missense_variant	-	ENSSSCG00000003781	10/16	1557	808	270	A/T	rs345405009	1
6:138588805-138588805	T	missense_variant	-	ENSSSCG00000003781	11/16	1735	986	329	Q/L	rs81392161	0
6:138596918-138596918	G	missense_variant	-	ENSSSCG00000003781	12/16	2124	1375	459	T/A	rs332694930	1
6:138596924-138596924	A	missense_variant	-	ENSSSCG00000003781	12/16	2130	1381	461	A/T	rs324008647	0,06
6:138596982-138596982	T	missense_variant	-	ENSSSCG00000003781	12/16	2188	1439	480	A/V	rs336106966	0,3
6:138597134-138597134	G	missense_variant	-	ENSSSCG00000003781	12/16	2340	1591	531	K/E	rs334265153	0,44
6:138606656-138606656	A	missense_variant	-	ENSSSCG00000003781	13/16	2355	1606	536	V/I	rs340544850	1
6:138606695-138606695	T	missense_variant	-	ENSSSCG00000003781	13/16	2394	1645	549	G/C	rs331479715	0,16
6:138606716-138606716	G	missense_variant	-	ENSSSCG00000003781	13/16	2415	1666	556	S/G	rs339029139	1
6:138610961-138610961	C	missense_variant	-	ENSSSCG00000003781	14/16	2641	1892	631	E/A	rs196953780	0,64
6:138611111-138611111	T	missense_variant	-	ENSSSCG00000003781	14/16	2791	2042	681	G/V	rs196956666	0,17
6:138703784-138703784	A	missense_variant	-	ENSSSCG00000033405	2/5	244	244	82	E/K	-	1
7:121383228-121383228	C	stop_lost	WDR25	ENSSSCG00000002517	5/5	1423	1423	475	*/R	-	-
7:35286654-35286654	C	3_prime_UTR_variant	-	ENSSSCG00000001596	11/11	2720	-	-	-	rs345873541	-
7:35286660-35286660	A	3_prime_UTR_variant	-	ENSSSCG00000001596	11/11	2714	-	-	-	rs325701187	-
7:35286761-35286761	C	3_prime_UTR_variant	-	ENSSSCG00000001596	11/11	2613	-	-	-	rs326273244	-
7:35286879-35286879	G	3_prime_UTR_variant	-	ENSSSCG00000001596	11/11	2495	-	-	-	rs80870581	-
7:35286935-35286935	T	3_prime_UTR_variant	-	ENSSSCG00000001596	11/11	2439	-	-	-	rs318292874	-
7:35286940-35286940	G	3_prime_UTR_variant	-	ENSSSCG00000001596	11/11	2434	-	-	-	rs332183228	-
7:35286941-35286941	C	3_prime_UTR_variant	-	ENSSSCG00000001596	11/11	2433	-	-	-	rs709814265	-
7:35286950-35286950	T	3_prime_UTR_variant	-	ENSSSCG00000001596	11/11	2424	-	-	-	rs342691920	-
7:35286998-35286998	C	3_prime_UTR_variant	-	ENSSSCG00000001596	11/11	2376	-	-	-	rs333680314	-

7:35287106-35287106	A	3_prime_UTR_variant	-	ENSSSCG0000001596	11/11	2268	-	-	-	rs325073735	-
7:35287237-35287237	C	3_prime_UTR_variant	-	ENSSSCG0000001596	11/11	2137	-	-	-	rs699018229	-
7:35287238-35287238	T	3_prime_UTR_variant	-	ENSSSCG0000001596	11/11	2136	-	-	-	rs705407908	-
7:36331390-36331390	T	3_prime_UTR_variant	UNC5CL	ENSSSCG00000027922	8/8	4406	-	-	-	rs709937745	-
7:36331395-36331395	A	3_prime_UTR_variant	UNC5CL	ENSSSCG00000027922	9/9	4357	-	-	-	rs697939879	-
7:36331414-36331414	A	3_prime_UTR_variant	UNC5CL	ENSSSCG00000027922	8/8	4382	-	-	-	rs707555989	-
7:36331641-36331641	C	3_prime_UTR_variant	UNC5CL	ENSSSCG00000027922	9/9	4111	-	-	-	rs342032111	-
7:36331695-36331695	T	3_prime_UTR_variant	UNC5CL	ENSSSCG00000027922	8/8	4101	-	-	-	rs346184708	-
7:36331734-36331734	T	3_prime_UTR_variant	UNC5CL	ENSSSCG00000027922	9/9	4018	-	-	-	rs335533792	-
7:36332038-36332038	T	3_prime_UTR_variant	UNC5CL	ENSSSCG00000027922	8/8	3758	-	-	-	rs325464996	-
7:36332071-36332071	T	3_prime_UTR_variant	UNC5CL	ENSSSCG00000027922	9/9	3681	-	-	-	rs336670948	-
7:36332132-36332132	A	3_prime_UTR_variant	UNC5CL	ENSSSCG00000027922	8/8	3664	-	-	-	rs320456218	-
7:36332264-36332264	A	3_prime_UTR_variant	UNC5CL	ENSSSCG00000027922	9/9	3488	-	-	-	rs336933569	-
7:36332488-36332488	A	3_prime_UTR_variant	UNC5CL	ENSSSCG00000027922	10/10	3121	-	-	-	rs318930121	-
7:36332969-36332969	C	3_prime_UTR_variant	UNC5CL	ENSSSCG00000027922	10/10	2640	-	-	-	rs332947455	-
7:36333004-36333004	T	3_prime_UTR_variant	UNC5CL	ENSSSCG00000027922	9/9	2748	-	-	-	rs322637060	-
7:36333070-36333070	C	3_prime_UTR_variant	UNC5CL	ENSSSCG00000027922	10/10	2539	-	-	-	rs344205500	-
7:36338563-36338563	T	missense_variant	UNC5CL	ENSSSCG00000027922	4/9	1216	718	240	A/T	rs344317151	0,76
7:36339348-36339348	T	missense_variant	UNC5CL	ENSSSCG00000027922	4/10	971	616	206	A/T	rs342074282	0,62
7:36341338-36341338	C	missense_variant	UNC5CL	ENSSSCG00000027922	1/8	822	280	94	I/V	rs322556836	0,58
7:36341341-36341341	C	missense_variant	UNC5CL	ENSSSCG00000027922	1/8	819	277	93	I/V	rs334169645	1
7:36341383-36341383	C	missense_variant	UNC5CL	ENSSSCG00000027922	2/9	733	235	79	S/A	rs342604717	1
7:36341506-36341506	A	missense_variant	UNC5CL	ENSSSCG00000027922	3/10	467	112	38	R/W	-	0,26
7:36351191-36351191	A	missense_variant	TSPO2	ENSSSCG00000038553	2/4	815	5	2	R/Q	rs326447422	0,74
7:36351266-36351266	G	missense_variant	TSPO2	ENSSSCG00000038553	2/4	890	80	27	M/R	rs338275993	0,1
7:36351349-36351349	C	missense_variant	TSPO2	ENSSSCG00000038553	2/4	973	163	55	S/P	rs337679459	0,06
7:36351850-36351850	A	missense_variant	TSPO2	ENSSSCG00000038553	3/4	1117	307	103	A/T	rs345568753	0,4
7:36375062-36375062	T	missense_variant	APOBEC2	ENSSSCG00000036018	1/2	6004	68	23	R/L	rs1107918361	0,15
7:36437431-36437431	A	missense_variant	-	ENSSSCG0000001612	6/18	1139	1139	380	L/Q	rs80941501	0,63
7:36442289-36442289	G	missense_variant	-	ENSSSCG0000001612	9/18	1475	1475	492	L/R	rs80889625	0,05



7:36448781-36448781	G	missense_variant	-	ENSSSCG00000001612	16/18	2444	2444	815	Q/R	rs330776277	1
7:36449540-36449540	A	missense_variant	-	ENSSSCG00000001612	27/28	5033	2531	844	R/Q	rs343961428	1
7:36449723-36449723	C	missense_variant	-	ENSSSCG00000001612	17/18	2840	2840	947	M/T	rs324833299	0
7:36456041-36456041	G	missense_variant	TREML1	ENSSSCG00000001613	6/6	817	766	256	S/P	rs325641107	0,09
7:36469497-36469497	C	missense_variant	TREM2	ENSSSCG000000039270	2/5	163	163	55	S/G	rs332094134	1
7:36485523-36485523	C	missense_variant	TREML2	ENSSSCG00000001616	3/5	653	448	150	I/V	rs320079266	1
7:36490577-36490577	G	missense_variant	TREML2	ENSSSCG00000001616	2/5	372	167	56	V/A	rs326714578	0,13
7:36490590-36490590	T	missense_variant	TREML2	ENSSSCG00000001616	2/5	359	154	52	V/M	rs338501507	0,16
7:36546518-36546518	C	missense_variant	-	ENSSSCG00000001618	2/5	388	388	130	Y/H	rs345146269	0,18
7:36560228-36560228	T	missense_variant	-	ENSSSCG00000001618	5/5	689	689	230	T/I	rs80854282	0,13
7:36560261-36560261	G	missense_variant	-	ENSSSCG00000001618	5/5	722	722	241	S/W	rs324508129	0,43
7:36560284-36560284	G	missense_variant	-	ENSSSCG00000001618	5/5	745	745	249	K/E	rs332396398	1
7:36781399-36781399	A	missense_variant	FOXP4	ENSSSCG00000001619	7/16	1788	808	270	A/T	rs340251982	0,2
7:36975973-36975973	A	missense_variant	-	ENSSSCG000000034190	8/10	1323	877	293	A/T	-	0,13
7:37086377-37086377	C	missense_variant	BYSL	ENSSSCG000000026752	7/8	1278	1219	407	S/P	rs328772640	0,29
7:37182590-37182590	C	missense_variant	TAF8	ENSSSCG000000022224	2/10	596	32	11	G/A	-	0,99
7:37225252-37225252	G	missense_variant	C6orf132	ENSSSCG000000036801	4/5	2459	2057	686	V/A	-	-
7:37225273-37225273	C	missense_variant	C6orf132	ENSSSCG000000036801	4/5	2438	2036	679	E/G	-	-
7:37225286-37225286	G	missense_variant	C6orf132	ENSSSCG000000036801	4/5	2425	2023	675	K/Q	-	-
7:37225313-37225313	T	missense_variant	C6orf132	ENSSSCG000000036801	4/5	2398	1996	666	A/T	-	-
7:37225852-37225852	C	missense_variant	C6orf132	ENSSSCG000000036801	4/5	1859	1457	486	V/G	-	-
7:37226468-37226468	T	missense_variant	C6orf132	ENSSSCG000000036801	4/5	1243	841	281	L/M	rs695876155	-
7:37301520-37301520	A	missense_variant	GUCA1B	ENSSSCG00000001637	2/4	230	230	77	P/L	rs325609339	1
7:37345766-37345766	T	missense_variant	TRERF1	ENSSSCG00000001639	12/14	3022	3022	1008	V/M	rs337403608	0,05
7:37379148-37379148	A	missense_variant	TRERF1	ENSSSCG00000001639	4/16	1562	1048	350	P/S	rs332218035	0,09
7:37570147-37570147	C	missense_variant	-	ENSSSCG000000034453	2/7	56	56	19	V/A	rs340388435	1
7:37570425-37570425	T	missense_variant	-	ENSSSCG000000034453	4/7	282	282	94	R/S	rs333036484	0
7:37570766-37570766	C	missense_variant	-	ENSSSCG000000034453	6/7	584	584	195	V/A	rs332449983	0,77
7:37573623-37573623	G	stop_lost	-	ENSSSCG000000034453	7/7	667	667	223	*/E	rs334099333	-
7:37739917-37739917	G	missense_variant	UBR2	ENSSSCG00000001641	22/46	2861	2413	805	T/A	rs321902569	0,54

7:37767426-37767426	G	missense_variant	UBR2	ENSSSCG00000001641	37/46	4551	4103	1368	N/S	rs337481913	0,55
7:37769662-37769662	T	missense_variant	UBR2	ENSSSCG00000001641	40/46	4700	4444	1482	A/S	rs339195919	0,8
7:37823484-37823484	G	missense_variant	TBCC	ENSSSCG000000026741	1/1	511	494	165	G/A	rs328644586	0,49
7:37823536-37823536	C	missense_variant	TBCC	ENSSSCG000000026741	1/1	459	442	148	I/V	rs339819402	0,17
7:54161986-54161986	A	missense_variant	-	ENSSSCG000000033284	1/2	161	161	54	G/E	rs80977845	0,8
7:54162046-54162046	G	missense_variant	-	ENSSSCG000000033284	1/2	221	221	74	N/S	rs323927491	0,19
7:54162111-54162111	C	missense_variant	-	ENSSSCG000000033284	1/2	286	286	96	F/L	rs332584727	1
7:54162115-54162115	C	missense_variant	-	ENSSSCG000000033284	1/2	290	290	97	F/S	rs691370528	1
7:54162208-54162208	A	missense_variant	-	ENSSSCG000000033284	1/2	383	383	128	T/N	rs340497630	0,15
7:68412730-68412730	C	stop_lost	HECTD1	ENSSSCG00000001971	36/43	6970	6419	2140	*/S	-	-
7:77064917-77064917	A	stop_lost	-	ENSSSCG000000023624	1/1	117	117	39	*/Y	-	-
7:77216540-77216540	T	stop_lost	-	ENSSSCG000000039839	4/4	391	391	131	*/K	rs325597085	-
8:10808944-10808944	T	missense_variant, splice_region_variant	CC2D2A	ENSSSCG00000008738	4/35	281	281	94	P/L	rs712639410	0,2
8:10816428-10816428	A	missense_variant	CC2D2A	ENSSSCG00000008738	6/35	619	619	207	A/T	rs81316899	1
8:10833254-10833254	A	missense_variant	CC2D2A	ENSSSCG00000008738	13/35	1721	1721	574	R/Q	rs322041672	0,6
8:108382419-108382419	T	stop_gained	-	ENSSSCG000000038163	2/5	262	262	88	R/*	rs708401550	-
8:10867919-10867919	G	missense_variant	CC2D2A	ENSSSCG00000008738	23/35	3108	3108	1036	D/E	rs333338217	0,39
8:10874177-10874177	T	missense_variant	CC2D2A	ENSSSCG00000008738	27/35	3520	3520	1174	V/F	rs319717016	0,79
8:10874193-10874193	A	missense_variant	CC2D2A	ENSSSCG00000008738	27/35	3536	3536	1179	A/E	rs330969332	1
8:10880382-10880382	A	missense_variant	CC2D2A	ENSSSCG00000008738	28/35	3649	3649	1217	A/T	rs320634746	0,49
8:10915806-10915806	A	3_prime_UTR_variant	FBXL5	ENSSSCG00000008740	11/11	2632	-	-	-	rs345686673	-
8:11011715-11011715	T	missense_variant	BST1	ENSSSCG000000028359	1/7	144	61	21	I/F	-	0,06
8:11011731-11011731	G	missense_variant	BST1	ENSSSCG000000028359	1/7	160	77	26	Q/R	rs340511541	0,52
8:11011842-11011842	G	missense_variant	BST1	ENSSSCG000000028359	1/7	271	188	63	H/R	rs346072948	0,33
8:11029356-11029356	T	3_prime_UTR_variant	BST1	ENSSSCG000000028359	7/7	1021	-	-	-	rs320788323	-
8:11029471-11029471	T	3_prime_UTR_variant	BST1	ENSSSCG000000028359	7/7	1136	-	-	-	rs338895469	-
8:11029512-11029512	C	3_prime_UTR_variant	BST1	ENSSSCG000000028359	7/7	1177	-	-	-	rs336347027	-
8:11029560-11029560	G	3_prime_UTR_variant	BST1	ENSSSCG000000028359	7/7	1225	-	-	-	rs340044646	-
8:11029588-11029588	C	3_prime_UTR_variant	BST1	ENSSSCG000000028359	7/7	1253	-	-	-	rs334175979	-
8:11029617-11029617	A	3_prime_UTR_variant	BST1	ENSSSCG000000028359	7/7	1282	-	-	-	rs333540718	-

8:11029664-11029664	A	3_prime_UTR_variant	BST1	ENSSSCG00000028359	7/7	1329	-	-	-	rs346097470	-
8:11029702-11029702	A	3_prime_UTR_variant	BST1	ENSSSCG00000028359	7/7	1367	-	-	-	rs341225418	-
8:11030001-11030001	T	3_prime_UTR_variant	BST1	ENSSSCG00000028359	7/7	1666	-	-	-	rs713933926	-
8:11030047-11030047	A	3_prime_UTR_variant	BST1	ENSSSCG00000028359	7/7	1712	-	-	-	rs341092026	-
8:11030084-11030084	C	3_prime_UTR_variant	BST1	ENSSSCG00000028359	7/7	1749	-	-	-	rs346107744	-
8:11030186-11030186	C	3_prime_UTR_variant	BST1	ENSSSCG00000028359	7/7	1851	-	-	-	rs345324415	-
8:11030231-11030231	A	3_prime_UTR_variant	BST1	ENSSSCG00000028359	7/7	1896	-	-	-	rs344283708	-
8:11030237-11030237	C	3_prime_UTR_variant	BST1	ENSSSCG00000028359	7/7	1902	-	-	-	rs693719303	-
8:11030291-11030291	A	3_prime_UTR_variant	BST1	ENSSSCG00000028359	7/7	1956	-	-	-	rs703401642	-
8:11110786-11110786	T	missense_variant	CD38	ENSSSCG00000008742	2/8	723	362	121	T/I	rs321511786	0,05
8:11118007-11118007	A	missense_variant	CD38	ENSSSCG00000008742	4/8	885	524	175	L/H	rs335118010	0,05
8:11129483-11129483	A	3_prime_UTR_variant	CD38	ENSSSCG00000008742	8/8	1510	-	-	-	rs337105733	-
8:11129506-11129506	G	3_prime_UTR_variant	CD38	ENSSSCG00000008742	8/8	1533	-	-	-	rs320846944	-
8:11129580-11129580	G	3_prime_UTR_variant	CD38	ENSSSCG00000008742	8/8	1607	-	-	-	rs340003527	-
8:11129805-11129805	C	3_prime_UTR_variant	CD38	ENSSSCG00000008742	8/8	1832	-	-	-	rs323827627	-
8:11129839-11129839	G	3_prime_UTR_variant	CD38	ENSSSCG00000008742	8/8	1866	-	-	-	rs335920586	-
8:11129931-11129931	C	3_prime_UTR_variant	CD38	ENSSSCG00000008742	8/8	1958	-	-	-	rs690331905	-
8:11129933-11129933	T	3_prime_UTR_variant	CD38	ENSSSCG00000008742	8/8	1960	-	-	-	rs701448033	-
8:11129934-11129934	G	3_prime_UTR_variant	CD38	ENSSSCG00000008742	8/8	1961	-	-	-	rs709619747	-
8:11130010-11130010	A	3_prime_UTR_variant	CD38	ENSSSCG00000008742	8/8	2037	-	-	-	rs326254665	-
8:11130032-11130032	G	3_prime_UTR_variant	CD38	ENSSSCG00000008742	8/8	2059	-	-	-	rs333998125	-
8:11130043-11130043	G	3_prime_UTR_variant	CD38	ENSSSCG00000008742	8/8	2070	-	-	-	rs344106990	-
8:11130310-11130310	T	3_prime_UTR_variant	CD38	ENSSSCG00000008742	8/8	2337	-	-	-	rs337117012	-
8:11130476-11130476	G	3_prime_UTR_variant	CD38	ENSSSCG00000008742	8/8	2503	-	-	-	rs318802626	-
8:11130535-11130535	A	3_prime_UTR_variant	CD38	ENSSSCG00000008742	8/8	2562	-	-	-	rs343931402	-
8:11130658-11130658	C	3_prime_UTR_variant	CD38	ENSSSCG00000008742	8/8	2685	-	-	-	rs709909949	-
8:11130678-11130678	G	3_prime_UTR_variant	CD38	ENSSSCG00000008742	8/8	2705	-	-	-	rs699578321	-
8:11187106-11187106	C	missense_variant	FGFBP1	ENSSSCG00000008741	2/2	426	332	111	K/R	rs323212790	0,1
8:11199637-11199637	C	missense_variant	-	ENSSSCG00000040149	1/2	20	20	7	L/S	rs707038658	0,14
8:11199700-11199700	G	missense_variant	-	ENSSSCG00000040149	1/2	83	83	28	L/R	rs327948510	0,04

8:11243103-11243103	T	missense_variant	PROM1	ENSSSCG00000008745	11/24	1596	1169	390	G/E	rs341812147	0,95
8:11252814-11252814	C	missense_variant	PROM1	ENSSSCG00000008745	8/23	1191	896	299	H/R	rs338726201	0,34
8:121173426-121173426	G	stop_gained	ADH6	ENSSSCG00000039017	3/10	126	126	42	Y/*	-	-
8:124106936-124106936	A	stop_gained	-	ENSSSCG00000035632	2/3	112	112	38	Q/*	rs1113101070	-
8:12541403-12541403	T	missense_variant	QDPR	ENSSSCG00000034811	3/8	735	226	76	A/T	rs196950080	0,16
8:12552040-12552040	T	missense_variant	CLRN2	ENSSSCG00000023941	1/3	149	149	50	T/M	rs327727654	0
8:126045053-126045053	A	stop_gained	-	ENSSSCG00000009199	5/6	229	229	77	Q/*	rs344580626	-
8:12647129-12647129	G	missense_variant	FAM184B	ENSSSCG00000026232	12/18	2687	2305	769	W/R	rs345118103	0,32
8:12751383-12751383	G	missense_variant	DCAF16	ENSSSCG00000040393	2/2	551	551	184	H/P	rs325283333	0,2
8:12751482-12751482	G	missense_variant	-	ENSSSCG00000032553	1/1	103	103	35	R/G	rs337143235	-
8:12751575-12751575	C	missense_variant	DCAF16	ENSSSCG00000040393	2/2	359	359	120	N/S	rs319404303	1
8:12751689-12751689	T	missense_variant	DCAF16	ENSSSCG00000040393	2/2	245	245	82	G/E	rs329899902	1
8:14781870-14781870	A	missense_variant	-	ENSSSCG00000029791	5/5	451	451	151	A/T	rs333938110	-
8:14989868-14989868	C	missense_variant	SLIT2	ENSSSCG00000008749	1/33	74	74	25	V/A	rs337473719	0,01
8:15147958-15147958	G	missense_variant	SLIT2	ENSSSCG00000008749	34/34	4277	4277	1426	N/S	rs327869416	1
8:16667725-16667725	A	missense_variant	ADGRA3	ENSSSCG00000008753	13/13	2551	2551	851	A/S	rs324350840	0,3
8:16691060-16691060	T	missense_variant	ADGRA3	ENSSSCG00000008753	8/13	1387	1387	463	V/I	-	0,13
8:17867068-17867068	T	missense_variant	PPARGC1A	ENSSSCG00000029275	8/13	1547	1288	430	C/S	rs45430917	0,47
8:18810600-18810600	A	missense_variant	CCDC149	ENSSSCG00000023001	13/13	1341	1273	425	P/S	rs323690682	0,27
8:19103417-19103417	A	missense_variant	SEPSECS	ENSSSCG00000025315	10/11	1218	1146	382	K/N	rs330286802	0,37
8:19235492-19235492	A	missense_variant	-	ENSSSCG00000027340	10/11	1666	1666	556	V/M	rs326912759	0,06
8:19263608-19263608	T	missense_variant	ZCCHC4	ENSSSCG00000028581	4/13	369	344	115	T/I	rs344719517	1
8:19263728-19263728	G	missense_variant	ZCCHC4	ENSSSCG00000028581	4/13	489	464	155	K/R	rs708786531	0,83
8:19331295-19331295	A	missense_variant	ANAPC4	ENSSSCG00000025902	20/28	1495	1495	499	D/N	rs712793046	0,44
8:19331314-19331314	G	missense_variant	ANAPC4	ENSSSCG00000025902	19/27	1457	1457	486	N/S	rs698624933	1
8:19450441-19450441	A	missense_variant	-	ENSSSCG00000028632	1/1	752	752	251	A/V	rs328507353	1
8:19450454-19450454	G	missense_variant	-	ENSSSCG00000028632	1/1	739	739	247	N/H	rs340835779	0,08
8:19450589-19450589	C	missense_variant	-	ENSSSCG00000028632	1/1	604	604	202	N/D	rs337738339	0,59
8:19450972-19450972	A	missense_variant	-	ENSSSCG00000028632	1/1	221	221	74	P/L	rs342418141	1
8:19451048-19451048	T	missense_variant	-	ENSSSCG00000028632	1/1	145	145	49	E/K	rs324907731	1

8:19932576-19932576	C	5_prime_UTR_variant	RBPJ	ENSSSCG00000022155	1/11	111	-	-	-	rs334082295	-
8:19932618-19932618	G	5_prime_UTR_variant	RBPJ	ENSSSCG00000022155	1/12	153	-	-	-	rs335288304	-
8:19932753-19932753	C	5_prime_UTR_variant	RBPJ	ENSSSCG00000022155	1/11	288	-	-	-	rs318639472	-
8:19932898-19932898	A	5_prime_UTR_variant	RBPJ	ENSSSCG00000022155	1/12	433	-	-	-	rs336635062	-
8:20063265-20063265	A	5_prime_UTR_variant	RBPJ	ENSSSCG00000022155	1/12	116	-	-	-	rs339922431	-
8:20189527-20189527	A	3_prime_UTR_variant	CCKAR	ENSSSCG00000023501	5/5	1938	-	-	-	rs329784558	-
8:20189564-20189564	G	3_prime_UTR_variant	CCKAR	ENSSSCG00000023501	5/5	1901	-	-	-	rs339032662	-
8:20189685-20189685	C	3_prime_UTR_variant	CCKAR	ENSSSCG00000023501	5/5	1780	-	-	-	rs335049141	-
8:20189760-20189760	C	3_prime_UTR_variant	CCKAR	ENSSSCG00000023501	5/5	1705	-	-	-	rs321915361	-
8:20189766-20189766	C	3_prime_UTR_variant	CCKAR	ENSSSCG00000023501	5/5	1699	-	-	-	rs334871548	-
8:20189783-20189783	C	3_prime_UTR_variant	CCKAR	ENSSSCG00000023501	5/5	1682	-	-	-	rs343286940	-
8:20477641-20477641	A	5_prime_UTR_variant	STIM2	ENSSSCG00000008762	1/13	188	-	-	-	rs341698491	-
8:20607717-20607717	G	3_prime_UTR_variant	STIM2	ENSSSCG00000008762	13/13	1847	-	-	-	rs81242163	-
8:20608059-20608059	G	3_prime_UTR_variant	STIM2	ENSSSCG00000008762	12/12	2694	-	-	-	rs323812716	-
8:20608764-20608764	G	3_prime_UTR_variant	STIM2	ENSSSCG00000008762	13/13	2894	-	-	-	rs45434003	-
8:20608916-20608916	G	3_prime_UTR_variant	-	ENSSSCG00000032413	4/4	1654	-	-	-	rs45434005	-
8:23875056-23875056	A	3_prime_UTR_variant	-	ENSSSCG00000035424	1/1	1659	-	-	-	rs332811508	-
8:26790641-26790641	T	stop_gained	-	ENSSSCG00000008767	1/1	40	40	14	R/*	rs323411587	-
8:26790713-26790713	T	missense_variant	-	ENSSSCG00000008767	1/1	112	112	38	R/W	rs339394402	0
8:26791037-26791037	A	missense_variant	-	ENSSSCG00000008767	1/1	436	436	146	V/M	rs330299542	0,11
8:26791060-26791060	C	stop_lost	-	ENSSSCG00000008767	1/1	459	459	153	*/Y	rs332940343	-
8:27771589-27771589	A	3_prime_UTR_variant	ARAP2	ENSSSCG00000008768	33/33	7696	-	-	-	rs335708504	-
8:27771793-27771793	A	3_prime_UTR_variant	ARAP2	ENSSSCG00000008768	31/31	8488	-	-	-	rs328001433	-
8:27772843-27772843	T	3_prime_UTR_variant	ARAP2	ENSSSCG00000008768	33/33	6442	-	-	-	rs318685588	-
8:27779045-27779045	C	3_prime_UTR_variant	ARAP2	ENSSSCG00000008768	30/31	6238	-	-	-	rs327384248	-
8:27788533-27788533	C	3_prime_UTR_variant	ARAP2	ENSSSCG00000008768	28/31	5913	-	-	-	rs334808541	-
8:27788602-27788602	T	3_prime_UTR_variant	ARAP2	ENSSSCG00000008768	28/31	5844	-	-	-	rs332850754	-
8:27788634-27788634	G	3_prime_UTR_variant	ARAP2	ENSSSCG00000008768	28/31	5812	-	-	-	rs345447280	-
8:27788667-27788667	A	3_prime_UTR_variant	ARAP2	ENSSSCG00000008768	28/31	5779	-	-	-	rs339465871	-
8:27788671-27788671	G	3_prime_UTR_variant	ARAP2	ENSSSCG00000008768	28/31	5775	-	-	-	rs345536985	-

8:27788737-27788737	C	3_prime_UTR_variant	ARAP2	ENSSSCG0000008768	28/31	5709	-	-	-	rs328219890	-
8:27788797-27788797	A	3_prime_UTR_variant	ARAP2	ENSSSCG0000008768	28/31	5649	-	-	-	rs329043273	-
8:27788946-27788946	T	3_prime_UTR_variant	ARAP2	ENSSSCG0000008768	28/31	5500	-	-	-	rs337641021	-
8:27789142-27789142	T	3_prime_UTR_variant	ARAP2	ENSSSCG0000008768	28/31	5304	-	-	-	rs330598879	-
8:27789213-27789213	A	3_prime_UTR_variant	ARAP2	ENSSSCG0000008768	28/31	5233	-	-	-	rs339484087	-
8:27789304-27789304	G	3_prime_UTR_variant	ARAP2	ENSSSCG0000008768	28/31	5142	-	-	-	rs321125905	-
8:27789352-27789352	A	3_prime_UTR_variant	ARAP2	ENSSSCG0000008768	28/31	5094	-	-	-	rs331406754	-
8:27789365-27789365	C	3_prime_UTR_variant	ARAP2	ENSSSCG0000008768	28/31	5081	-	-	-	rs339653003	-
8:27789382-27789382	T	3_prime_UTR_variant	ARAP2	ENSSSCG0000008768	28/31	5064	-	-	-	rs321686527	-
8:27789402-27789402	C	3_prime_UTR_variant	ARAP2	ENSSSCG0000008768	28/31	5044	-	-	-	rs332966889	-
8:27789442-27789442	T	3_prime_UTR_variant	ARAP2	ENSSSCG0000008768	28/31	5004	-	-	-	rs341274663	-
8:27798141-27798141	C	missense_variant	ARAP2	ENSSSCG0000008768	28/33	4929	4397	1466	A/G	rs322849707	0,89
8:27832607-27832607	T	missense_variant	ARAP2	ENSSSCG0000008768	22/33	4221	3689	1230	R/H	rs319851384	0,27
8:27846159-27846159	G	missense_variant	ARAP2	ENSSSCG0000008768	20/33	3846	3314	1105	I/T	rs330328663	1
8:27943469-27943469	A	missense_variant	ARAP2	ENSSSCG0000008768	2/33	1345	813	271	Q/H	rs324958025	0,1
8:27944035-27944035	T	missense_variant	ARAP2	ENSSSCG0000008768	2/31	758	247	83	A/T	rs336592977	0,84
8:27998592-27998592	A	missense_variant	DTHD1	ENSSSCG00000028471	1/10	293	67	23	V/M	rs318598143	0,03
8:27998719-27998719	C	missense_variant	DTHD1	ENSSSCG00000028471	1/10	420	194	65	I/T	rs331755755	0,39
8:28005086-28005086	T	missense_variant	DTHD1	ENSSSCG00000028471	3/10	1191	965	322	H/L	rs338105327	0,12
8:28009748-28009748	G	missense_variant	DTHD1	ENSSSCG00000028471	5/10	1723	1497	499	H/Q	rs332926502	0,05
8:28024842-28024842	G	missense_variant	DTHD1	ENSSSCG00000028471	6/10	1973	1747	583	M/V	rs342830478	1
8:28034396-28034396	C	missense_variant	DTHD1	ENSSSCG00000028471	8/10	2502	2276	759	C/S	rs341888006	1
8:28066880-28066880	A	missense_variant	DTHD1	ENSSSCG00000028471	9/10	2575	2349	783	N/K	rs336596629	1
8:28066891-28066891	C	missense_variant	DTHD1	ENSSSCG00000028471	9/10	2586	2360	787	R/P	rs318457747	0
8:28306376-28306376	C	missense_variant	-	ENSSSCG0000008769	1/6	361	137	46	V/A	rs327196006	0,87
8:28306568-28306568	G	missense_variant	-	ENSSSCG0000008769	2/6	469	245	82	D/G	rs327958340	1
8:28306979-28306979	T	missense_variant	-	ENSSSCG0000008769	3/7	440	440	147	T/M	rs322548218	0,26
8:28307521-28307521	A	missense_variant	-	ENSSSCG0000008769	4/6	1115	891	297	M/I	rs345696199	0,17
8:28308297-28308297	G	3_prime_UTR_variant	-	ENSSSCG0000008769	6/6	1481	-	-	-	rs196950523	-
8:28308454-28308454	T	3_prime_UTR_variant	-	ENSSSCG0000008769	6/6	1638	-	-	-	rs81246184	-

8:28308851-28308851	C	3_prime_UTR_variant	-	ENSSSCG00000008769	6/6	2035	-	-	-	rs341797416	-
8:28309353-28309353	T	3_prime_UTR_variant	-	ENSSSCG00000008769	6/6	2537	-	-	-	rs346030183	-
8:28309356-28309356	T	3_prime_UTR_variant	-	ENSSSCG00000008769	6/6	2540	-	-	-	rs326264050	-
8:28309452-28309452	T	3_prime_UTR_variant	-	ENSSSCG00000008769	6/6	2636	-	-	-	rs330535987	-
8:28309910-28309910	A	3_prime_UTR_variant	-	ENSSSCG00000008769	6/6	3094	-	-	-	rs330090074	-
8:28310049-28310049	G	3_prime_UTR_variant	-	ENSSSCG00000008769	6/6	3233	-	-	-	rs324946849	-
8:28312237-28312237	A	3_prime_UTR_variant	-	ENSSSCG00000008769	6/6	5421	-	-	-	rs81398723	-
8:28312284-28312284	G	3_prime_UTR_variant	-	ENSSSCG00000008769	6/6	5468	-	-	-	rs327705959	-
8:28312503-28312503	T	3_prime_UTR_variant	-	ENSSSCG00000008769	6/6	5687	-	-	-	rs704274349	-
8:28312709-28312709	C	3_prime_UTR_variant	-	ENSSSCG00000008769	6/6	5893	-	-	-	rs336169030	-
8:28313058-28313058	T	3_prime_UTR_variant	-	ENSSSCG00000008769	6/6	6242	-	-	-	rs343141419	-
8:29003580-29003580	A	missense_variant	NWD2	ENSSSCG00000008770	6/6	1801	1801	601	E/K	rs335531699	0,15
8:29003605-29003605	T	missense_variant	NWD2	ENSSSCG00000008770	7/7	1943	1943	648	S/F	rs319597964	0,03
8:29005185-29005185	G	missense_variant	NWD2	ENSSSCG00000008770	6/6	3406	3406	1136	S/A	rs81492456	0,12
8:29006912-29006912	A	3_prime_UTR_variant	NWD2	ENSSSCG00000008770	2/2	4428	-	-	-	rs345358306	-
8:29007019-29007019	A	3_prime_UTR_variant	NWD2	ENSSSCG00000008770	2/2	4535	-	-	-	rs328507733	-
8:29007154-29007154	T	3_prime_UTR_variant	NWD2	ENSSSCG00000008770	2/2	4670	-	-	-	rs338233785	-
8:29007409-29007409	C	3_prime_UTR_variant	NWD2	ENSSSCG00000008770	2/2	4925	-	-	-	rs321057822	-
8:29007667-29007667	T	3_prime_UTR_variant	NWD2	ENSSSCG00000008770	2/2	5183	-	-	-	rs339648184	-
8:29101734-29101734	G	missense_variant	C4orf19	ENSSSCG00000008771	5/5	603	352	118	K/E	rs340039723	0,67
8:29101738-29101738	C	missense_variant	C4orf19	ENSSSCG00000008771	4/4	463	356	119	L/P	rs323620020	1
8:29101774-29101774	G	missense_variant	C4orf19	ENSSSCG00000008771	4/4	499	392	131	A/G	rs331215623	0,03
8:29101818-29101818	A	missense_variant	C4orf19	ENSSSCG00000008771	2/2	458	436	146	A/T	rs341248300	1
8:29101834-29101834	A	missense_variant	C4orf19	ENSSSCG00000008771	5/5	703	452	151	C/Y	rs713305761	0,71
8:29101878-29101878	A	missense_variant	C4orf19	ENSSSCG00000008771	2/2	518	496	166	D/N	rs712893486	0,13
8:29101879-29101879	G	missense_variant	C4orf19	ENSSSCG00000008771	5/5	748	497	166	D/G	rs698458153	0,12
8:29102082-29102082	G	missense_variant	C4orf19	ENSSSCG00000008771	5/5	951	700	234	N/D	rs345302250	0,08
8:29102167-29102167	G	missense_variant	C4orf19	ENSSSCG00000008771	2/2	807	785	262	D/G	rs337520257	0,15
8:29102233-29102233	G	missense_variant	C4orf19	ENSSSCG00000008771	5/5	1102	851	284	E/G	rs325665596	1
8:29155501-29155501	T	missense_variant	RELL1	ENSSSCG00000008772	2/7	246	112	38	H/N	rs331874380	0,11

8:29331833-29331833	A	missense_variant	PGM2	ENSSSCG0000008773	9/14	1200	1138	380	A/T	rs340670960	0,58
8:29336565-29336565	T	missense_variant	PGM2	ENSSSCG0000008773	13/14	1753	1649	550	T/I	rs329896112	0,17
8:29381550-29381550	A	stop_gained	PGM2	ENSSSCG0000008773	14/14	1883	1779	593	W/*	rs338126113	-
8:29509375-29509375	C	missense_variant	TBC1D1	ENSSSCG00000028983	9/18	1282	1235	412	L/P	rs341488281	0,4
8:29513997-29513997	G	missense_variant	TBC1D1	ENSSSCG00000028983	10/18	1479	1432	478	S/G	rs335298904	0,52
8:29548906-29548906	A	missense_variant	TBC1D1	ENSSSCG00000028983	11/18	1596	1549	517	S/T	-	0,41
8:30125447-30125447	T	missense_variant	TLR10	ENSSSCG00000027877	2/2	2315	1765	589	E/K	rs81218846	0,11
8:30125448-30125448	A	missense_variant	TLR10	ENSSSCG00000027877	2/2	2314	1764	588	Q/H	-	1
8:30125459-30125459	G	missense_variant	TLR10	ENSSSCG00000027877	2/2	2303	1753	585	N/H	rs81218844	0,03
8:30125686-30125686	A	missense_variant	TLR10	ENSSSCG00000027877	2/2	2076	1526	509	Y/F	rs81218842	0,01
8:30125765-30125765	C	missense_variant	TLR10	ENSSSCG00000027877	2/2	1997	1447	483	T/A	rs81218841	0,13
8:30126912-30126912	A	missense_variant	TLR10	ENSSSCG00000027877	2/2	850	300	100	Q/H	rs345618858	0,53
8:30126991-30126991	T	missense_variant	TLR10	ENSSSCG00000027877	2/2	771	221	74	V/D	rs81218838	0,07
8:30148693-30148693	A	missense_variant	TLR1	ENSSSCG00000026583	3/3	1719	1499	500	T/I	rs344399897	0,66
8:30229921-30229921	A	missense_variant	FAM114A1	ENSSSCG0000008785	5/14	710	622	208	V/M	rs322984357	0,9
8:30259497-30259497	C	missense_variant	FAM114A1	ENSSSCG0000008785	10/11	2027	1402	468	A/P	rs321567873	0,41
8:30306658-30306658	C	missense_variant	TMEM156	ENSSSCG00000023785	4/7	877	652	218	I/V	rs337326495	1
8:30309813-30309813	G	missense_variant	TMEM156	ENSSSCG00000023785	3/6	448	409	137	A/P	rs326987203	0,26
8:31122607-31122607	C	stop_lost	-	ENSSSCG00000036025	2/2	258	258	86	*/Y	rs336654126	-
8:31903048-31903048	A	stop_gained	-	ENSSSCG00000039302	1/2	151	151	51	R/*	rs325641850	-
8:52354281-52354281	T	stop_gained	TMA16	ENSSSCG0000008890	2/7	163	34	12	Q/*	-	-
8:66356066-66356066	A	stop_gained	-	ENSSSCG00000037521	2/5	85	85	29	Q/*	-	-
8:73410875-73410875	A	stop_gained	MRPL1	ENSSSCG0000008990	10/10	1451	905	302	L/*	rs340729607	-
8:8942845-8942845	T	3_prime_UTR_variant	-	ENSSSCG00000022358	8/8	1130	-	-	-	rs336799421	-
8:8951013-8951013	A	3_prime_UTR_variant	-	ENSSSCG00000022358	7/7	1335	-	-	-	rs320444052	-
8:9029432-9029432	A	3_prime_UTR_variant	-	ENSSSCG00000022358	7/7	4812	-	-	-	rs345411631	-
8:9029658-9029658	A	3_prime_UTR_variant	-	ENSSSCG00000022358	7/7	4586	-	-	-	rs326649404	-
8:9029859-9029859	T	3_prime_UTR_variant	-	ENSSSCG00000022358	7/7	4385	-	-	-	rs342830872	-
8:9030035-9030035	A	3_prime_UTR_variant	-	ENSSSCG00000022358	7/7	4209	-	-	-	rs335074844	-
8:9030298-9030298	T	3_prime_UTR_variant	-	ENSSSCG00000022358	7/7	3946	-	-	-	rs323848295	-



8:9030369-9030369	T	3_prime_UTR_variant	-	ENSSSCG00000022358	7/7	3875	-	-	-	rs332301480	-
8:9031943-9031943	G	3_prime_UTR_variant	-	ENSSSCG00000022358	7/7	2301	-	-	-	rs328029870	-
8:9032013-9032013	C	3_prime_UTR_variant	-	ENSSSCG00000022358	7/7	2231	-	-	-	rs335085013	-
8:9032247-9032247	C	3_prime_UTR_variant	-	ENSSSCG00000022358	8/8	1660	-	-	-	rs339314061	-
8:9195086-9195086	A	3_prime_UTR_variant	BOD1L1	ENSSSCG00000008735	26/26	12392	-	-	-	rs322851920	-
8:9195211-9195211	A	3_prime_UTR_variant	BOD1L1	ENSSSCG00000008735	26/26	12267	-	-	-	rs333495127	-
8:9195388-9195388	T	3_prime_UTR_variant	BOD1L1	ENSSSCG00000008735	26/26	12090	-	-	-	rs332907699	-
8:9195423-9195423	C	3_prime_UTR_variant	BOD1L1	ENSSSCG00000008735	26/26	12055	-	-	-	rs81497535	-
8:9196264-9196264	C	3_prime_UTR_variant	BOD1L1	ENSSSCG00000008735	26/26	11214	-	-	-	rs324332932	-
8:9197563-9197563	G	3_prime_UTR_variant	BOD1L1	ENSSSCG00000008735	26/26	9915	-	-	-	rs326083407	-
8:9197679-9197679	G	3_prime_UTR_variant	BOD1L1	ENSSSCG00000008735	26/26	9799	-	-	-	rs336644355	-
8:9221700-9221700	T	missense_variant	BOD1L1	ENSSSCG00000008735	12/26	8040	7857	2619	F/L	rs331158814	-
8:9225645-9225645	T	missense_variant	BOD1L1	ENSSSCG00000008735	10/26	7310	7127	2376	G/E	rs322788499	-
8:9225718-9225718	C	missense_variant	BOD1L1	ENSSSCG00000008735	10/26	7237	7054	2352	T/A	rs334211586	-
8:9227886-9227886	G	missense_variant	BOD1L1	ENSSSCG00000008735	10/26	5069	4886	1629	M/T	rs334867089	-
8:9241404-9241404	G	missense_variant	BOD1L1	ENSSSCG00000008735	4/26	931	748	250	T/P	rs342141094	-
8:9241422-9241422	T	missense_variant	BOD1L1	ENSSSCG00000008735	4/26	913	730	244	G/S	rs322599879	-
8:9241430-9241430	C	missense_variant	BOD1L1	ENSSSCG00000008735	4/26	905	722	241	T/S	rs331106848	-
8:9241593-9241593	C	missense_variant	BOD1L1	ENSSSCG00000008735	4/26	742	559	187	P/A	rs341222162	-
10:42741828-42741828	A	missense_variant	-	ENSSSCG00000039308	14/14	2639	2639	880	K/M	rs321197904	0,79
10:42741840-42741840	G	missense_variant	-	ENSSSCG00000039308	14/14	2627	2627	876	L/P	rs329696528	0,01
10:42741889-42741889	T	missense_variant	-	ENSSSCG00000039308	14/14	2578	2578	860	A/T	rs334703123	0,19
10:42741924-42741924	G	missense_variant	-	ENSSSCG00000039308	14/14	2543	2543	848	V/A	rs322628750	0,92
10:42742077-42742077	A	missense_variant, splice_region_variant	-	ENSSSCG00000039308	13/14	2407	2407	803	R/W	rs321569405	0
10:42742108-42742108	T	missense_variant	-	ENSSSCG00000039308	12/14	2393	2393	798	R/H	rs329958843	0,05
10:42742312-42742312	C	missense_variant	-	ENSSSCG00000039308	11/14	2216	2216	739	T/R	rs322972972	0
10:42742343-42742343	C	missense_variant	-	ENSSSCG00000039308	11/14	2185	2185	729	S/G	rs344361654	0
10:42742409-42742409	G	missense_variant	-	ENSSSCG00000039308	11/14	2119	2119	707	W/R	rs333013515	1
10:42742432-42742432	C	missense_variant	-	ENSSSCG00000039308	11/14	2096	2096	699	P/R	rs343848802	1
10:42742502-42742502	T	missense_variant	-	ENSSSCG00000039308	11/14	2026	2026	676	V/I	rs335456449	0,3

10:42742745-42742745	T	missense_variant	-	ENSSSCG00000039308	11/14	1783	1783	595	E/K	rs331517825	0,09
10:42742924-42742924	C	missense_variant	-	ENSSSCG00000039308	10/14	1643	1643	548	L/R	rs339688171	0
10:42742928-42742928	T	missense_variant	-	ENSSSCG00000039308	10/14	1639	1639	547	V/I	rs324485834	1
10:42742951-42742951	T	missense_variant	-	ENSSSCG00000039308	10/14	1616	1616	539	R/H	rs335069735	0,22
10:42742969-42742969	T	missense_variant	-	ENSSSCG00000039308	10/14	1598	1598	533	R/H	rs343473077	0,15
10:42743217-42743217	T	missense_variant	-	ENSSSCG00000039308	8/14	1411	1411	471	E/K	rs343794012	0,44
10:42743252-42743252	A	missense_variant	-	ENSSSCG00000039308	8/14	1376	1376	459	R/L	rs319376965	0,21
10:42744653-42744653	T	missense_variant	-	ENSSSCG00000039308	3/14	159	159	53	M/I	rs332272317	0
10:42835390-42835390	C	3_prime_UTR_variant	EPC1	ENSSSCG00000011028	14/14	4345	-	-	-	rs326266569	-
10:42964757-42964757	C	missense_variant	-	ENSSSCG00000036380	1/2	32	32	11	L/P	rs324960636	-
10:42964766-42964766	G	missense_variant	-	ENSSSCG00000036380	1/2	41	41	14	N/S	rs332598557	-
10:42964796-42964796	G	missense_variant	-	ENSSSCG00000036380	1/2	71	71	24	L/R	rs341266307	-
10:42964798-42964798	A	missense_variant	-	ENSSSCG00000036380	1/2	73	73	25	A/T	rs345380837	-
10:42967235-42967235	A	missense_variant	-	ENSSSCG00000036380	2/2	88	88	30	A/T	rs341598412	-
10:42967296-42967296	A	missense_variant	-	ENSSSCG00000036380	2/2	149	149	50	R/H	rs344575957	-
10:42967319-42967319	T	missense_variant	-	ENSSSCG00000036380	2/2	172	172	58	L/F	rs702942075	-
10:42967383-42967383	A	stop_retained_variant	-	ENSSSCG00000036380	2/2	236	236	79	*	rs342393728	-
10:42967495-42967495	G	missense_variant	-	ENSSSCG00000040023	1/1	53	53	18	D/G	rs325956995	1
10:42967569-42967569	A	missense_variant	-	ENSSSCG00000040023	1/1	127	127	43	G/R	rs320583244	1
10:42967572-42967572	T	missense_variant	-	ENSSSCG00000040023	1/1	130	130	44	R/C	rs331718508	
10:43168065-43168065	C	missense_variant	CUBN	ENSSSCG00000011030	57/68	8884	8884	2962	L/V	rs341199309	0,05
10:43175503-43175503	C	missense_variant	CUBN	ENSSSCG00000011030	55/68	8578	8578	2860	S/G	rs81210077	1
10:43182479-43182479	C	missense_variant	CUBN	ENSSSCG00000011030	52/68	8038	8038	2680	I/V	rs81210076	0,63
10:43185239-43185239	G	missense_variant	CUBN	ENSSSCG00000011030	50/68	7640	7640	2547	V/A	rs340541406	0,01
10:43207630-43207630	T	missense_variant	CUBN	ENSSSCG00000011030	42/68	6248	6248	2083	R/Q	rs334895833	0,07
10:43224120-43224120	T	missense_variant	CUBN	ENSSSCG00000011030	38/68	5419	5419	1807	G/S	rs339606270	0,02
10:43242399-43242399	G	missense_variant	CUBN	ENSSSCG00000011030	35/68	5024	5024	1675	S/T	rs81210074	1
10:43247629-43247629	T	missense_variant	CUBN	ENSSSCG00000011030	33/68	4856	4856	1619	R/Q	rs337666608	0,46
10:43349299-43349299	C	missense_variant	CUBN	ENSSSCG00000011030	22/68	2825	2825	942	I/R	-	0,03
10:43412884-43412884	A	missense_variant	CUBN	ENSSSCG00000011030	4/68	345	345	115	E/D	rs321372179	0,4

10:43450608-43450608	A	missense_variant	TRDMT1	ENSSSCG00000021309	5/12	2062	275	92	T/M	rs693182052	0,03
10:43530185-43530185	A	3_prime_UTR_variant	VIM	ENSSSCG00000011033	9/9	6091	-	-	-	-	-
10:43676323-43676323	T	missense_variant	ST8SIA6	ENSSSCG00000011034	3/8	325	256	86	D/N	rs702328444	0,46
10:43894649-43894649	G	missense_variant	-	ENSSSCG00000026941	4/5	100	100	34	I/V	rs332479087	0,31
10:43894649-43894649	G	missense_variant	-	ENSSSCG00000026941	4/4	100	100	34	I/V	rs332479087	0,3
10:43897211-43897211	C	3_prime_UTR_variant	HACD1	ENSSSCG00000038007	8/8	6768	-	-	-	rs335610564	-
10:43901434-43901434	C	3_prime_UTR_variant	HACD1	ENSSSCG00000038007	7/8	2821	-	-	-	rs324889798	-
10:43901441-43901441	A	3_prime_UTR_variant	HACD1	ENSSSCG00000038007	7/8	2814	-	-	-	rs329719660	-
10:43901453-43901453	C	3_prime_UTR_variant	HACD1	ENSSSCG00000038007	7/8	2802	-	-	-	rs340563698	-
13:160404513-160404513	G	3_prime_UTR_variant	CLDND1	ENSSSCG00000011980	6/6	1332	-	-	-	rs337870183	-
13:160407003-160407003	C	3_prime_UTR_variant	CLDND1	ENSSSCG00000011980	5/5	3813	-	-	-	rs326498619	-
13:160407456-160407456	A	3_prime_UTR_variant	CLDND1	ENSSSCG00000011980	5/5	4266	-	-	-	rs335968635	-
13:160425826-160425826	C	missense_variant	-	ENSSSCG00000035026	1/1	392	392	131	Q/R	rs338334700	0,11
13:160454908-160454908	A	missense_variant	-	ENSSSCG00000038050	1/1	821	821	274	A/V	rs322127041	0,66
13:160455711-160455711	T	missense_variant	-	ENSSSCG00000038050	1/1	18	18	6	H/Q	rs344263028	0,49
13:160683102-160683102	G	missense_variant	-	ENSSSCG00000037486	2/3	665	665	222	K/T	rs692747476	0,11
13:160683181-160683181	A	missense_variant	-	ENSSSCG00000037486	2/3	577	577	193	I/L	-	0,04
13:160683691-160683691	G	missense_variant	-	ENSSSCG00000037486	1/3	89	89	30	V/A	rs699987011	0,04
13:160969309-160969309	G	missense_variant	-	ENSSSCG00000036574	1/1	511	511	171	S/P	rs322952132	0,05
13:160992795-160992795	T	missense_variant	OR5K4	ENSSSCG00000011986	1/1	209	209	70	G/D	rs346342268	1
13:160992834-160992834	A	missense_variant	OR5K4	ENSSSCG00000011986	1/1	170	170	57	T/I	rs342429566	0,29
13:161104444-161104444	A	missense_variant	-	ENSSSCG00000034620	1/1	659	659	220	P/L	rs696310895	0,99
13:161153072-161153072	T	missense_variant	-	ENSSSCG00000038996	1/1	809	809	270	G/D	rs321771184	1
13:161309681-161309681	T	missense_variant	-	ENSSSCG00000040221	1/1	915	915	305	F/L	-	1
13:161363216-161363216	A	missense_variant	-	ENSSSCG00000036840	3/4	665	665	222	A/V	-	0,33
13:161363216-161363216	A	missense_variant	-	ENSSSCG00000036840	3/4	671	671	224	A/V	-	0,34

13:161469459-161469459	A	missense_variant	-	ENSSSCG00000032300	1/1	701	701	234	T/I	rs703234252	0,51
13:161469886-161469886	C	missense_variant	-	ENSSSCG00000032300	1/1	274	274	92	T/A	rs710248784	1
13:161519602-161519602	T	missense_variant	-	ENSSSCG00000040784	2/3	500	500	167	V/D	-	1
13:161554274-161554274	A	missense_variant	-	ENSSSCG00000038798	1/2	431	431	144	T/M	-	0,34
13:161595952-161595952	C	missense_variant	-	ENSSSCG00000034533	3/3	848	848	283	H/R	-	0
13:161610637-161610637	C	missense_variant	-	ENSSSCG00000032817	2/3	587	587	196	V/G	rs705463908	0
13:161610850-161610850	C	missense_variant	-	ENSSSCG00000032817	1/3	448	448	150	T/A	rs329048549	0,22
13:161611006-161611006	T	missense_variant	-	ENSSSCG00000032817	1/3	292	292	98	V/I	rs711963585	0,24
13:161680007-161680007	T	missense_variant	-	ENSSSCG00000040573	1/1	799	799	267	V/I	rs712956866	0,06
13:161733160-161733160	G	missense_variant	-	ENSSSCG00000040046	1/1	386	386	129	G/A	rs702913281	1
13:161733170-161733170	C	missense_variant	-	ENSSSCG00000040046	1/1	376	376	126	L/V	rs694634160	0,34
13:161754961-161754961	C	missense_variant	-	ENSSSCG00000031766	1/1	697	697	233	S/G	rs337015410	1
13:161755092-161755092	G	missense_variant	-	ENSSSCG00000031766	1/1	566	566	189	C/S	rs694590021	0
13:161755428-161755428	T	missense_variant	-	ENSSSCG00000031766	1/1	230	230	77	V/E	rs319006087	0
13:161755501-161755501	A	missense_variant	-	ENSSSCG00000031766	1/1	157	157	53	P/S	rs335823387	0,29
13:161790932-161790932	C	missense_variant	-	ENSSSCG00000033782	2/2	691	691	231	T/A	-	0,38
13:161791007-161791007	T	missense_variant	-	ENSSSCG00000033782	2/2	616	616	206	L/I	-	0,85
13:162069952-162069952	G	missense_variant	-	ENSSSCG00000038781	1/3	41	41	14	V/A	-	1
13:162069982-162069982	G	missense_variant	-	ENSSSCG00000038781	1/3	11	11	4	L/P	-	0,23
13:162069982-162069982	G	missense_variant	-	ENSSSCG00000038781	1/3	17	17	6	L/P	-	0,16
13:162092326-162092326	T	missense_variant	-	ENSSSCG00000033060	1/1	37	37	13	V/I	-	0,82
13:162217439-162217439	A	missense_variant	-	ENSSSCG00000038935	1/1	115	115	39	V/L	-	0,1
13:162354455-162354455	T	3_prime_UTR_variant	RIOX2	ENSSSCG00000034501	10/11	1844	-	-	-	-	-
13:162508415-162508415	C	3_prime_UTR_variant	ARL6	ENSSSCG00000027278	8/8	2537	-	-	-	-	-

13:162720478-162720478	A	missense_variant	-	ENSSSCG00000033884	1/1	212	212	71	R/K	rs329921086	1
13:162720588-162720588	T	missense_variant	-	ENSSSCG00000033884	1/1	322	322	108	A/S	rs333391249	0,66
13:162720886-162720886	T	missense_variant	-	ENSSSCG00000033884	1/1	620	620	207	T/I	rs345785625	1
13:162721069-162721069	A	missense_variant	-	ENSSSCG00000033884	1/1	803	803	268	G/E	rs700173951	0,01
13:162721199-162721199	C	missense_variant	-	ENSSSCG00000033884	1/1	933	933	311	E/D	rs700258833	1
13:162721209-162721209	T	missense_variant	-	ENSSSCG00000033884	1/1	943	943	315	H/Y	rs708533693	1
13:162721279-162721279	C	missense_variant	-	ENSSSCG00000033884	1/1	1013	1013	338	V/A	rs700942105	0,02
13:162721305-162721305	A	missense_variant	-	ENSSSCG00000033884	1/1	1039	1039	347	G/S	rs693345145	-
13:165056301-165056301	G	missense_variant	-	ENSSSCG00000033076	5/5	674	674	225	E/A	rs333032278	0,06
13:165764514-165764514	C	3_prime_UTR_variant	NSUN3	ENSSSCG00000027432	6/6	2139	-	-	-	rs319321350	-
13:165764779-165764779	T	3_prime_UTR_variant	NSUN3	ENSSSCG00000027432	6/6	1874	-	-	-	rs337542467	-
13:165872661-165872661	C	missense_variant	STX19	ENSSSCG00000025170	1/1	868	868	290	C/R	rs343787594	0,04
13:166594366-166594366	G	3_prime_UTR_variant	EPHA3	ENSSSCG00000026130	17/17	3960	-	-	-	rs338807218	-
13:167810204-167810204	C	3_prime_UTR_variant	ZNF654	ENSSSCG00000036885	8/8	3915	-	-	-	-	-
13:167958224-167958224	T	missense_variant	HTR1F	ENSSSCG00000030227	1/1	532	532	178	H/N	rs332575672	0,09
13:167958358-167958358	T	missense_variant	HTR1F	ENSSSCG00000030227	1/1	398	398	133	R/Q	rs324303045	0,15
13:168527081-168527081	T	3_prime_UTR_variant	POU1F1	ENSSSCG00000011996	6/6	1730	-	-	-	rs327559231	-
13:168531845-168531845	T	3_prime_UTR_variant	CHMP2B	ENSSSCG00000011997	6/6	4557	-	-	-	rs340497521	-
13:168532685-168532685	T	3_prime_UTR_variant	CHMP2B	ENSSSCG00000011997	6/6	3717	-	-	-	rs335876104	-
13:168533507-168533507	C	3_prime_UTR_variant	CHMP2B	ENSSSCG00000011997	6/6	2895	-	-	-	rs336614509	-
13:173910615-173910615	A	3_prime_UTR_variant	GBE1	ENSSSCG00000012000	17/17	3883	-	-	-	rs341523165	-
13:176450890-176450890	C	missense_variant	ROBO1	ENSSSCG00000012001	24/29	3615	3343	1115	T/P	rs318889696	0,22
13:176450890-176450890	C	missense_variant	ROBO1	ENSSSCG00000012001	24/29	3935	3583	1195	T/P	rs318889696	0,22
13:176462550-176462550	G	missense_variant	ROBO1	ENSSSCG00000012001	26/29	4156	3884	1295	A/G	rs333769517	1

13:176658767-176658767	A	missense_variant	-	ENSSSCG00000033742	1/1	40	40	14	V/I	rs81441424	0,11
13:176659086-176659086	A	missense_variant	-	ENSSSCG00000033742	1/1	359	359	120	R/K	rs335235313	1
13:178319188-178319188	G	stop_lost	-	ENSSSCG00000038214	4/4	514	514	172	*/E	rs327263348	-
13:179126899-179126899	C	3_prime_UTR_variant	LIPI	ENSSSCG00000012004	10/10	3929	-	-	-	rs331432158	-
13:179126901-179126901	T	3_prime_UTR_variant	LIPI	ENSSSCG00000012004	10/10	3927	-	-	-	rs339087058	-
13:179126936-179126936	T	3_prime_UTR_variant	LIPI	ENSSSCG00000012004	10/10	3892	-	-	-	rs694164353	-
13:179127226-179127226	A	3_prime_UTR_variant	LIPI	ENSSSCG00000012004	10/10	3952	-	-	-	rs344047012	-
13:179127288-179127288	C	3_prime_UTR_variant	LIPI	ENSSSCG00000012004	9/9	3785	-	-	-	rs324706126	-
13:179127364-179127364	C	3_prime_UTR_variant	LIPI	ENSSSCG00000012004	9/9	3709	-	-	-	rs332568773	-
13:194087627-194087627	C	stop_lost	-	ENSSSCG00000037613	1/1	186	186	62	*/C	-	-

## 4. Considerazioni generali e prospettive future

Nel corso del triennio di Dottorato l'impianto originario del progetto ha subito una rivisitazione rispetto quanto inizialmente previsto, in modo da poter ampliare il campo di indagine. In una prima fase era stato ipotizzato un disegno sperimentale di "Targeted resequencing", che prevedeva il sequenziamento di specifiche regioni in un set di geni di particolare interesse, su soggetti selezionati in accordo con l'approccio di selective genotyping. A tale scopo, era stata progettata una library con tecnologia AmpliSeq DNA. Con l'ausilio del software Ion Ampliseq Designer (Thermo Fisher Scientific) era stata utilizzata la pipeline *DNA Gene designs* per il disegno di un pannello custom che comprendesse 35 regioni genomiche correlate al metabolismo lipidico. In realtà, ricorrere ad un disegno di questo tipo, escludendo delle regioni a priori, avrebbe causato una perdita di informazioni. Si è dunque proceduto a una variazione della strategia di sequenziamento che potesse, in qualche modo, evitare una distorsione nel dato. Abbiamo, pertanto, modificato l'impianto originario del progetto focalizzandoci non più su un pannello di geni, ma optando per una strategia di SNP discovery nell'intero genoma, seppur "ridotto" nella sua complessità.

Abbiamo, inoltre, sequenziato l'intero genoma di un verro producendo, per la prima volta, un dato che riguardasse il genoma di suino Nero Siciliano nella sua interezza. Le raw reads sono disponibili presso l'NCBI SRA (accession number SRX3406507).

In questa fase è stata, tuttavia, condotta solo un'analisi preliminare che ha previsto il mapping vs il genoma di riferimento *Sscrofa 11.1* ed il variant calling.

Il prossimo passo riguarderà la ricostruzione dell'intero genoma mediante una strategia che integra l'assembly *ref-guided* e quello *de-novo*. A tal fine impiegheremo Reconstructor, una pipeline automatizzata in *silico*, che mira alla generazione di una sequenza completa del genoma di un individuo, a partire dai dati del genoma di riferimento e da quelli di risequenziamento. Reconstructor si basa su due *step* principali: mapping iterativo delle reads prodotte dal sequenziamento e assembly *de novo* delle reads non mappate nel primo *step*. La pipeline di Reconstructor è stata sviluppata dalla sequentia biotech ([www.sequentiabiotech.com](http://www.sequentiabiotech.com)).

Questo lavoro può rappresentare un primo contributo alla caratterizzazione genetica della razza, ma ulteriori studi saranno necessari per rendere trasferibili i risultati con possibili ricadute pratiche per l'intera filiera. Fino ad ora la gestione della razza ha riguardato la necessità di incrementare la consistenza della popolazione contenendo al contempo il tasso

di consanguineità. Tuttavia, per garantire il rinsanguamento e migliorare alcune caratteristiche commerciali, sono stati condotti incroci indiscriminati che hanno rischiato di inquinare il patrimonio genetico originario. È noto come la tutela di una razza sia strettamente connessa alla valorizzazione delle sue produzioni ed al miglioramento della qualità delle stesse, uno sforzo congiunto da parte di tutti i portatori di interesse è dunque auspicabile. Il riconoscimento della DOP, e l'applicazione del disciplinare, imporranno il rispetto di specifici requisiti, determinando la standardizzazione dei processi e, in ultima analisi, del prodotto.

L'ottenimento del marchio di qualità, incrementando il valore del prodotto lungo la filiera, ne consentirà un nuovo posizionamento in segmenti di mercato più ricchi. Tale impulso all'economia locale risulterebbe fondamentale soprattutto per le aree più depresse e marginali dell'entroterra nebroido.



## 5. Appendice

**Materiale supplementare 5.1: Accession number degli Assembly (a livello di cromosomi, scaffold o contig) presenti in NCBI per Sus scrofa (www.ncbi.nlm.nih.gov)**

### [Sscrofa11.1](#)

Sscrofa11 with Y sequences from WTSI\_X\_Y\_pig V2  
Organism: **Sus scrofa** (pig)  
Infraspecific name: Breed: Duroc  
Sex: female  
Submitter: The Swine Genome Sequencing Consortium (SGSC)  
Date: 2017/02/07  
Assembly level: Chromosome  
Genome representation: full  
RefSeq category: representative genome  
Synonyms: susScr11  
GenBank assembly accession: GCA\_000003025.6 (**latest**)  
RefSeq assembly accession: GCF\_000003025.6 (**latest**)  
IDs:1004191 [UID] 4121818 [GenBank] 4192498 [RefSeq]

### [USMARCv1.0](#)

Organism: **Sus scrofa** (pig)  
Infraspecific name: Breed: Cross-bred  
Sex: male  
Submitter: USDA ARS  
Date: 2017/12/20  
Assembly level: Chromosome  
Genome representation: full  
GenBank assembly accession: GCA\_002844635.1 (**latest**)  
RefSeq assembly accession: n/a  
IDs:1493281 [UID] 5798698 [GenBank]

### [ss10.2\\_mar2013](#)

Organism: **Sus scrofa** (pig)  
Infraspecific name: Breed: Goettingen  
Sex: female  
Submitter: F. Hoffmann - La Roche AG  
Date: 2015/09/16  
Assembly level: Scaffold  
Genome representation: full  
GenBank assembly accession: GCA\_001292865.1 (**latest**)  
RefSeq assembly accession: n/a  
IDs:485711 [UID] 2325318 [GenBank]

### [minipig\\_v1.0](#)

Organism: **Sus scrofa** (pig)  
Infraspecific name: Breed: Wuzhishan  
Sex: male  
Submitter: BGI-shenzhen  
Date: 2015/03/18  
Assembly level: Scaffold  
Genome representation: full  
GenBank assembly accession: GCA\_000325925.2 (**latest**)  
RefSeq assembly accession: n/a  
IDs:313291 [UID] 1690118 [GenBank]

### [SscrofaMinipig](#)

Organism: **Sus scrofa** (pig)  
Infraspecific name: Breed: Ellegaard Gottingen minipig  
Sex: female  
Submitter: GlaxoSmithKline  
Date: 2013/01/10  
Assembly level: Contig  
Genome representation: full  
GenBank assembly accession: GCA\_000331475.1 (**latest**)  
RefSeq assembly accession: n/a  
IDs:524158 [UID] 524158 [GenBank]

### [Tibetan Pig\\_v2](#)

Organism: **Sus scrofa** (pig)  
Infraspecific name: Breed: Tibetan  
Sex: female  
Submitter: Novogene  
Date: 2016/08/08  
Assembly level: Scaffold  
Genome representation: full  
GenBank assembly accession: GCA\_000472085.2 (**latest**)  
RefSeq assembly accession: n/a  
IDs:778631 [UID] 3376028 [GenBank]

### [Large White\\_v1](#)

Organism: **Sus scrofa** (pig)  
Infraspecific name: Breed: LargeWhite  
Sex: female  
Submitter: Novogene  
Date: 2016/08/05  
Assembly level: Scaffold  
Genome representation: full  
GenBank assembly accession: GCA\_001700135.1 (**latest**)

RefSeq assembly accession: n/a  
IDs:778301 [UID] 3374548 [GenBank]

[Rongchang pig v1](#)

Organism: **Sus scrofa** (pig)  
Intraspecific name: Breed: Rongchang  
Sex: female  
Submitter: Novogene  
Date: 2016/08/05  
Assembly level: Scaffold  
Genome representation: full  
GenBank assembly accession: GCA\_001700155.1 (**latest**)  
RefSeq assembly accession: n/a  
IDs:778311 [UID] 3374558 [GenBank]

[Hampshire pig v1](#)

Organism: **Sus scrofa** (pig)  
Intraspecific name: Breed: Hampshire  
Sex: female  
Submitter: Novogene  
Date: 2016/08/05  
Assembly level: Scaffold  
Genome representation: full  
GenBank assembly accession: GCA\_001700165.1 (**latest**)  
RefSeq assembly accession: n/a  
IDs:778321 [UID] 3374568 [GenBank]

[Meishan pig v1](#)

Organism: **Sus scrofa** (pig)  
Intraspecific name: Breed: Meishan  
Sex: female  
Submitter: Novogene  
Date: 2016/08/05  
Assembly level: Scaffold  
Genome representation: full  
GenBank assembly accession: GCA\_001700195.1 (**latest**)  
RefSeq assembly accession: n/a  
IDs:778331 [UID] 3374578 [GenBank]

[Landrace pig v1](#)

Organism: **Sus scrofa** (pig)  
Intraspecific name: Breed: Landrace  
Sex: female  
Submitter: Novogene  
Date: 2016/08/05  
Assembly level: Scaffold

Genome representation: full  
GenBank assembly accession: GCA\_001700215.1 (**latest**)  
RefSeq assembly accession: n/a  
IDs:778341 [UID] 3374598 [GenBank]

[Bamei\\_pig\\_v1](#)

Organism: **Sus scrofa** (pig)  
Infraspecific name: Breed: Bamei  
Sex: female  
Submitter: Novogene  
Date: 2016/08/05  
Assembly level: Scaffold  
Genome representation: full  
GenBank assembly accession: GCA\_001700235.1 (**latest**)  
RefSeq assembly accession: n/a  
IDs:778351 [UID] 3374618 [GenBank]

[Pietrain\\_pig\\_v1](#)

Organism: **Sus scrofa** (pig)  
Infraspecific name: Breed: Pietrain  
Sex: female  
Submitter: Novogene  
Date: 2016/08/05  
Assembly level: Scaffold  
Genome representation: full  
GenBank assembly accession: GCA\_001700255.1 (**latest**)  
RefSeq assembly accession: n/a  
IDs:778361 [UID] 3374638 [GenBank]

[Jinhua\\_pig\\_v1](#)

Organism: **Sus scrofa** (pig)  
Infraspecific name: Breed: Jinhua  
Sex: female  
Submitter: Novogene  
Date: 2016/08/05  
Assembly level: Scaffold  
Genome representation: full  
GenBank assembly accession: GCA\_001700295.1 (**latest**)  
RefSeq assembly accession: n/a  
IDs:778381 [UID] 3374678 [GenBank]

[Berkshire\\_pig\\_v1](#)

Organism: **Sus scrofa** (pig)  
Infraspecific name: Breed: Berkshire  
Sex: female  
Submitter: Novogene  
Date: 2016/08/05  
Assembly level: Scaffold  
Genome representation: full  
GenBank assembly accession: GCA\_001700575.1 (**latest**)  
RefSeq assembly accession: n/a  
IDs:778521 [UID] 3374958 [GenBank]

[WTSI X Y pig V2](#)

Organism: **Sus scrofa** (pig)  
Submitter: SC  
Date: 2017/02/08  
Assembly level: Chromosome  
Genome representation: partial  
GenBank assembly accession: GCA\_900119615.2 (**latest**)  
RefSeq assembly accession: n/a  
IDs:1002641 [UID] 4118188 [GenBank]

## Materiale supplementare 5.2: Illumina® TruSeq® DNA PCR-Free Library Prep

Il protocollo TruSeq DNA PCR-Free non prevede PCR amplification per abilitare la generazione dei *cluster*, evitando dunque le distorsioni ed artefatti tipicamente indotti dalla PCR, ed utilizza una *size selection* basata su microsfere al posto del gel. Tali microsfere, Sample Purification Beads (SPB), sono utilizzate anche per gli *step* di purificazione, non richiedendo dunque l'utilizzo di microsfere AMPure come previsto invece nel caso del Ion Torrent PGM System. La fase di frammentazione per sonicazione (Covaris® S220) è ottimizzata per progetti di whole genome resequencing con inserti di 350 o 550 bp. Nel presente lavoro di tesi la libreria paired-end è stata generata con frammenti di 350 pb. Il protocollo prevede l'aggiunta di una serie di controlli in-line, Il CTE (End Repair Control), CTA (A-Tailing Control) e CTL (Ligation Control) allo scopo di monitorare le attività dell'ERP 2 (End Repair Mix 2), ATL (A-Tailing Mix) e LIG 2 (Ligation Mix 2). Ogni controllo in-line contiene frammenti di dsDNA progettati per segnalare il successo o il fallimento di una specifica attività enzimatica e viene aggiunto alla reazione prima del relativo corrispondente *step* nel protocollo. Le molecole di controllo presentano strutture terminali corrispondenti alle strutture terminali di una molecola di DNA che non ha superato lo *step*. Se il passaggio ha successo, la molecola di controllo viene modificata per partecipare alle reazioni a valle della generazione della libreria e risulterà nei dati di sequenziamento. Se il passaggio non riesce, la molecola di controllo non procede nel processo e non vengono generati dati di sequenziamento.

Di seguito la descrizione delle molecole controllo.

Reagent	Function	Control	Structure of Control DNA Ends
ERP 2	End repair: Generate blunt ended fragments by 3'→ 5' exonuclease and 5'→ 3' polymerase activities	End repair control	5' overhang at 1 end, 3' overhang at other end
ERP 2	End repair: Add 5'-phosphate groups needed for downstream ligation	End repair control	Blunt with 5'-OH group
ATL	A-tailing: Make fragments compatible with adapters and prevent self-ligation by adding a 3'-A overhang	A-tailing control	Blunt with 5'- phosphate group
LIG 2	Ligation: Join 3'-T overhang adapters to 3'-A overhang inserts	Ligation control	Single-base 3' 'A' base overhang

## **Frammentazione gDNA e purificazione del gDNA frammentato**

1 µg di gDNA, dopo quantifica in fluorimetria, è stato normalizzato mediante aggiunta di Resuspension Buffer (RSB) fino al raggiungimento del volume finale di 55 µl ed utilizzato per la frammentazione mediante sonicazione (Covaris® S220). Questa tecnica genera frammenti con estremità protudenti (sticky end) al 5' o 3', è dunque richiesto end repair.

Brevemente, 52.5 µl DNA sono stati aggiunti ad un tubo Covaris, centrifugati a 280 x g per 5 secondi e frammentati per 50 secondi in modalità Frequency sweeping, a 6 °C, seguendo le indicazioni per inserti di 350 bp.

Il campione è stato poi centrifugato a 280 x g per 5 secondi e 50 µl di surnatante sono stati trasferiti nel Clean Up Sheared DNA Plate (CSP) per lo *step* di purificazione. 80 µl di Sample Purification Beads (SPB) sono stati addizionati al tubo, mixati ed incubati a temperatura ambiente per 5 minuti trascorsi i quali il tubo è stato posizionato su un rack magnetico, finché la soluzione non si è schiarita. Dopo ~ 7 minuti il tubo è stato rimosso dal rack, il surnatante scartato ed il pellet lavato due volte mediante aggiunta di 200 µl EtOH al 80%. È seguita una fase di incubazione sul rack magnetico per 30 secondi con successiva eliminazione del surnatante.

Dopo aver rimosso l'etanolo residuo la provetta è stata asciugata all'aria sul rack magnetico per 5 minuti, trascorsi i quali 52.5 µl di RSB sono stati aggiunti al tubo e mixati. Dopo rimozione dal rack, miscelazione ed incubazione a temperatura ambiente per 2 minuti, il tubo è stato nuovamente posto sul rack magnetico finché la soluzione non è diventata chiara (~ 4 minuti). 50 µl di surnatante sono stati dunque trasferiti nel Insert Modification Plate (IMP).

## **Repair Ends e Select Library Size**

In questa fase le estremità protudenti (sticky ends) risultanti dalla frammentazione sono convertite in estremità smussate (blunt ends) utilizzando End Repair Mix 2 (ERP2). L'attività esonucleasica 3'→5' di questa miscela rimuove le sporgenze al 3', mentre l'attività polimerasica 5'→3' riempie i gap al 5'. Dopo la riparazione, viene selezionata la dimensione appropriata della libreria utilizzando diversi rapporti del SPB.

Brevemente, 10 µl di End Repair Control (CTE) sono stati aggiunti, dopo centrifugazione a 600 × g per 5 secondi, al pozzetto contenente i 50 µl di surnatante, seguiti da 40 µl di ERP2, anch'essi dopo centrifugazione a 600 × g per 5 secondi, per un volume finale di 100 µl. Il campione è stato trattato a 30°C per 30 minuti, con lid preriscaldato a 100°C, e

mantenuto a 4°C su un termociclatore 2720 Thermal Cycler (Applied Biosystems, Foster City – USA).

Dopo la fase di end repair si è proceduto alla rimozione dei frammenti di grandi dimensioni in un primo *step* e di quelli di piccole dimensioni in un secondo *step*.

Brevemente, dopo aver miscelato 109.25 µl di SPB precedentemente agitato su vortex e 74.75 µl di PCR grade water, 160 µl di soluzione di SPB diluita sono stati aggiunti al campione e pipettati, incubati a temperatura ambiente per 5 minuti e posti su un rack magnetico finché la soluzione non è diventata chiara (~5 minuti).

250 µl di surnatante sono poi stati trasferiti al corrispondente pozzetto nel Clean Up End Repair Plate (CEP).

Nel secondo *step*, 30 µl di SPB non diluito, precedentemente agitati su vortex per risospendere le biglie, sono stati addizionati al tubo, miscelati ed incubati a temperatura ambiente per 5 minuti e posti su un rack magnetico finché la soluzione non è diventata chiara (~5 minuti).

Il tubo è stato rimosso dal rack, il surnatante scartato ed il pellet lavato due volte mediante aggiunta di 200 µl EtOH al 80%, incubazione sul rack magnetico per 30 secondi ed eliminazione del surnatante.

Dopo aver rimosso l'etanolo residuo la provetta è stata asciugata all'aria sul rack magnetico per 5 minuti trascorsi i quali, 17.5 µl di RSB sono stati aggiunti al tubo e mixati. Dopo rimozione dal rack, miscelazione ed incubazione a temperatura ambiente per 2 minuti, il tubo è stato nuovamente posto sul rack magnetico finché la soluzione non è diventata chiara (~ 4 minuti). 15 µl di surnatante sono stati dunque trasferiti nel corrispondente pozzetto del Adapter Ligation Plate (ALP).

### **Adenilazione al 3'end**

In questa fase una singola Adenina è addizionata all'estremità 3' dei frammenti blunt ends per impedire che si leghino l'un l'altro durante la reazione di ligazione dell'adattatore. Allo stesso modo, una Timina all'estremità 3' dell'adattatore fornisce una sporgenza complementare per garantire la ligazione dello stesso al frammento. Questa strategia garantisce un ridotto tasso di formazione di chimere (modello concatenato).

Brevemente, 2.5 µl di A-Tailing Control (CTA) sono stati addizionati ai 15 µl campione previa centrifugazione a  $600 \times g$  per 5 secondi seguiti da 12.5 µl di A-Tailing Mix (ATL) pretrattati allo stesso modo, per un volume finale di 30 µl. La mix così ottenuta è stata sottoposta al



seguito protocollo termico su un termociclatore 2720 Thermal Cycler (Applied Biosystems, Foster City – USA):

Stage	Temperature	Time
Hold	37° C	30 min
Hold	70° C	5 min
Hold	4° C	5 min

**Tab.5.2.1:** Protocollo termico per la reazione di Adenilazione. Il lid era preriscaldato a 100° C

### **Ligazione adattatori**

In questa fase avviene la ligazione degli adattatori alle estremità dei frammenti di DNA, allo scopo di prepararli per la successiva ibridazione sulla *flow cell*. Gli adattatori sono complementari ai primer di sequenziamento immobilizzati sulla *flow cell*, dunque il campione è pronto per la generazione dei *cluster* senza necessità di uno *step* aggiuntivo di PCR.

Brevemente, gli adattatori ed il Ligation Control (CTL) sono stati centrifugati a 600 x g per 5 secondi e 2.5 µl di ciascuno è stato addizionato al tubo, seguito dall'aggiunta di 2.5 µl di Ligation Mix 2 (LIG2). La mix è stata dunque miscelata, centrifugata a 280 x g per 1 minuto e trattata a 30°C per 10 minuti, con *lid* preriscaldato a 100°C, e mantenuto a 4°C su un termociclatore 2720 Thermal Cycler (Applied Biosystems, Foster City – USA). Il volume finale nel tubo era di 37.5 µl.

Al termine della termociclazione 5 µl Stop Ligation Buffer (STL) sono stati aggiunti alla mix previa centrifugazione a 600 x g per 5 secondi, e pipettati.

### **Purificazione frammenti ligati**

Dopo aver agitato su vortex le SPB, 42.5 µl sono stati aggiunti al campione, pipettati e incubati a temperatura ambiente per 5 minuti e posti su un rack magnetico finché la soluzione non è diventata chiara (~5 minutes). Dunque il surnatante è stato scartato ed il pellet lavato due volte mediante aggiunta di 200 µl EtOH al 80%, incubazione sul rack magnetico per 30 secondi ed eliminazione del surnatante.

Dopo aver rimosso l'etanolo residuo la provetta è stata asciugata all'aria sul rack magnetico per 5 minuti trascorsi i quali, 52.5 µl di RSB sono stati aggiunti al tubo e mixati. Dopo rimozione dal rack, miscelazione ed incubazione a temperatura ambiente per 2 minuti, il tubo è stato nuovamente posto sul rack magnetico finché la soluzione non è

diventata chiara (~ 4 minuti). 50 µl di surnatante sono stati dunque trasferiti nel corrispondente pozzetto del Clean Up ALP Plate (CAP). La procedura è stata ripetuta aggiungendo questa volta 50 µl di SPB e 22.5 µl di RSB rispettivamente. Sono stati mantenuti i volumi di EtOH.

20 µl surnatante sono stati infine trasferiti nel corrispondente pozzetto del Target Sample Plate (TSP1).

### **Validazione librerie**

Per ottenere dati di altissima qualità sulle piattaforme di sequenziamento Illumina, è importante creare densità di *cluster* ottimali su ogni corsia della cella di flusso. L'ottimizzazione della densità dei *cluster* richiede una quantifica accurata delle librerie mediante qPCR.

Il protocollo TruSeq DNA PCR-Free non prevede una fase di arricchimento per PCR in modo da selezionare i prodotti completamente ligati, pertanto frammenti che potrebbero non avere adattatori su entrambe le estremità possono essere presenti e sarebbero non clusterizzabili.

La qPCR quantifica solo le librerie con ligazione completa degli adattatori, che avranno dunque successo nella fase di cluster generation e successiva bridge amplification.

Nel presente lavoro di tesi la quantifica della libreria è avvenuta per qPCR, mediante KAPA Library Quantification Kit.

**Materiale supplementare 5.3: Mangime impiegato per l'accrescimento/ingrasso nelle aziende in cui sono stati effettuati i campionamenti**

Composizione: Orzo, Favino, Granturco, Cruschello di Frumento, polpe essiccate di Agrumi, Carrube, Melasso di Canna da zucchero, Carbonato di Calcio, Fosfato Bicalcico, Sodio Cloruro. Componenti additivati (per kg): vitamine, provitamine e sostanze con effetto analogo. (E672) Vitamina A 7.500 u.i, (E671) Vitamina D3 1.200 u.i, (3a700) Vitamina E tutto-rac-alfa-tocoferile acetate 18 mg/kg, M.S.B (Vitamina K3 50%) 1,5 18 mg/kg; Vitamina B2 2,7 mg/kg, Vitamina B6 2,1 mg/kg, Vitamina B12 0,02 mg/kg, Vitamina PP 15 mg/kg, Ac. d-Pantotenico 12 mg/kg, Ac. Folico 0,6 mg/kg, Colina 300 mg/kg. Composti di oligoelementi: E1 Solfato ferroso monoidrato 155 mg/kg, E5 Solfato manganoso monoidrato 139 mg/kg, E4 Solfato rameico pentaidrato 35 mg/kg. E5 Ossido di zinco 74 mg/kg, E2 Ioduro di potassio 0,8 mg/kg, E8 Selenito di sodio 0,4 mg/kg. Stabilizzatori della flora intestinale: 4a1600 3-fitasi 500 FTU.

## **Materiale supplementare 5.4: Preparazione Ion Xpress™ Plus gDNA Fragment Library**

### **Frammentazione enzimatica**

200 ng di gDNA purificato da gel, collezionati in 1.5-mL Eppendorf LoBind®, sono stati enzimaticamente frammentati con l'ausilio del Ion Xpress™ Plus Fragment Library Kit (Thermo Fisher Scientific), secondo quanto riportato in tabella 5.4.1. Dopo incubazione in presenza di 10 µL di Ion Shear™ Plus Enzyme Mix II (5 minuti a 37° C per 200 base-read library), 5 µL di Ion Shear™ Plus Stop Buffer sono stati aggiunti alla miscela di reazione per bloccare la digestione.

<b>Component</b>	<b>Volume</b>
gDNA,	30 µL
Ion Shear™ Plus 10X Reaction Buffer	5 µL
Nuclease-free Water	5 µL
Total	40 µL

**Tab.5.4.1:** Componenti di reazione per la frammentazione del gDNA

La frammentazione, ad opera di una mix di esonucleasi aspecifiche fornita con il kit, genera frammenti blunt-ended. Il gDNA frammentato è dunque pronto per la ligazione degli adattatori, non è richiesto end repair.

### **Purificazione del DNA frammentato**

99 µL (1.8X del volume del campione) di Agencourt® AMPure® XP Reagent (Beckman Coulter) sono stati aggiunti alla mix e lasciati in incubazione a temperatura ambiente per 5 minuti, trascorsi i quali i campioni sono stati alloggiati in un rack magnetico (DynaMag™-2, Life Technologies) per ~3 minuti, finchè la soluzione è diventata chiara e le biglie sono rimaste adese alla parete magnetica. Dopo rimozione del surnatante, avendo cura di non disturbare il pellet, sono stati aggiunti 500 µL di etanolo al 70% mantenendo il campione in incubazione nel rack per 30 secondi e scartando ogni volta il surnatante. Dopo aver effettuato un secondo lavaggio ed eliminato l'etanolo residuo, il campione è stato lasciato asciugare per 4 minuti a temperatura ambiente.

Dopo aver rimosso il tubo dal rack magnetico sono stati dispensati 25 µL di Low TE direttamente sul pellet per disperdere le biglie e, dopo aver pipettato e agitato su vortex la sospensione per 10 secondi, il tubo è stato nuovamente alloggiato nel rack magnetico fino a quando la soluzione non si è schiarita. Dopo circa 1 minuto il surnatante, contenente il

DNA eluito, è stato trasferito in una nuova provetta per PCR (0,2 ml) senza disturbare il pellet.

### **Ligazione adattatori, nick-repair e purificazione del DNA ligato**

Alla purificazione del DNA è seguita la ligazione degli adattatori con l'ausilio del kit Ion Plus Fragment Library Kit (Thermo Fisher Scientific), il nick-repair per rendere la ligazione covalente su entrambi gli strands (Ion Xpress™ Plus Fragment Library Kit, Thermo Fisher Scientific) ed un ulteriore *step* di purificazione del DNA ligato mediante il kit AMPure XP (Agencourt) ed Etanolo al 70%.

Le condizioni della reazione di ligazione (tabella 5.4.2) ed il relativo protocollo termico (tabella 5.4.3) sono riportate nelle tabelle seguenti.

<b>Component</b>	<b>Volume</b>
DNA	~25 µL
10X Ligase Buffer	10 µL
Adapters	10 µL
dNTP Mix	2 µL
Nuclease-free Water	41 µL
DNA Ligase	4 µL
Nick Repair Polymerase	8 µL
Total	100 µL

**Tab.5.4.2:** Reagenti di ligazione e relativo volume.

<b>Stage</b>	<b>Temperature</b>	<b>Time</b>
Hold	25° C	15 min
Hold	72° C	5 min
Hold	4° C	Hold

**Tab.5.4.3:** Protocollo termico per la reazione di ligazione degli adattatori

## **Materiale supplementare 5.5: quantifica delle libraries: Agilent 2100 Bioanalyzer, Qubit 2.0 Fluorometer e qPCR**

Un aspetto fondamentale del protocollo di preparazione delle library è l'utilizzo di una precisa concentrazione di templatò nella fase di amplificazione clonale attraverso Emulsion PCR, per l'influenza che questa riveste nella resa ed efficienza della corsa di sequenziamento. Una concentrazione troppo elevata genererebbe infatti policlonalità, che produce dati non analizzabili; al contrario, una concentrazione troppo bassa può causare la presenza di segnali troppo deboli e un'imprecisa rilevazione da parte dello strumento. Nel presente lavoro di tesi le fragment libraries, costruite a partire dalle 2 RRLs, sono state quantificate con tre diversi metodi analitici e con l'ausilio di tre strumenti al fine di assicurare una quantifica accurata, come raccomandato nei protocolli Thermo Fisher Scientific: 2100 Bioanalyzer Instrument (Agilent), Qubit 2.0 Fluorometer (Life Technologies) e Real-Time PCR System (Applied Biosystems StepOne™). La determinazione del fattore di diluizione per la normalizzazione delle librerie alla concentrazione di 100 pM è avvenuta mediando i risultati delle tre misurazioni.

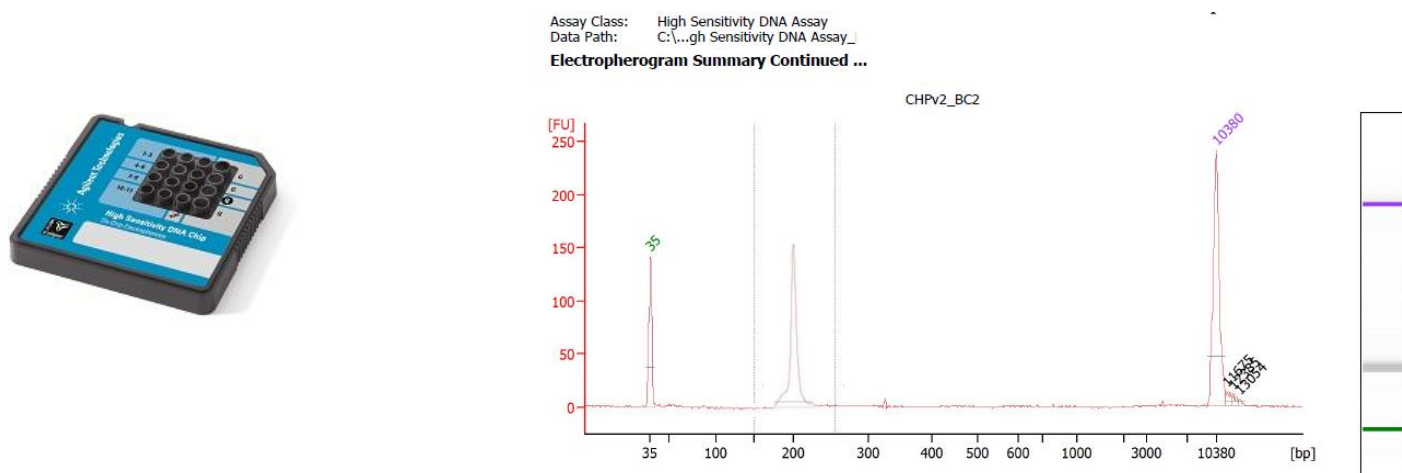
### **5.5.1 Agilent® 2100 Bioanalyzer®**

Lo strumento Agilent 2100 Bioanalyzer permette l'ottenimento di informazioni circa la dimensione e quantifica di biomolecole dotate di carica, sfruttando i principi dell'elettroforesi e le tecnologie di microfluidica. Il chip (figura 5.5.1a) è costituito da 16 pozzetti (11 per i campioni, 4 per il gel, 1 per il ladder che funge da standard esterno) interconnessi tra loro tramite una rete di micro-canali in vetro. Durante la preparazione del chip, i micro-canali sono riempiti con una matrice polisaccaridica e un colorante a fluorescenza. Una volta riempiti i pozzetti e i canali, il chip diventa un circuito elettrico integrato e le molecole dotate di carica sono sottoposte a differenza di potenziale che ne determina la migrazione. L'emissione della fluorescenza indotta da un laser, determinata dall'intercalazione delle molecole fluorescenti tra i dsDNA, è rilevata da un detector.

I dati sono poi elaborati dal software dello strumento in mappe densitometriche, simili a gel, e in elettroferogrammi (figura 5.5.1b). Con l'ausilio di un ladder che contiene frammenti di dimensioni e concentrazioni note, viene costruita una curva standard del tempo di migrazione rispetto alla lunghezza dei frammenti. Dal tempo di migrazione misurato per ogni frammento nel campione, se ne calcola la size. Due marcatori (low- e high-molecular weight) vengono fatti co-migrare con ciascuno dei campioni e

rappresentano degli standard interni utilizzati per allineare i dati ladder con i dati dei pozzetti dei campioni.

Nel presente lavoro di tesi, 1  $\mu\text{L}$  di ciascuna libreria amplificata è stato analizzato mediante Agilent® Bioanalyzer® con il kit Agilent® High Sensitivity DNA al fine di determinarne la concentrazione molare. La molarità è stata calcolata automaticamente in  $\text{pmol} / \text{L}$  (pM) con l'ausilio del 2100 Expert Software version B.02.08.



**Fig. 5.5.1a** – Chip High Sensitivity

**Fig. 5.5.1b**- Output Bioanalyzer®. Elettroferogramma degli ampliconi della fragment library (pool 1, 200-base-read library) e relativa mappa densitometrica. I picchi a 35 e 10380 bp rappresentano low- e high-molecular weight markers.

## Preparazione e caricamento Gel-Dye Mix

Dopo aver riequilibrato i reagenti a temperatura ambiente, 15  $\mu\text{L}$  High Sensitivity DNA dye concentrate sono stati miscelati alla High Sensitivity DNA gel matrix in una vial, agitati su vortex e centrifugati a  $2240 \text{ g} \pm 20 \%$  (6000 rpm per Eppendorf microcentrifuge). Il caricamento del gel su un High Sensitivity DNA chip è avvenuto in accordo con il protocollo fornito da Agilent. Brevemente, dopo aver alloggiato il chip nella chip priming station, 9  $\mu\text{L}$  di Gel-Dye Mix sono stati pipettati sui pozzetti marcati G (figura 5.5.1a). La posizione marcata in nero è stata pressurizzata per 60 sec. 6  $\mu\text{L}$  di marker sono stati caricati in tutti i pozzetti rimanenti (solo 5  $\mu\text{L}$  di marker nei pozzetti che avrebbero alloggiato i campioni, ai quali è stato aggiunto 1  $\mu\text{L}$  di libreria), 1  $\mu\text{L}$  di ladder è stato pipettato nel pozzetto identificato dal relativo simbolo. Infine il chip è stato agitato su vortex per 60 secondi a 2400 rpm, su un IKA vortex mixer (Agilent) e alloggiato nello strumento per

l'avvio della corsa, al termine della quale lo strumento ha restituito in output la concentrazione delle libraries.

### **5.5.2. Qubit® 2.0 Fluorometer**

Questo tipo di quantifica permette di ottenere la concentrazione dei campioni utilizzando un principio diverso da quello impiegato in spettrofotometria. Il Qubit™ dsDNA HS Assay Kit è infatti ottimizzato per la quantifica di DNA a doppio filamento (dsDNA), mediante l'utilizzo di un fluoroforo intercalante degli acidi nucleici che si lega specificatamente al dsDNA.

Lo svantaggio principale delle spettroscopie a UV per la quantificazione del dsDNA è rappresentato infatti dal contributo di segnali provenienti da DNA a singolo filamento (ssDNA) che può quindi determinare la sovrastima del campione, e altri contaminanti, come proteine e buffer di estrazione.

Nel presente lavoro di tesi, 10 µL di ciascuna libreria amplificata sono stati quantificati con l'ausilio del kit Qubit® dsDNA HS Assay, secondo il protocollo fornito dalla Ditta produttrice come di seguito riportato.

#### **Calibrazione Qubit® 2.0 Fluorometer e determinazione della concentrazione delle librerie amplificate**

In un 1,5 mL DNA Lobind Tube sono stati mixati 199 µL di Qubit™ dsDNA HS buffer e 1 µL di Qubit™ dsDNA HS reagent per la preparazione della Qubit™ working solution. Per le esigenze sperimentali del presente lavoro di tesi sono stati dunque miscelati 796 µL di buffer e 4 µL di dsDNA HS reagent.

Per la preparazione dei Qubit™ standard (due standard sono richiesti), 190 µL di Qubit™ working solution sono stati miscelati con 10 µL di ciascun Qubit™ dsDNA HS standard, agitati su vortex per 3 secondi e incubati a temperatura ambiente per 2 minuti. Infine, 190 µL di Qubit™ working solution sono stati miscelati con 10 µL di ciascuna libreria, agitati su vortex per 3 secondi e incubati a temperatura ambiente per 2 minuti. Dopo aver effettuato la calibrazione dello strumento e aver proceduto alla lettura dei campioni, lo strumento ha restituito in output la concentrazione delle libraries.



### 5.5.3. qPCR

La PCR quantitativa (qPCR) è in grado di determinare il quantitativo di template amplificabile in una Ion Library, mediante l'ausilio del Ion Library Quantitation Kit (Life Technologies) con un elevato livello di specificità ed accuratezza anche a partire da bassissime concentrazioni di DNA iniziale. Questo approccio permette dunque la quantifica delle library anche quando il materiale input di partenza è insufficiente per la rilevazione con altri strumenti, ad esempio il Qubit 2.0 Fluorometer o l'Agilent 2100 Bioanalyzer.

La qPCR è stata condotta sul termociclatore Applied Biosystems StepOne Plus™ seguendo il protocollo fornito dalla ditta produttrice (Life Technologies). I Reagenti impiegati (Ion Library TaqMan® Quantitation Kit) consistono in due primer e in una sonda TaqMan®. I primer sono disegnati per l'amplificazione del target mentre la sonda è disegnata per l'ibridazione del target e per la generazione del segnale di fluorescenza quando questo viene amplificato.

I reagenti TaqMan utilizzano quindi una sonda fluorogenica per l'abilitazione al rilevamento di uno specifico prodotto della PCR accumulatosi durante i cicli PCR.

Inizialmente sono state preparate le diluizioni seriali di *E. coli* DH10B Control Library, come di seguito riportato:

Standard	Control Library	Nuclease-free water	Fold dilution	Concentration
1	5 µL (undiluted)	45 µL	0.1	6.8 pM
2	5 µL Std 1	45 µL	0.01	0.68 pM
3	5 µL Std 2	45 µL	0.001	0.068 pM
4	5 µL Std 3	45 µL	0.0001	0.0068 pM
5	5 µL Std 4	45 µL	0.00001	0.00068 pM

Diluizioni seriali di *E. coli* DH10B Control Library

Tali diluizioni rappresentano i riferimenti a concentrazione nota per la determinazione della quantità di molecole di un campione ignoto, mediante costruzione di una curva standard, che è ottenuta dall'interpolazione dei valori di fluorescenza misurata nei campioni a concentrazione nota.

Anche la Library deve essere diluita prima di essere processata, come riportato in tabella 5.5.3a. Le condizioni di reazione ed il protocollo termico sono riportati nelle tabelle 5.5.3b e 5.5.3c rispettivamente.

Dilution	Library input	Nuclease-free water
1:2000	1 $\mu$ L of 1:20	99 $\mu$ L
1:20000	5 $\mu$ L of 1:2000	45 $\mu$ L

Tab. 5.5.3a: Diluizioni Library per qPCR

Component	Volume for 20 $\mu$ L reaction
Ion Library TaqMan® qPCR Mix, 2X	10 $\mu$ L
Ion Library TaqMan® Quantitatio Assay, 20X	1 $\mu$ L
Nuclease-free water	To 15 $\mu$ L

Tab. 5.5.3b: Componenti di reazione e relativi volumi

Temperature	Time
95° C	1 sec
60° C	20 sec

Tab. 5.5.3c: Protocollo termico qPCR, 40 cicli

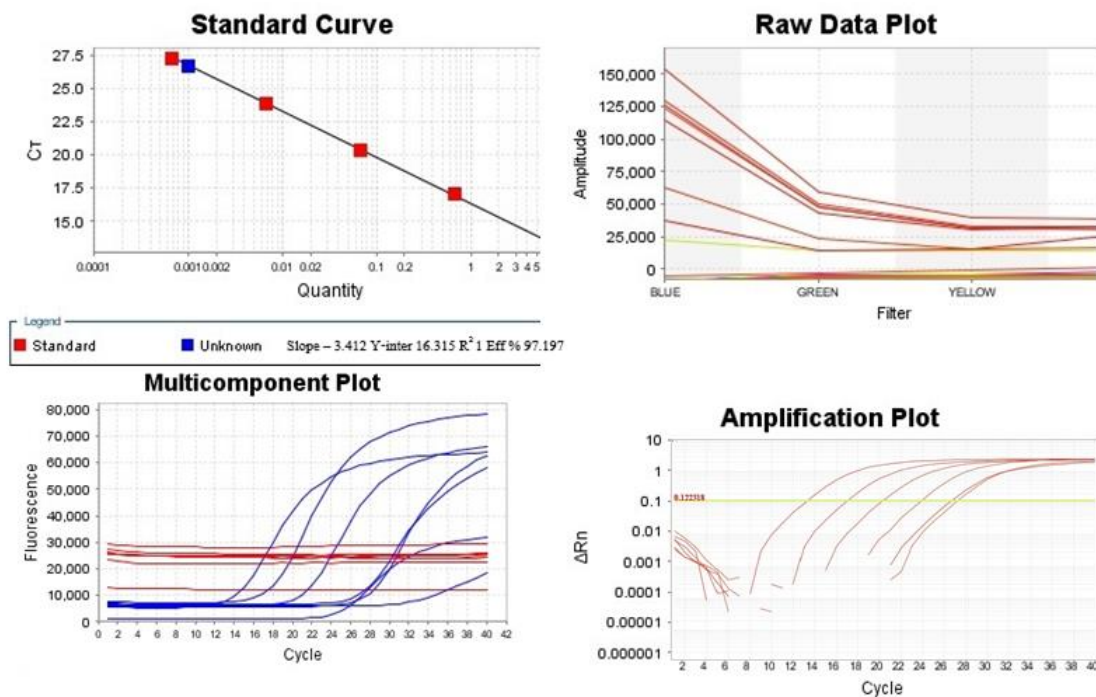


Fig. 5.5.3a: StepOnePlus™ Real-Time PCR System output per pool 1-library

Tale saggio TaqMan, essendo disegnato specificatamente per la quantifica delle library Ion Torrent (primer disegnati contro le sequenze degli adattatori), risulta essere il metodo analitico più performante in quanto in grado di quantificare solo i prodotti legati agli adattatori e quindi i soli sequenziabili. Ai fini del presente lavoro di tesi il calcolo del TDF (Template Dilution Factor), per ciascuna libreria, è stato eseguito mediando i risultati delle tre misurazioni.

### **Materiale supplementare 5.6: preparazione e arricchimento delle template-positive Ion PGM™ Hi-Q™ Ion Sphere™, sequenziamento su Ion PGM™**

Dopo aver quantificato e calcolato i TDF (Template Dilution Factor, sulla media delle letture di quantifica) per normalizzare a 100 pM le due Libraries, queste sono state diluite alla concentrazione finale di 26 pM, come indicato per le gDNA Fragment Libraries dal protocollo fornito dalla ditta produttrice. In questa fase lo scopo è di massimizzare gli eventi monoclonali rispetto i policlonali, intervenendo sulla concentrazione delle librerie (le ISPs e le molecole di DNA sono distribuite secondo la legge di Poisson, la quantità delle ISPs è fissa). Dopo aver equalizzato alla stessa concentrazione molare le due libraries, la fase di preparazione ed arricchimento del template è avvenuta separatamente per ciascuna libreria come di seguito riportato. 25 µl di ciascuna libreria diluita è stata addizionata alla rispettiva soluzione di amplificazione che rappresenta la fase acquosa per la emPCR. Brevemente, in ciascun tubo da 2 ml contenente 800 µl di Ion PGM™ Hi-Q™ Reagent Mix, sono stati addizionati 25 µl di Nuclease-free Water, 50 µl Ion PGM™ Hi-Q™ Enzyme Mix, 25 µl delle libraries diluite e 100 µl di Ion PGM™ Hi-Q™ ISPs (Ion Sphere™ Particles) precedentemente agitate su vortex per 1 minuto, per un volume finale totale di 1000 µl. La reazione è stata allestita in due distinti tubi, ciascuno per ogni library. La soluzione di amplificazione così preparata (l'intero volume) è stata poi agitata su vortex e caricata attraverso la sample port del Ion OneTouch™ Reaction Filter. Dalla stessa sample port sono stati caricati, in due *step* successivi, 850 µl di Ion OneTouch™ Reaction Oil. Il filtro è stato poi montato sul Ion OneTouch™2 Instrument per la reazione di emPCR, al termine della quale tutte le sfere (monoclonali e policlonali, vuote e non template) sono state recuperate in 150 µl Ion OneTouch™ Breaking Solution, contenente SDS concentrata che ha lo scopo di dissolvere l'emulsione, in due Ion OneTouch™ Recovery Tubes. Da ciascuna provetta è stato scartato il surnatante e mantenuti 100 µl in cui è stato risospeso il pellet, contenente le ISPs. A ciascun recovery tube sono stati aggiunti 500 µL di Ion OneTouch™ Wash Solution e ogni sospensione è stata trasferita in un unico tubo Eppendorf LoBind®, centrifugato per 2,5 minuti a  $15.500 \times g$ . Il surnatante è stato rimosso, mantenendo solo 100 µl di wash solution, avendo cura di non disturbare il pellet. Un'aliquota di 2 µL delle ISPs non ancora arricchite è stata destinata alla valutazione di qualità mediante Qubit® 2.0 Fluorometer, mentre l'intero volume residuo è stato trasferito nel pozzetto n. 1 del 8-well strip (Ion OneTouch™ ES Supplies Kit) per l'arricchimento

delle template-positive ISPs con Ion OneTouch™ ES. Questa fase ha lo scopo di collezionare tutte le sfere template rimuovendo quelle non template e di rendere single strand la popolazione dei frammenti templati mediante denaturazione basica. Le sfere template sono biotinilate (perché il primer A è biotinilato) e dunque trattenute da Ion OneTouch™ ES che sfrutta le biglie paramagnetiche streptavidinate (Dynabeads® MyOne™ Streptavidin C1 Beads, Thermo Fisher Scientific). Queste microsfeere, del diametro di 1 µm, presentano un monostrato di streptavidina ricombinante accoppiata covalentemente alla superficie con siti di legame alla biotina stericamente disponibili per il legame, non solo della biotina libera, ma anche per il legame di target biotinilati.

Al termine dell'arricchimento il campione includerà dunque solo le sfere template, sia policlonali che monoclonali.

Si riporta di seguito la preparazione dei reagenti ed il caricamento del 8-well strip.

Il protocollo è stato eseguito allo stesso modo per le due libraries.

### **Preparazione Melt-Off Solution**

<b>Component</b>	<b>Volume</b>
Tween® Solution	280 µL
1 M NaOH	40 µL
Total	320 µL

### **Preparazione Dynabeads® MyOne™ Streptavidin C1 Beads**

Al fine di risospendere completamente le microsfeere, il tubo contenente Dynabeads® MyOne™ Streptavidin C1 Beads è stato agitato su vortex e centrifugato rispettivamente per 30 e per 2 secondi. Dopo aver pipettato la sospensione per disperdere il pellet, 13 µL di Dynabeads® MyOne™ Streptavidin C1 Beads sono stati trasferiti in un nuovo tubo Eppendorf LoBind® da 1,5 ml posizionato su un magnete DynaMag™ -2 e ivi mantenuti per 2 minuti, trascorsi i quali il surnatante è stato scartato avendo cura di non disturbare il pellet. Dopo aver aggiunto 130 µL di MyOne™ Beads Wash Solution, il tubo è stato rimosso dal magnete, agitato su vortex 30 secondi e centrifugato per 2 secondi.

## Caricamento 8-well strip

Well number	Reagent to dispense in well
Well 1	Intero volume (100 µl) template-positive ISPs
Well 2	130 µL di Dynabeads® MyOne™ Streptavidin C1 Beads risospese in MyOne™ Beads Wash Solution
Well 3	300 µL di Ion OneTouch™ Wash Solution
Well 4	300 µL di Ion OneTouch™ Wash Solution
Well 5	300 µL di Ion OneTouch™ Wash Solution
Well 6	-
Well 7	300 µL di Melt- Off Solution
Well 8	-

10 µL di Neutralization Solution sono stati aggiunti ad un tubo da PCR 0.2-mL destinato a collezionare le ISPs arricchite, alloggiato poi nella rispettiva posizione (alla base del Tip Loader, Ion OneTouch™ ES).

Dopo l'esecuzione della corsa, durata 35 minuti, al tubo da PCR 0.2-mL contenente le ISPs arricchite sono stati addizionati 5 µL di Control Ion Sphere™ Particles (Ion PGM™ Controls Kit v2). Il tubo è stato poi centrifugato per 2 minuti a  $15,500 \times g$ , il surnatante scartato avendo cura di non disturbare il pellet, mantenendo nel tubo 15 µL.

Dopo aver effettuato i lavaggi, inizializzato lo strumento ed effettuato il Chip Check (*Ion PGM™ Hi-Q™ Sequencing Kit User Guide*) 12 µL di Sequencing Primer e 2 µL di Annealing Buffer (Ion PGM™ Hi-Q™ Sequencing Kit) sono stati addizionati al tubo contenente le ISPs, poi processato a 95°C per 2 minuti seguiti da un ciclo termico a 37°C per 2 minuti. Dopo aver rimosso le ISPs dal termociclatore, 3 µL di Ion PGM™ Hi-Q™ Sequencing Polymerase sono stati aggiunti al tubino lasciato poi in incubazione per 5 minuti a T° ambiente, al termine dei quali 30 µL di ISPs sono stati caricati su un Ion 318™ Chip V2 BC.

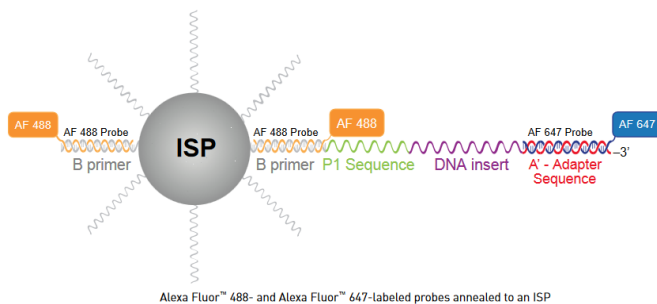
La metodica è stata eseguita allo stesso modo per le due libraries.

### **Materiale supplementare 5.7: valutazione dell'efficienza di templaggio con il kit Ion Sphere™ Quality Control e il Qubit® 2.0 Fluorometer**

Al termine della fase di amplificazione clonale un'aliquota delle ISPs non arricchite, in uscita da Ion OneTouch™2 Instrument, è stata utilizzata per valutarne l'efficienza di templaggio. A 2 µL di ciascuna libreria sono stati aggiunti 19 µL di Annealing Buffer e 1 µL di Ion Probes (Ion Sphere™ Quality Control Kit) e, dopo aver agitato su vortex, i due tubini sono stati denaturati a 95°C per 2 minuti e sottoposti ad un ciclo termico di 37°C per 2 minuti per far avvenire l'annealing delle sonde. Al termine del protocollo termico sono stati eseguiti tre lavaggi consecutivi mediante aggiunta di 200 µL di Wash Buffer, agitazione su vortex e centrifuga a 15500 x g per 1,5 minuti. Ad ogni *step* di lavaggio il surnatante è stato scartato, avendo cura di non disturbare il pellet di ISPs, e sono stati mantenuti nel tubino 10 µL. Infine 190 µL di Wash Buffer sono stati aggiunti alle due librerie, agitati su vortex e trasferiti in due Qubit™ assay tubes.

Il controllo negativo è stato preparato aggiungendo 200 µL di Quality Control Wash Buffer in un nuovo Qubit™ Assay Tube.

Il dosaggio Ion Sphere™ consente una valutazione rapida delle ISPs prima di performare la corsa su PGM, misurando la fluorescenza emessa dalle template-positive ISPs marcate con due fluorofori: Alexa Fluor™ 488 e Alexa Fluor™ 647. La sonda marcata con AF 488 si annila al sito B e dunque a tutte le ISPs presenti, mentre la sonda marcata con AF 647 al sito A (figura 5.7.1) e dunque alle sole ISPs con template estesi. Il ratio tra la fluorescenza emessa da AF 647 (ISPs template) e la fluorescenza emessa da AF 488 (tutte le ISPs presenti) permette di stimare la percentuale di sfere template tenendo conto di un fattore di calibrazione, variabile in funzione dello strumento utilizzato per la determinazione e di un fattore di conversione variabile in funzione della size della libreria, del kit e dei reagenti impiegati per il dosaggio, calcolati con l'ausilio del tool on line Qubit® Conversion Factors ([www.thermofisher.com](http://www.thermofisher.com)).



Alexa Fluor™ 488- and Alexa Fluor™ 647-labeled probes annealed to an ISP

**Fig. 5.7.1:** Probes Alexa Fluor™ 488 e Alexa Fluor™ 647 coniugati alle ISPs (www. Thermofisher.com)

### Misurazione calibration standard e ISPs

Dopo aver effettuato la taratura dello strumento contro il controllo negativo (bianco) si è proceduto alla preparazione degli standard di calibrazione. 200 µL di ciascun Alexa Fluor™ 488 e Alexa Fluor™ 647 Calibration Standard reagents (Ion Sphere™ Quality Control Kit) sono stati trasferiti in due separati Qubit™ assay tubes e sottoposti a lettura della fluorescenza su Qubit® 2.0 Fluorometer. Per ciascuna lettura è stato registrato il valore di RFU (raw fluorescence) nel Qubit™ Easy Calculator Microsoft™ Excel™ Spreadsheet e ottenuto il fattore di calibrazione specifico dello strumento. Si è proceduto poi alla lettura di ciascun campione per stimare la percentuale di ISPs template, risultata nel range atteso (10%-30%). Si è proceduto dunque al sequenziamento delle due librerie.



**Materiale supplementare 5.8: Genotyping mediante sequenziamento Sanger degli SNPs localizzati in geni annotati, che mostravano differenze significative nelle due RRLs.**

Il gDNA individuale dei 192 soggetti campionati all'atto della macellazione, compresi quelli inclusi nella costruzione dei due pool, è stato amplificato in sei differenti reazioni di PCR, in un 2720 Thermal Cycler (Applied Biosystems, Foster City – USA). Le amplificazioni sono state performati in 30 µl di volume. Le coppie di primer utilizzate, progettate con il software Primer 3 (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>), e le condizioni di reazione sono riportati in tabella 5.8.2.

Il *locus MC4R* è stato amplificato nella regione *downstream*, a 49 bp di distanza dal trascritto. La mix di reazione comprendeva ~ 50 ng/µl di gDNA, 1 x PCR buffer, 2 mM di MgCl<sub>2</sub>, 0.5 U di Taq polymerase (Bioline), 5% di DMSO, 10 pmol di ciascun primer (Eurofins) e 200 µM di ciascun dNTPs (Promega).

Il *locus FASN* è stato amplificato nella regione *upstream*, a 4558 bp di distanza dal trascritto. La mix di reazione comprendeva ~ 50 ng/µl di gDNA, 1 x PCR buffer, 3 mM di MgCl<sub>2</sub>, 50mM KCl, 2.5 U di Taq polymerase (Bioline), 10 pmol di ciascun primer (Eurofins) e 200 µM di ciascun dNTPs (Promega).

Il *locus FFAR4* è stato amplificato nella regione 5' UTR. La mix di reazione comprendeva ~ 50 ng/µl di gDNA, 1 x PCR buffer, 25 mM di MgCl<sub>2</sub>, 2.5 U di Taq polymerase (Bioline), 10 pmol di ciascun primer (Eurofins) e 200 µM di ciascun dNTPs (Promega).

Il *locus SCD* è stato amplificato nella regione *upstream*, a 4637 bp di distanza dal trascritto. La mix di reazione comprendeva ~ 50 ng/µl di gDNA, 1 x PCR buffer, 3 mM di MgCl<sub>2</sub>, 10 mM TRIS HCl (pH 9.0), 1 U di Taq polymerase (Bioline), 10 pmol di ciascun primer (Eurofins) e 200 µM di ciascun dNTPs (Promega).

Il *locus LEP* è stato amplificato nella regione del II introne. La mix di reazione comprendeva ~ 50 ng/µl di gDNA, 1 x PCR buffer, 25 mM di MgCl<sub>2</sub>, 2.5 U di Taq polymerase (Bioline), 10 pmol di ciascun primer (Eurofins) e 200 µM di ciascun dNTPs (Promega).

Il locus *LEPR* è stato amplificato nella regione del III introne. La mix di reazione comprendeva ~ 50 ng/μl di gDNA, 1 x PCR buffer, 25 mM di MgCl<sub>2</sub>, 2.5 U di Taq polymerase (Bioline), 10 pmol di ciascun primer (Eurofins) e 200 μM di ciascun dNTPs (Promega).

I prodotti di amplificazione sono stati visualizzati su gel di agarosio al 2% e purificati con Wizard® SV Gel and PCR Clean-Up System (Promega), allo scopo di eliminare le porzioni di acidi nucleici non coinvolte nell'amplificazione. Il purificato è stato quindi risospeso in 20μl di acqua.

Ciascuna delle regioni genomiche in esame è stata sottoposta ad una amplificazione di sequenza impiegando un kit di dideossinucleotidi trifosfato marcati con fluorofori (BigDye Terminator v3.1 Ready Reaction, ABI PRISM®), seguendo il protocollo riportato in tabella 5.8.1.

Component	Volume (μl)
dH <sub>2</sub> O	14.5
Ready MIX	2
PCR Buffer (10X)	2
Primer singolo (F o R)	0.5
PCR purificata	1

**Tab. 5.8.1:** Componenti e volume di reazione

La reazione, avvenuta con l'ausilio di un termociclatore 2720 Thermal Cycler (Applied Biosystems, Foster City – USA), è stata effettuata secondo i seguenti cicli termici: 96° C x 1', 96° C x 10'', 50° C x 5'', 60° C x 4'.

Le catene di terminazione ottenute sono state purificate con l'impiego di un kit per la rimozione dei Dye in eccesso (Dye Terminator Removal Kit, CentriSep™), che sfrutta le proprietà filtranti della resina SEPHADEX. 10 μl di prodotto sono stati denaturati per 2 minuti a 95°C con 10μl di Formamide e sequenziati. Il sequenziamento è stato effettuato con il metodo a terminazione di catena (Sanger et al., 1977a) per mezzo del sequenziatore automatico capillare ABI PRISM® 3500 (Applied Biosystems, Foster City – USA).

Le sequenze prodotte, visualizzate mediante il software Sequencing Analysis (Applied Biosystems), sono state allineate con l'ausilio del software CLUSTALX 1.8 (Thompson et al., 1997).

GENE	Position	Acc. Num.	Primers (5'-3')	Amplification Reaction	Amplicon size
MC4R	1:160774173	<a href="#">397359</a>	Fw-TGATTCTAGCATTGAGGGGA Rev-ACCTAAGAGCATGGCAAGCT	Cycle 1: 94°C x 30'' Cycle 2-30: 94°C x 1'; <b>58°C x 2'</b> ; 72°C x 3'' Cycle 31: 72°C x 5'	316
LEPR	6: 146893537	<a href="#">396836</a>	Fw-CCAAGAAAAAACAACACTCGTGT Rev-TTAAATCAGAAAGGCACAAA	Cycle 1: 96°C x 30'' Cycle 2-30: 94°C x 30''; <b>58.5°C x 30''</b> ; 72°C x 40'' Cycle 31: 72°C x 3'	300
LEP	18: 20123273	<a href="#">396832</a>	Fw-CAGTTGCGCAAGTTGCGCTC Rev-CAGGAGACTTCCAGAGCCGC	Cycle 1-34: 95°C x 1'; <b>61°C x 1'</b> ; 72°C x 1'	358
FFAR4	14: 105011727	<a href="#">100135678</a>	Fw-ATCAATTTTAGCTAAATC Rev-AGAGATGTCGCAACCGCTT	Cycle 1: 96°C x 30'' Cycle 2-30: 94°C x 30''; <b>60°C x 30''</b> ; 72°C x 1' Cycle 31: 72°C x 6'	304
SCD	14: 111461097	<a href="#">396670</a>	Fw- AGTTCAGCAGTACGCATCTAT Rev- AAACTTCCCTAGTGCCCATCCTTT	Cycle 1: 96°C x 30'' Cycle 2-30: 94°C x 1'; <b>60°C x 2''</b> ; 72°C x 2' Cycle 31: 72°C x 5'	275
FASN	12: 921860	<a href="#">397561</a>	Fw-ACCAGGAGTGTTGTCCAGGT Rev-TTCTCTTAACAAGCTGAC	Cycle 1: 96°C x 30'' Cycle 2-30: 94°C x 30''; <b>58°C x 30''</b> ; 72°C x 40'' Cycle 31: 72°C x 5'	350

**Tab. 5.8.2:** Accession number, primers Fw+Rev, protocollo termico per le reazioni di amplificazione e size del prodotto di PCR (bp), per ciascuno dei polimorfismi analizzati

## 6. Allegati

### 6.1 Whole genome SNPs discovery and analysis of genetic diversity in Nero Siciliano pig

Domenico Giosa<sup>1\*</sup>, Alessandro Zumbo<sup>2</sup>, Irene Sapienza<sup>2</sup>, Letterio Giuffrè<sup>2</sup>, Riccardo Aiese Cigliano<sup>3</sup>, Orazio Romeo<sup>1</sup>, Enrico D'Alessandro<sup>2</sup>.

<sup>1</sup> Department of Chemical, Biological, Pharmaceutical and Environmental Sciences, University of Messina, Italy.

<sup>2</sup> Department of Veterinary Sciences, Unit of Animal Production, University of Messina, Italy.

<sup>3</sup> Sequentia Biotech SL, Calle Comte D'Urgell 240, 08036, Barcelona, Spain.

\*Corresponding author: E-mail: domenico.giosa@gmail.com

#### Abstract

Nero Siciliano is a local pig breed reared under extensive or semi-extensive management mainly in the Nebrodi mountains of Sicily (Italy). This breed has been rediscovered by farmers who have established a consortium with the aim to protect, exploit and characterize the products obtained from it.

In order to discover typical Nero Siciliano signatures, here we present a whole genome SNPs discovery investigation against the *Sus Scrofa* reference genome (SScrofa 11.1). Genomic DNA was extracted from leukocytes of fresh whole blood sample from a male of Nero Siciliano pig with Wizard® Genomic DNA Purification Kit (Promega Corporation), then quantified with Qubit 2.0 Fluorometer (ThermoFisher). DNA quality was evaluated by Nanophotometer P-330 (Implen GmbH) and 1µg di gDNA was used for TruSeq DNA PCR-Free Library preparation (insert size 350bp) following the protocol provided by Illumina.

Illumina HiSeq X Ten sequencing produced 346.832.803 paired-end raw reads and after cleaning, performed with Trimmomatic (v.0.36), 346.573.870 (99,92%) good quality reads (phred score  $\geq 30$ ) were mapped using bwa aligner against the reference genome (Genbank: GCA\_000003025.6). SNPs discovery was performed using SUPER-CAP and the potential effects of detected mutations were evaluated by SnpEff software (v4\_3m\_core).

Analysis resulting in 9.302.670 SNPs, 1.336.511 insertions and 614.764 deletions. More than 5 million of those SNPs were alternative homozygous. Furthermore, over 40% of intronic variants, ~38% of non coding transcript variants and ~8% of intergenic variants were identified. The estimated variant rate was of 1 variant every 222 bases, showing a very high degree of variability among studied breeds.

**90° Convegno della Società Italiana di Biologia Sperimentale (SIBS),**

"Biologia sperimentale nella ricerca di base e applicata all'ambiente e all'uomo".

27-28/10/2017, Università degli Studi di Palermo-Polo Territoriale Universitario di  
Trapani

## 6.2 *In silico* comparison of meat quality candidate genes among Nero Siciliano, Duroc, Large White and Landrace genomes

Sapienza Irene<sup>1\*</sup>, Giosa Domenico<sup>2</sup>, Giuffrè Letterio<sup>1</sup>, Romeo Orazio<sup>2</sup>, Chiofalo Vincenzo<sup>2</sup>, Zumbo Alessandro<sup>1</sup>, D'Alessandro Enrico<sup>1</sup>

<sup>1</sup>Dipartimento di Scienze Veterinarie, Unità di Produzione Animale, Università di Messina, IT, [isapienza@unime.it](mailto:isapienza@unime.it), [lgiuffre@unime.it](mailto:lgiuffre@unime.it), [zumbo@unime.it](mailto:zumbo@unime.it), [edalessandro@unime.it](mailto:edalessandro@unime.it),

<sup>3</sup>Dipartimento di Scienze Chimiche, Biologiche, Farmaceutiche ed Ambientali, Università di Messina, IT, [dgiosa@unime.it](mailto:dgiosa@unime.it), [oromeo@unime.it](mailto:oromeo@unime.it), [vincenzo.chiofalo@unime.it](mailto:vincenzo.chiofalo@unime.it)

\*Corresponding author: E-mail: [isapienza@unime.it](mailto:isapienza@unime.it)

### Abstract

In recent decades, many local breeds have been subjected to genetic erosion and loss of biodiversity resulting in the impoverishment of a precious gene pool that has mainly affected marginal areas and low input breeding systems.

Nero Siciliano pig is a local breed reared under semi-extensive system mainly in the Nebrodi Mountains of Sicily (Italy). In 2003 was established a Consortium for the valorization of its productions and a request to label the fresh Nero Siciliano meat with the Protected Denomination of Origin (PDO) was issued in 2005. The request for the PDO has been started also for Nero Siciliano's cured ham in 2011.

In this study we report an *in silico* comparison of 48 candidate genes involved in meat quality traits, retrieved by using Nero Siciliano (NS), Large White (LW), Landrace (LAN) and Duroc (DU) genomes. The latter is the reference genome for pig (*S. scrofa*11.1).

In particular, we focused on genes related to muscle mass deposition and carcass fatness as these traits influence technological processes adopted for long matured pork meats products such as cured ham. The reads, downloaded from NCBI Sequence Read Archive (SRA) database, checked using FastQC program and cleaned with Trimmomatic v. 0.36, were first mapped against the *S. scrofa* reference genome (Genbank: GCA\_000003025.6) using the BWA aligner and then used for variant calling analysis. The reads of NS were obtained by an our recent study (unpublished data, manuscript in review) and used for the genome-wide analysis carried out to identify variants potentially breed-related. Genomic DNA (gDNA) was extracted from a blood sample of a male Nero Siciliano pig using the Wizard® Genomic DNA Purification Kit (Promega Corporation, Italy) and quantified by Qubit 2.0 Fluorometer with Qubit dsDNA HS Assay Kit (Thermo Fisher, Italy). PCR-Free Library preparation was performed using the TruSeq DNA kit (insert size 350 bp) and 1 µg

of gDNA, following the protocol provided by Illumina. Then, paired-sequencing was carried out using the HiSeqX platform (Illumina®). SNPs and short INDELs discovery in 48 genes related to lipid metabolism and fat deposition was performed using SUPER-CAP and the potential effects of resulting mutations were evaluated by SnpEff software (v4\_3m\_core).

More than twenty thousand variants were identified by comparing the gene set of each breed to the reference genome assembly. Of these ~22,000 were SNPs, ~3,000 short insertions and ~1,400 short deletions. Transitions / transversions ratio was 2.650 while missense/silent ratio resulting in 0.526. Furthermore, over 40% of intronic variants and ~45% of non coding transcript variants were also identified.

Among all variants detected in this study, more than 3,000 were shared among NS, LW and LAN while ~7,000 were unique for NS, ~2,000 for LW and ~6,000 for LAN, showing a high degree of genetic variability among studied breeds.

**XII° Convegno Nazionale “Biodiversità, Ambienti, Salute”.**

13-14-15/06 2018, Campus A. Saliceti, Università degli Studi di Teramo.

## 7. Bibliografia

Aho A. V., Weinberger B. W., Keringhan P. J. (1979): Awk - A Pattern Scanning and Processing Language. *SOFTWARE-PRACTICE AND EXPERIENCE*, VOL. 9. 261-279.

Albers C.A., Lunter G., MacArthur D.G., McVean G., Ouwehand W.H., et al. (2011): Dindel: accurate indel calls from short-read data. *Genome Research* 21: 961–973.

Albrechtsen A., Nielsen F.C., Nielsen R. (2010): Ascertainment biases in SNP chips affect measures of population divergence. *Mol Biol Evol*, 27:2534–2547.

Altshuler D. *et al.* (2000): An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature* 407, 513–516.

Amaral A. J. *et al.* (2009): Application of massive parallel sequencing to whole genome SNP discovery in the porcine genome. *BMC Genomics* 10, 374.

Amaral A. J. *et al.* (2011): Genome-wide footprints of pig domestication and selection revealed through massive parallel sequencing of pooled DNA. *PLoS ONE* 6, e14782.

Amills M., Villalba D., Tor M., Mercadé A., Gallardo D., Cabrera B., Jiménez N., Noguera J.L., Sánchez A. & Estany J. (2008): Plasma leptin levels in pigs with different leptin and leptin receptor genotypes. *Journal of Animal Breeding and Genetics* 15, 228–33.

Associazione Nazionale Allevatori Suini (ANAS) 2016. ANAS Notizie: La selezione per il prosciutto DOP. Una sfida tecnica per un orizzonte dilungo periodo. Retrieved on 15 January 2018 from <http://www.anas.it/circolari/20170001A.PDF> Google Scholar

Anderson S.I., Lopez-Corrales N.L., Gorick B., Archibald A.L. (2000): A large fragment porcine genomic library resource in a BAC vector. *Mamm Genome* 11: 811–814.

Andolfatto P., Davison D., Erezyilmaz D., Hu T. T., Mast J., Sunayama-Morita T., & Stern D. L. (2011): Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *Genome Research*, 21(4), 610–617.

Andrés A. I., Cava R., Mayoral A. I., Tejada J. F., Morcuende D., & Ruiz J. (2001): Oxidative stability and fatty acid composition of pig muscles as affected by rearing system, crossbreeding and metabolic type of muscle fibre. *Meat Science*, 59, 39–47.

Ansorge W. J. (2009): Next-generation DNA sequencing techniques. *New Biotechnology* Volume 25, Number 4

Archibald A.L., Bolund L., Churcher C., Fredholm M., Groenen M. AM., Harlizius B., Kyung-Tai Lee, Milan D., Rogers J., Rothschild M.F., Uenishi H., Jun Wang, Schook L.B., the Swine Genome Sequencing Consortium (2010): Pig genome sequence - analysis and publication strategy. *BMC Genomics* 2010, 11:438.



Arden W., Brillouët M., Cogez P., Graef M., Huizing B., Mahnkopf R. (Eds.), ITRS (2010): More-than-Moore Whitepaper. The International Technology Roadmap for Semiconductors Consortium.

Aslam M. L., Bastiaansen J.W. M., Elferink M. G. et al. (2012): Whole genome SNP discovery and analysis of genetic diversity in Turkey (*Meleagris gallopavo*), BMC Genomics, vol. 13, no. 1, article 391.

Au K.F., Underwood J.G., Lee L., Wong W.H. (2012): Improving PacBio Long Read Accuracy by Short Read Alignment. PLoS ONE 7(10): e46679.

Bader J. S., Bansal A. & Sham P. C. (2001): Efficient SNP-based tests of association for quantitative phenotypes using pooled DNA. *GeneScreen* 1, 143–150.

Bader J. S. & Sham P. C. (2002): Family-based association tests for quantitative traits using pooled DNA. *Eur. J. Hum. Genet.*

Baird N. A. et al. (2008): Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3, e3376.

Barb C.R., Kraeling R.R., Rampacek G.B., Hausman G.J. (2006): The role of neuropeptide Y and interaction with leptin in regulating feed intake and luteinizing hormone and growth hormone secretion in the pig. *Reproduction*, 131, 1127-1135.

Barratt B. J. et al. (2002): Identification of the sources of error in allele frequency estimations from pooled DNA indicates an optimal experimental design. *Ann. Hum. Genet.*

Baxter S. W. et al. (2011): Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism. PLoS ONE 6, e19315.

Bentley D.R., Balasubramanian S., Swerdlow H.P., Smith G.P., Milton J., Brown C.G., Hall K.P., Evers D.J., Barnes C.L., Bignell H.R., et al. (2008): Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456(7218):53–59.

Benzinou M, Creemers J.W., Choquet H., Lobbens S., Dina C., Durand E., Guerardel A., Boutin P., Jouret B., Heude B., Balkau B., et al. (2008): Common nonsynonymous variants in PCSK1 confer risk of obesity. *Nature Genetics* 40, 943–945.

Berka J., Chen Y. J., Leamon J. H. et al. (2005): Bead emulsion nucleic acid amplification, U.S. Patent Application.

Bertolini F., Schiavo G., Scotti E., Ribani A., Martelli P.L., Casadio R. & Fontanesi L. (2014): High-throughput SNP discovery in the rabbit (*Oryctolagus cuniculus*) genome by next-generation semiconductor-based sequencing. *Animal Genetics* 45, 304–7.

Bertolini F., Scimone C., Geraci C., Schiavo G., Utzeri V.J., Chiofalo V. & Fontanesi L. (2015): Next generation semiconductor based sequencing of the donkey (*Equus asinus*) genome provided comparative sequence data against the horse genome and a fewmillions of single nucleotide polymorphisms. PLoS One 10, e0131925.

Bidanel J.P., Milan D., Iannuccelli N., Amigues Y., Boscher M.Y., Bourgeois F., Caritez J.C., Gruand J., Le Roy P., Lagant H., Quintanilla R., Renard C., Gellin J., Ollivier L., Chevalet C. (2001): Detection of quantitative trait loci for growth and fatness in pigs. *Genet Sel Evol.* 33(3):289-309.

Biodiversity International (2017). *Mainstreaming Agrobiodiversity in Sustainable Food Systems: Scientific Foundations for an Agrobiodiversity Index* (Biodiversity International, Fiumicino, Italy).

Bolger A.M., Lohse M., Usadel B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* Oxford University Press; 30:2114–20.

Bonneau M. & Lebret B. (2010): Production systems and influence on eating quality of pork. *Meat Science*, 84, 293–300.

Bosi P. & Russo V. (2004): The production of the heavy pig for high quality processed products. *Ital. J. Anim. Sci.* 3:309–321.

Bosse M., Megens H.J., Madsen O., Paudel Y., Frantz L.A., Schook L.B., et al. (2012). Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape. *PLoS Genet*; 8:e1003100.

Bosse M., Megens H.J., Frantz L.A., Madsen O., Larson G., Paudel Y., et al. (2014): Genomic analysis reveals selection for Asian genes in European pigs following human-mediated introgression. *Nat Commun.* 5:4392.

Bosse M., Megens H.J., Madsen O., Crooijmans R.P., Ryder O.A., Austerlitz F., et al. (2015): Using genome-wide measures of coancestry to maintain diversity and fitness in endangered and domestic pig populations. *Genome Res.* 25:970–81.

Bovo S., Bertolini F., Schiavo G., Mazzoni G., Dall'Olio S., & Fontanesi L. (2015): Reduced Representation Libraries from DNA Pools Analysed with Next Generation Semiconductor Based-Sequencing to Identify SNPs in Extreme and Divergent Pigs for Back Fat Thickness. *Int J Genomics.* 2015: 950737.

Bragg L. M., Stone G., Butler M. K., Hugenholtz P., & Tyson G.W. (2013): Shining a light on dark sequencing: characterizing errors in Ion Torrent PGM data. *PLoS Computational Biology*, vol. 9, no. 4, Article ID e1003031.

Brandl E. J., Tiwari A. K., Lett T. A., Shaikh S. A., Lieberman J. A., Meltzer H. Y., Kennedy J. L. & Müller D. J. (2013): Exploratory study on association of genetic variation in *TBC1D1* with antipsychotic-induced weight gain. *Hum. Psychopharmacol Clin Exp*, 28: 183-187.

Branton D, *et al.*, (2008): The potential and challenges of nanopore sequencing. *Nat Biotechnol* 26 (10):1146–1153.

Bruun C. S., Jørgensen C. B., Nielsen V. H., Andersson L., & Fredholm M. (2006):

Evaluation of the porcine melanocortin 4 receptor (MC4R) gene as a positional candidate for a fatness QTL in a cross between Landrace and Hampshire. *Animal Genetics*, 37(4), 359–362.

Cava R., Ruiz J., Ventanas J., & Antequera T. (1999): Oxidative and lipolytic changes during ripening of Iberian hams as affected by feeding regime: Extensive feeding and a tocopheryl acetate supplementation. *Meat Science*, 52, 165–172.

Cava R., Ventanas J., Ruiz J., Andres A. I., & Antequera T. (2000a). Sensory characteristics of Iberian ham: Influence of rearing system and muscle location. *Food Science and Technology International*, 6, 235–242

Cava R., Ventanas J., Tejada J. F., Ruiz J., & Antequera T. (2000b). Effect of free-range rearing and  $\alpha$ -tocopherol and copper supplementation on fatty acid profiles and susceptibility to lipid oxidation of fresh meat from Iberian pigs. *Food Chemistry*, 68, 51–59.

Chadt A., Leicht K., Deshmukh A., et al. (2008): Tbc1d1 mutation in lean mouse strain confers leanness and protects from diet-induced obesity. *Nat Genet* 40: 1354–1359

Chao Z., Wang F., Deng C.Y., Wei L.M., Sun R.P., Liu H.L., Liu Q.W., Zheng X.L. (2012): Distribution and linkage disequilibrium analysis of polymorphisms of MC4R, LEP, H-FABP genes in the different populations of pigs, associated with economic traits in DIV2 line. *Mol Biol Rep* 19, 111–120

Chen C.C., Chang T., Su H.Y. (2004): Genetic polymorphisms in porcine leptin gene and their association with reproduction and production traits. *Aust. J. Agric. Res.*, 55, 699–704.

Chessa S., Criscione A., Moretti R., Bordonaro S., Marletta D., Castiglioni B. (2013): Estimation of linkage disequilibrium in the Nero Siciliano Italian autochthonous breed using the Illumina 60K SNP array. 8th International Symposium on the Mediterranean Pig, Slovenia, Ljubljana, October 10th–12th, 2013.

Chiofalo B., Liotta L., Piccolo D., & Chiofalo V. (2005): “Características nutricionales del jamon Nero Siciliano”, III Congreso Mundial del jamon. Teruel (ES) 18-20 mayo 2005. 404-406.

Chiofalo B., Lo Presti V., Piccolo D., & Arena G. (2007): Nero Siciliano pig: Effect of the diet on meat quality. *Italian Journal of Animal Science*, 2(Suppl. 1), 679.

Chiofalo L. (2007): Nero Siciliano pig. Proceedings of 6th International Symposium on the Mediterranean Pig, Capo d’Orlando (ME), Italy, October 11-13, 2007.

Chiofalo V., Zumbo A., Liotta L., Chiofalo B. (2002): In vitam performances and carcass traits of “Nero Siciliano” pigs reared outdoor and plein air. 27<sup>th</sup> World Veterinary Congress. Tunis, September 25-29, 2002.

Chiofalo V., Chiofalo B., Liotta L., & Piccolo D. (2005): “Características nutricionales del jamon Nero Siciliano”, III Congreso mundial del Jamon Teruel, pg. 404-405.

Choi J.W., Chung W.H., Lee K.T., Cho E.S., Lee S.W., Choi B.H., et al. (2015): Whole genome resequencing analyses of five pig breeds, including Korean wild and native, and three European origin breeds. *DNA Res.* 22:259–67.

Chutimanitsakun Y. et al. (2011): Construction and application for QTL analysis of a restriction site associated DNA (RAD) linkage map in barley. *BMC Genomics* 12, 4.

Cingolani P., Platts A., Wang L.L., Coon M., Nguyen T., Wang L., et al. (2012a): A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*. 6:80–92.

Cingolani P., Patel V.M., Coon M., Nguyen T., Land S.J., Ruden D.M., et al. (2012b): Using *Drosophila melanogaster* as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. *Front. Genet. Frontiers*; 3:35.

Cloonan N. et al. (2008): Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat. Methods* 5, 613–619.

Convention on Biological Diversity's Strategic Plan for Biodiversity 2011–2020 Sustainable Development Goals, 2015. Available: <http://www.un.org>.

Corella D., Sorli Q.L., Godoy J.V., Portoles D., Coltell O. (2005): Obese subjects carrying the 11482G>A polymorphism at the perilipin locus are resistant to weight loss after dietary energy restriction. *J. Clin. Endocrinol. Metab.*, 90, 5121–5126.

Domestic Animal Diversity Information System (DAD-IS). Available: <http://dad.fao.org/>

D'Alessandro E., Fontanesi L., Liotta L., Davoli R., Chiofalo V., Russo V. (2007a): Analysis of the MC1R gene in the nero siciliano pig breed and usefulness of this locus for breed traceability. *Vet. Res. Commun.* 31:389–92.

D'Alessandro E., Fontanesi L., Liotta L., Davoli R., Chiofalo V., Russo V. (2007b): Nero Siciliano pig: analysis of coat colour affecting genes and perspectives for breed traceability, *Italian Journal of Animal Science*, 6: suppl. 105.

D'Andrea M., Pilla F., Giuffra E., Waddington D., & Archibald A. L. (2008): Structural analysis and haplotype diversity in swine LEP and MC4R genes. *Journal of Animal Breeding and Genetics = Zeitschrift Für Tierzucht Und Züchtungsbiologie*, 125 (2), 130–136.

Davoli R., Fontanesi L., Zambonelli P., Bigi D., Gellin J., Yerle M., Milc J., Braglia S., Cenci V., Cagnazzo M., Russo V. (2002): Isolation of porcine expressed sequence tags for the construction of a first genomic transcript map of the skeletal muscle in pig. *Anim Genet* 33:3–18).

Davoli R., Fontanesi L., Braglia S., Nisi I., Scotti E., Buttazzoni L., Russo V. (2006): Investigation of SNPs in the ATP1A2, CA3 and DECR1 genes mapped to porcine chromosome 4: analysis in groups of pigs divergent for meat production and quality traits. *ITAL.J.ANIM.SCI.* VOL. 5, 249-263.

Davoli R., Zambonelli P., San-Cristobal M., Scotti E., Fontanesi L., Colombo M., Dall'Olio S., Braglia S., Russo V. (2007): SNPs and microsatellite markers analysis for genetic diversity study in Italian pig breeds. Proceedings of 6th International Symposium on the Mediterranean Pig, Capo d'Orlando (ME), Italy, October 11-13, 2007.

Davoli R., Gandolfi G., Braglia S., Comella M., Zambonelli P., Buttazzoni L., Russo V. (2011): New SNP of the porcine perilipin 2 (PLIN2) gene, association with carcass traits and expression analysis in skeletal muscle. *Mol Biol Rep.* 38(3):1575-83.

Davoli R., Braglia S., Valastro V., et al. (2012): Analysis of MC4R polymorphism in Italian Large White and Italian Duroc pigs: association with carcass traits. *MeatSci.* 90(4):887–892.

Davoli R., Gaffo E., Zappaterra M., Bortoluzzi S., Zambonelli P. (2018): Identification of differentially expressed small RNAs and prediction of target genes in Italian Large White pigs with divergent backfat deposition. *Anim Genet.* 49(3):205-214.

Davey J. W. & Blaxter M. L. (2010). RADSeq: next-generation population genetics. *Brief. Funct. Genomics* 9, 416–423.

de Oliveira Peixoto J., Facioni Guimaraes S.E., Savio Lopes P., Menck Soares M.A., Vieira Pires A., Gualberto Barbosa M.V., de Almeida Torres R., de Almeida e Silva M. (2006): Associations of leptin gene polymorphisms with production traits in pigs. *J. Anim. Breed. Genet.*, 123, 378–383.

DePristo M. A., Banks E., Poplin R. et al. (2011): A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, vol. 43, no. 5, pp. 491–498.

Diaz I., Garcia Rigueiro J. A., Casillas M., & De Pedro E. (1996): Triglyceride composition of fresh ham fat from Iberian pigs produced with different systems of animal nutrition. *Food Chemistry*, 55, 383–387.

Di Rosa A.R, Chiofalo B., Lo Presti V., Zumbo A. (2009): Dietary acorns vs. barley in "Nero Siciliano" pigs: effect on the volatile compounds of salami. *Italian Journal of Animal Science*, vol. 8 (SUPPL. 2); p. 569.

Di Rosa A.R., Zumbo A., Furfaro M. E., Carcione G., D'Angelo G., V. Chiofalo V. (2013): Effect of two different rearing systems on quality of dry cured coppa of Nero Siciliano pig. *Acta Agriculturae Slovenica - Suppl.*4.

Di Rosa A.R, Magazzù G., Lo Presti V., Velez-Cervera N., Zumbo A., (2015): Chemical and technological traits of Nero Siciliano dry-cured ham. *Italian Journal of Animal Science*, vol.14: s 1 pag.91.

Dong D., Jones G., Zhang S. (2009): Dynamic evolution of bitter taste receptor genes in vertebrates. *BMC Evol. Biol.* 9.

Dressman D. et al. (2003): Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc. Natl. Acad. Sci. U. S. A.* 100, 8817–8822.

Dvořáková V., Bartenschlager H., Stratil A., Horák P., Stupka R., Čítek J., Šprysl M., Hrdlicová A. & Geldermann H.(2012): Association between polymorphism in the FTO gene and growth and carcass traits in pig crosses. *Genetics Selection Evolution*, 44:13.

Elshire R. J. et al. (2011): A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6, e19379.

Emerson K. J. et al. (2010): Resolving postglacial phylogeography using high-throughput sequencing. *Proc. Natl Acad. Sci. USA* 107, 16196–16200.

Esteve-Codina A., Kofler R., Himmelbauer H., Ferretti L., Vivancos A.P., Groenen M.A.M., Folch J.M., Rodríguez M.C. & Pérez-Enciso M. (2011): Partial short-read sequencing of a highly inbred Iberian pig and genomics inference thereof. *Heredity*; 107, 256–264.

Esteve-Codina A., Paudel Y., Ferretti L., Raineri E., Megens H.J., Silio L., et al. (2013): Dissecting structural and nucleotide genome-wide variation in inbred Iberian pigs. *BMC Genomics*.14:148.

Etter P.D., Bassham S., Hohenlohe P. A., Johnson E.A., & Cresko W.A. (2011): SNP Discovery and Genotyping for Evolutionary Genetics Using RAD Sequencing. *Methods Mol Biol.* 772: 157–178.

Faelli F., (1928). *Razze bovine, equine, suine, ovine, caprine*. Hoepli, Milano.

Fahrenkrug S.C., Rohrer G.A., Freking B.A., Smith T.P., Osoegawa K., Shu C.L., Catanese J.J., de Jong P.J. (2001): A porcine BAC library with tenfold genome coverage: a resource for physical and genetic map integration. *Mamm Genome*, 12:472-474.

Fahrenkrug S. C., Freking B. A., Smith T. P. L., Rohrer G. A. & Keele J. W. (2002): Single nucleotide polymorphism (SNP) discovery in porcine expressed genes. *Animal Genetics*. 33, 186–195.

Fan B., Du Z. Q., & Rothschild M. F. (2009): The fat mass and obesity-associated (FTO) gene is associated with intramuscular fat content and growth rate in the pig. *Anim. Biotechnol.* 20:58–70.

Fang X., Mou Y., Huang Z., Li Y., Han L., Zhang Y., et al. (2012): The sequence and analysis of a Chinese pig genome. *Gigascience*. 1:16.

Fernández A. I., Muñoz M., García F., Núñez Y., Geraci C., Crovetto A., García-Casco J., Alves E., Skrlep M., Riquet J., M.J., Bozzi R., Candek-Potokar M., Fontanesi L., Óvilo C. (2017): Distribution of polymorphisms in major and candidate genes for productive and domestication-related traits in European local pig breeds. 36th International Society for Animal Genetics Conference (ISAG 2017), Dublin, Ireland, July 16-21, 2017.

Fischer A., Gilad Y., Man O., Pääbo S. (2005): Evolution of bitter taste receptors in humans and apes. *Mol. Biol. Evol.* 22:432–6.

Fontanesi L., Davoli R., Nanni Costa L., Scotti E., Russo V. (2003): Study of candidate genes for glycolytic potential of porcine skeletal muscle: identification and analysis of mutations, linkage and physical mapping and association with meat quality traits in pigs. *Cytogenet Genome Res* 102:145–151.

Fontanesi L., Scotti E., Buttazzoni L., Davoli R. and Russo V. (2008a): The porcine fat mass and obesity associated (FTO) gene is associated with fat deposition in Italian Duroc pigs. *Animal Genetics*, 40, 90–93.

Fontanesi L., Davoli R., Nanni Costa L., Beretti F., Scotti E., Tazzoli M., Tassone F., Colombo M., Buttazzoni L., Russo V. (2008b): Investigation of candidate genes for glycolytic potential of porcine skeletal muscle: Association with meat quality and production traits in Italian Large White pigs. *Meat Science* 80; 780–787.

Fontanesi L., D'Alessandro E., Scotti E., Liotta L., Crovetto A., Chiofalo V., & Russo V. (2010a): Genetic heterogeneity and selection signature at the KIT gene in pigs showing different coat colours and patterns. *Anim. Genet.* 41:478–92.

Fontanesi L., Speroni C., Buttazzoni L., Scotti E., Dall'Olio S., Nanni Costa L., Davoli R. & Russo V., (2010b): The insulin-like growth factor 2 (IGF2) gene intron3-g.3072G>A polymorphism is not the only Sus scrofa chromosome 2p mutation affecting meat production and carcass traits in pigs: Evidence from the effects of a cathepsin D (CTSD) gene polymorphism. *J Anim Sci.*88:2235-2245.

Fontanesi L., Colombo M., Scotti E., Buttazzoni L., Bertolini F., Dall'Olio S., Davoli R., Russo V. (2010c). The porcine tribbles homolog 3 (TRIB3) gene: identification of a missense mutation and association analysis with meat quality and production traits in Italian heavy pigs. *Meat Science.* 86(3):808-13.

Fontanesi L., Scotti E., Buttazzoni L., Dall'Olio S., Bagnato A., Lo Fiego D.P., Davoli R. & Russo V. (2010d): Confirmed association between a single nucleotide polymorphism in the FTO gene and obesity-related traits in heavy pigs. *Mol Biol Rep.* 37:461–466.

Fontanesi L., D'Alessandro E., Scotti E., Liotta L., Chiofalo V., Russo V. (2010e): Analysis of the KIT gene in a Sicilian pig population and identification of the Id allele at the dominant white locus. 6th International Symposium on the Mediterranean pig - 14-16 October 2010 – Cordoba, Spain, pp. 3.

Fontanesi L., D'Alessandro E., Scotti E., Liotta L., Chiofalo V., Russo V. (2011a) - Analysis of genome wide polymorphisms in the Nero Siciliano pig breed - *Ital J. Anim. Sci.* 2011, Vol. 10 ( Suppl. 1), pp.74.

Fontanesi L., Colombo M., Tognazzi L., Scotti E., Buttazzoni L., Dall'Olio S., Davoli R., Russo V. (2011b): The porcine TBC1D1 gene: mapping, SNP identification, and association study with meat, carcass and production traits in Italian heavy pigs. *Mol Biol Rep.* 38:1425–1431.

Fontanesi L., Bertolini F., Scotti E., Trevisi P., Buttazzoni L., Dall'Olio S., Davoli R., Bosi P. & Russo V. (2012a): Polymorphisms in an obesity-related gene (PCSK1) are associated with fat deposition and production traits in Italian heavy pigs. *Animal*, 6:12, pp 1913–1924 & The Animal Consortium 2012.

Fontanesi L., Schiavo G., Galimberti G. et al. (2012b): A genome wide association study for backfat thickness in Italian Large White pigs highlights new regions affecting fat deposition including neuronal genes. *BMC Genomics*, vol. 13, no. 1, article 583.

Fontanesi L., Bertolini F., Dall'Olio S., Buttazzoni L., Gallo M. & Russo V. (2012c): Analysis of Association Between the MUC4 g.8227C>G Polymorphism and Production Traits in Italian Heavy Pigs Using a Selective Genotyping Approach. *Animal Biotechnology* Vol. 23 , Iss. 3.

Fontanesi L., Galimberti G., Calò D. G., Fronza R., Martelli P. L., Scotti E., Colombo M., Schiavo G., Casadio R., Buttazzoni L. & Russo V. (2012d): Identification and association analysis of several hundred single nucleotide polymorphisms within candidate genes for back fat thickness in Italian Large White pigs using a selective genotyping approach. *J ANIM SCI* , 90:2450-2464.

Fontanesi L., D'Alessandro E., Scotti E., Liotta L., Chiofalo V., Russo V. (2012e): Analysis of the KIT gene in Sicilian pig population and identification of the Id allele and the Dominant white locus. *Options Méditerranées, SERIES A: Mediterranean Seminars 2012 – Number 101*, pp 25-29.

Fontanesi L., Buttazzoni L., Galimberti G., Calò D.G., Scotti E., Russo V. (2013): Association between melanocortin 4 receptor (MC4R) gene haplotypes and carcass and production traits in Italian Large White pigs evaluated with a selective genotyping approach. *Livestock Science* 157; 48–56.

Fontanesi L., Bertolini F., Scotti E., Schiavo G., Colombo M., Trevisi P., Ribani A., Buttazzoni L., Russo V. & Dall'Olio S. (2015): Next Generation Semiconductor Based-Sequencing of a Nutrigenetics Target Gene (GPR120) and Association with Growth Rate in Italian Large White Pigs. *Animal Biotechnology*, 26:92–97.

Fontanesi L., Scotti E., Gallo M., Nanni Costa L., Dall'Olio S. (2016). Authentication of “mono-breed” pork products: Identification of a coat colour gene marker in Cinta Senese pigs useful to this purpose. *LivestockScience*184; 71–77.

Fortina R., Barbera S., Lussiana C., Mimosi A., Tassone S., Rossi A., Zanardi E. (2005): Performance and meat quality of two Italian pig breeds fed diets for commercial hybrids. *Meat Science*, vol. 71, p. 713-718.

Frantz L.A.F., Schraiber J.G., Madsen O., Megens H.J., Cagan A., Bosse M., et al. (2015): Analyses of Eurasian wild and domestic pig genomes reveal long-term gene-flow and selection during domestication. *Nat Genet.* 47:1141–8.

Fuller CW, et al. (2009): The challenges of sequencing by synthesis. *Nat Biotechnol* 27(11):1013–1023.



Gaffo E., Zambonelli P., Bisognin A., Bortoluzzi S., Davoli R. (2014). miRNome of Italian Large White pig subcutaneous fat tissue: new miRNAs, isomiRs and moRNAs. *Anim Genet.* 45(5):685-98.

Galve A., Burgos C., Silió L., Varona L., Rodríguez C., Ovilo C., et al. (2012): The effects of leptin receptor (LEPR) and melanocortin-4 receptor (MC4R) polymorphisms on fat content, fat distribution and fat composition in a Duroc × Landrace/Large White cross. *Livest Sci.* 145:145–152.

García-González D.L., Aparicio R., Aparicio-Ruiz R. (2013): Volatile and amino acid profiling of dry cured hams from different swine breeds and processing methods. *Molecules.* 18:3927–47.

Giuliani (1940). Numero special dedicato alla suinicoltura italiana. *Riv. Di Zoot.*, 1, 1 –12.

Glenn T. C. (2011): Field guide to next-generation DNA sequencers. *Molecular Ecology Resources.* 11, 759–769.

Global Plan of Action for Animal Genetic Resources, 2007. International Technical Conference on Animal Genetic Resources for Food and Agriculture Interlaken, Switzerland, 3–7 September 2007. Available: <http://www.fao.org/docrep/010/a1404e/a1404e00.htm>

Go Y., Satta Y., Takenaka O., Takahata N. (2005): Lineage-specific loss of function of bitter taste receptor genes in humans and non human primates. *Genetics.* 170:313–26.

Gol S., Ros-Freixedes R., Zambonelli P., Tor M., Pena R.N., Braglia S., Zappaterra M., Estany J. & Davoli R. (2015): Relationship between perilipin genes polymorphisms and growth, carcass and meat quality traits in pigs. *J. Anim. Breed. Genet.* ISSN 0931-2668

Goodwin S., McPherson J.D. & McCombie W.R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics* volume 17, pages 333–351.

Gore M. A. *et al.* (2009): A first-generation haplotype map of maize. *Science* 326, 1115–1117.

Groenen M.A.M., Archibald A.L., Uenishi H., Tuggle C.K, Takeuchi Y., Rothschild M.F., et al. (2012): Analyses of pig genomes provide insight into porcine demography and evolution. *Nature.* 491:393–8.

Groenen M.A.M. (2016). A decade of pig genome sequencing: a window on pig domestication and evolution. *Genet. Sel. Evol. BioMed Central*;48:1–9.

Guastella A.M., Criscione A., Marletta D., Zuccaro A., Chies L., Bordonaro S. (2010): Molecular characterization and genetic structure of the Nero Siciliano pig breed. *Genet. Mol. Biol.* 33:650–6.

Guo S.W., & Thompson E.A. (1992): Performing the Exact Test of Hardy-Weinberg Proportion for Multiple Alleles. *Biometrics* 48: 361–72.

- Hernández-Sánchez J., Visscher P., Plastow Graham, & Haley, C. (2003). Candidate gene analysis for quantitative traits using the transmission disequilibrium test: The example of the melanocortin 4-receptor in pigs. *Genetics*, 164(2), 637–644.
- Herrero-Medrano J.M., Megens H-J, Groenen M.A.M., Bosse M., Pérez-Enciso M., Crooijmans R.P.M.A. (2014): Whole-genome sequence analysis reveals differences in population management and selection of European low-input pig breeds. *BMC Genomics*. 15:601.
- Hinchliff C.E., et al. (2015): Synthesis of phylogeny and taxonomy into a comprehensive tree of life. *Proc Natl Acad Sci USA* 112:12764–12769.
- Hirose K., Takizawa T., Fukawa K., Ito T., Ueda M., Hayashi Y., Tanaka K. (2011): Association of an SNP marker in exon 24 of a class 3 phosphoinositide-3-kinase (PIK3C3) gene with production traits in Duroc pigs. *Anim Sci J*. 82(1):46-51.
- Hirose K., Ito T., Fukawa K., Arakawa A., Mikawa S., Hayashi Y., Tanaka K., (2014): Evaluation of effects of multiple candidate genes (LEP, LEPR, MC4R, PIK3C3, and VRTN) on production traits in Duroc pigs. *Animal Science Journal*. 85, 198–206.
- Hohenlohe P. A. et al. (2010): Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet*. 6, e1000862.
- Hohenlohe P. A., Amish S. J., Catchen J. M., Allendorf F. W. & Luikart G. (2011): Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Mol. Ecol. Resour*. 11, 117–122.
- Homer N., Merriman B. TMAP: the Torrent Mapping Alignment Program. Available: <https://github.com/iontorrent/TMAP>.
- Huang X. et al. (2009): High-throughput genotyping by whole-genome resequencing. *Genome Res*. 19, 1068–1076.
- Humphray S.J., Scott C.E., Clark R., Marron B., Plumb R., Bender C., Camm N., Davis J., Jenks A., Noon A., Patel M., Sehra H., Yang F., Rogatcheva M.B., Milan D., Chardon P., Rohrer G., Nonneman D., de Jong P., Meyers S.N., Archibald A., Beever J.E., Schook L.B, Rogers J. (2007): A high utility integrated map of the pig genome. *Genome Biol*. 8(7):R139.
- Huse S. M., Huber J. A., Morrison H. G., Sogin M. L., & Welch D. M. (2007): Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology*, vol. 8, no. 7, article R143.
- Houston R. D., Cameron N. D., & Rance K. A. (2004): A melanocortin-4 receptor (MC4R) polymorphism is associated with performance traits in divergently selected Large White pig populations. *Animal Genetics*, 35(5), 386–390.

Ibeagha-Awemu E.M., Kgwatalala P., & Zhao X. (2008): A critical analysis of production-associated DNA polymorphisms in the genes of cattle, goat, sheep, and pig. *Mamm. Genome* 19, 591–617.

International Union for Conservation of Nature (2017) IUCN 2016: International Union for Conservation of Nature annual report 2016 (International Union for Conservation of Nature, Gland, Switzerland).

Jawaid A., Bader J. S., Purcell S., Cherny S. S. & Sham P. (2002): Optimal selection strategies for QTL mapping using pooled DNA samples. *Eur. J. Hum. Genet.*

Jeon J.T., Carlborg O., Tornsten A., Giuffra E., Amanger V., Chardon P., Andersson-Eklund L., Andersson K., Hansson I., Lundstrom K., Andersson L.(1999): A paternally expressed QTL affecting skeletal and cardiac muscle mass in pigs maps to the IGF2 locus. *Nat. Genet.* 21, 157-158.

Jeong H., Song K.D., Seo M., Caetano-Anollés K., Kim J., Kwak W., et al. (2015): Exploring evidence of positive selection reveals genetic basis of meat quality traits in Berkshire pigs through whole genome sequencing. *BMC Genet.* 16:104.

Jiang Z. H., Gibson J. P., (1999): Genetic polymorphisms in the leptin gene and their association with fatness in four pig breeds. *Mamm. Genome.* 10:191-193.

Jokubka R., Maak S., Kerziene S., & Swalve H. H. (2006): Association of a melanocortin 4 receptor (MC4R) polymorphism with performance traits in Lithuanian White pigs. *Journal of Animal Breeding and Genetics = Zeitschrift Für Tierzüchtung Und Züchtungsbiologie*, 123(1), 17–22.

Jungerius B.J., van Laere A.S., Pas M.F.W., van Oost B.A., Andersson L., Groenen M.A.M. (2004): The IGF2-intro-3-G3072A substitution explains a major imprinte QTL effect on backfat thickness in a Meishan×European white pig intercross. *Genetical Research.* 84 (2), 95-101.

Jungerius B.J., Jingjing G., Crooijmans R.P.M.A., Van Der Poel J.J., Groenen M.A.M., Van Oost B.A., et al. (2005): Estimation of the extent of linkage disequilibrium in seven regions of the porcine genome. *Anim. Biotechnol.* 16:41–54.

Kang H., Wang H., Fan Z., Zhao P., Khan A., Yin Z., et al. (2015): Resequencing diverse Chinese indigenous breeds to enrich the map of genomic variations in swine. *Genomics.* 106:286–94.

Kerstens H. H. *et al.* (2009): Large scale single nucleotide polymorphism discovery in unsequenced genomes using second generation high throughput sequencing technology: applied to turkey. *BMC Genomics* 10, 479.

Kim K. S., Larsen N., Short T., Plastow G., & Rothschild M. F. (2000): A missense variant of the porcine melanocortin-4 receptor (MC4R) gene is associated with fatness, growth, and feed intake traits. *Mammalian Genome: Official Journal of the International Mammalian Genome Society*, 11(2), 131–135.

- Kennes Y.M., Murphy B.D., Pothier F., Palin M.F. (2001): Characterization of swine leptin (LEP) polymorphisms and their association with production traits. *Anim. Genet.*, 32, 215–218.
- Kim J. H., Lee Y. S., Park E. W., Seo B. Y., Cho I. C., Lee J. G., Oh S. J., Lee J. H. & Jeon J. T. (2005a): Assignment of the phosphoinositide-3-kinase, class 3 (PIK3C3) gene to porcine chromosome 6q22→q23 by somatic cell and radiation hybrid panel mapping. *Cytogenet. Genome Res.* 108:362A.
- Kim J. H., Choi B. H., Lim H. T., Park E. W., Lee S. H., Seo B. Y., Cho I. C., Lee J. G., Oh S. J. & Jeon J. T. (2005b): Characterization of Phosphoinositide-3-kinase, Class 3 (PIK3C3) Gene and Association Tests with Quantitative Traits in Pigs. *Asian-Aust. J. Anim. Sci.* Vol 18, No. 12: 1701-1707.
- Kumar P., Henikoff S., Ng P.C. (2009): Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc.* 4(7):1073-81.
- Kumar S., Banks T.W. & Cloutier S. (2012): SNP Discovery through Next-Generation Sequencing and Its Applications. *International Journal of Plant Genomics*, Volume 2012, Article ID 831460, 15 pages.
- Kumar S., Stecher G., Tamura K. (2016). *MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets.* *Mol. Biol. Evol.* Oxford University Press; 33:1870–4.
- Kušec G., Dovč P., Karolyi D., Čandek P. M. (2015): Local pig breeds and pork products in Croatia and Slovenia – unexploited treasure. *Poljoprivreda* 21:(1) Supplement, 16-21.
- Labroue F., Goumy S., Gruand J., Mourot J., Neelz V., Legault C., (2000): Comparaison au Large White de quatre races locales porcine françaises pour les performance de croissance, de carcasse et de qualité de la viande. *Journées Recherche Porcine en France*, vol. 32, p 403-411.
- Lebret B. (2008): Effects of feeding and rearing systems on growth, carcass composition and meat quality in pigs. *Animal*, 2, 1548–1558.
- Lebret B., Dourmad J.Y., Mourot J., Pollet P.Y., Gondret F. (2014). Production performance, carcass composition, and adipose tissue traits of heavy pigs: influence of breed and production system. *JAnimSci.* 92(8):3543–3556. 28.
- Le Hellard S. *et al.* (2002). SNP genotyping on pooled DNAs: comparison of genotyping technologies and a semi automated method for data storage and analysis. *Nucleic Acids Res.*
- Leinonen R., Sugawara H., Shumway M., on behalf of the International Nucleotide Sequence Database Collaboration (2011): The Sequence Read Archive, *Nucleic Acids Research*, Volume 39, Issue suppl\_1, Pages D19–D21.
- Lewin H. A., Robinson G. E., Kress W. J., et al. (2018). Earth BioGenome Project: Sequencing life for the future of life. *PNAS.* 115 (17) 4325-4333.

- Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., et al. (2009a): The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 25:2078–9.
- Li Y., Willer C., Sanna S., Abecasis G. (2009b): Genotype imputation. *Annu Rev Genomics Hum Genet*. 10:387-406.
- Li H. (2011): A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 1;27(21):2987-93.
- Li M., Tian S., Jin L., Zhou G., Li Y., Zhang Y., et al. (2013): Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nat Genet*. 45:1431–8.
- Li M., Tian S., Yeung C.K., Meng X., Tang Q., Niu L., et al. (2014): Whole-genome sequencing of Berkshire (European native pig) provides insights into its origin and domestication. *Sci Rep*.4:4678.
- Li H., Durbin R. (2009): Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. Oxford University Press; 25:1754–60.
- Li H., Durbin R. (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26: 589–595.
- Liotta L., Chiofalo V., Zumbo A., & Chiofalo L. (2002). Nero Siciliano pigs reared in plain air and lived in extensive condition: data on tissue separation of fresh ham and shoulder. In 48th international congress of meat science and technology. Roma, 25–30 August 2002.
- Lin Liu, Yinhu Li, Siliang Li, Ni Hu, Yimin He, Ray Pong, Danni Lin, Lihua Lu, & Maggie Law (2012a): Comparison of Next-Generation Sequencing Systems. *Journal of Biomedicine and Biotechnology*, Volume 2012.
- Liu Q., Guo Y., Li J., Long J., Zhang B., Shyr Y. (2012b): Steps to ensure accuracy in genotype and SNP calling from Illumina sequencing data *BMC Genomics*; 13, Suppl. 8:S8.
- Luo Y., Qiu X., Li H., Zhang Q. (2010): Association between the polymorphism in FUT1 gene and the resistance to PWD and ED in three pig breeds. *Asian-Australasian J. Anim. Sci*. 23:1268–75.
- Madonia S., Porcu S., Liotta L., Margiotta S., Ligios S., Chiofalo V., Russo V. (2007a): Italian local breeds Sarda and Nero Siciliano: Typical breeding system, carcass traits and meat quality. Proceedings of the 6th International Symposium on the Mediterranean Pig, Messina, (p. 96). Capo d'Orlando, Italy (2007, October 11–13).
- Madonia S., Porcu S., Liotta L., Margiotta S., Chiofalo V., Ligios S. (2007b): Slaughter traits of “Sarda” and “Nero Siciliano” pigs reared outdoors. Preliminary results. *Italian Journal of Animal Science*, 6:sup1, 697.

Maiorano G. (2009): Swine production in Italy and research perspectives for the local breeds. *Slovak J. Anim. Sci.*, 42 (4): 159 – 166.

Mardis E. R. (2008): The impact of next-generation sequencing technology on genetics. *Trends in Genetics*, vol. 24, no. 3, pp. 133–141.

Margulies M., Egholm M., Altman W.E., Attiya S., Bader J.S. et al. (2005): Genome sequencing in microfabricated high-density picolitre reactors, *Nature* 437; 376–380.

McLaren W., Gil L., Hunt S.E., Riat H.S., Ritchie G.R., Thormann A., Flicek P., Cunningham F. (2016): The Ensembl Variant Effect Predictor. *Genome Biology* Jun 6;17(1):122.

Maxam A.M. & Gilbert W. (1977): A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America*, vol. 74, no. 2, pp. 560–564.

Malek M., Dekkers J.C.M., Lee H.K., Baas T.J., Prusa K., Huff-Lonergan E., Rothschild M.F. (2001a): A molecular genome scan analysis to identify chromosomal regions influencing economic traits in the pig. II. Meat and muscle composition. *Mammal Genome* 12:637–645.

Meidtner K., Wermter A. K., Hinney A., Remschmidt H., Hebebrand J., & Fries R. (2006): Association of the melanocortin 4 receptor with feed intake and daily gain in F2 Mangalitsa×Piétrain pigs. *Animal Genetics*, 37(3), 245–247.

Meijerink E., Neuenschwander S., Fries R., Dinter A., Bertschinger G.S., Vögeli P. (2000): A DNA polymorphism influencing  $\alpha(1,2)$ fucosyltransferase activity of the pig FUT1 enzyme determines susceptibility of small intestinal epithelium to *Escherichia coli* F18 adhesion. *Immunogenetics*. 52:129–36.

Menendez J.A., Vazquez-Martin A., Ortega F.J., Fernandez-Real J.M. (2009). Fatty acid synthase: Association with insulin resistance, type 2 diabetes, and cancer *Clinical Chemistry*. 55, pp. 425-438.

Metzker M. L. (2010): Sequencing technologies — the next generation. *Nature Reviews Genetics* volume 11, pages 31–46.

Miller M. R., Dunham J. P., Amores A., Cresko W. A. & Johnson E. A. (2007): Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res*. 17, 240–248.

Barry Merriman B. & Rothberg J. M. (2012): Progress in Ion Torrent semiconductor chip based sequencing. *Electrophoresis*. 33, 3397–3417.

Meyre D., Farge M., Lecoecur C., Proenca C. et al (2008): R125W coding variant in *TBC1D1* confers risk for familial obesity and contributes to linkage on chromosome 4p14 in the French population, *Human Molecular Genetics*, Volume 17, Issue 12, Pages 1798–1802.

- Molnár J., Nagy T., Stéger V., Tóth G., Marincs F., Barta E. (2014): Genome sequencing and analysis of Mangalica, a fatty local pig of Hungary. *BMC Genomics*. 15:761.
- Mortazavi A. et al. (2008): Mapping and quantifying mammalian transcriptomes by RNA Seq. *Nat. Methods* 5, 621–628.
- Mracek T., Ding Q., Tzanavari T., Kos K., Pinkney J., Wilding J., et al. (2010): The adipokine zinc- $\alpha$ 2-glycoprotein (ZAG) is downregulated with fat mass expansion in obesity. *Clin. Endocrinol. (Oxf)*. 72:334–341.
- Muir W.M., Wong G.K., Zhang Y., Wang J., Groenen M.A.M., Crooijmans R.P.M.A., et al. (2008): Genome-wide assessment of worldwide chicken SNP genetic diversity indicates significant absence of rare alleles in commercial breeds.
- Muñoz G., Óvilo C., Noguera J.L., Sánchez, Rodríguez C., Silió L. (2003): Assignment of the fatty acid synthase (FASN) gene to pig chromosome 12 by physical and linkage mapping *Animal Genetics*, 34, pp. 234-235.
- Muñoz G., Alcazar E., Fernandez A., Barragan C., Carrasco A., de Pedro E., Silió L., Sanchez J.L., Rodriguez M.C. (2011): Effects of porcine MC4R and LEPR polymorphisms, gender and Duroc sire line on economic traits in Duroc X Iberian crossbred pigs. *Meat Science* 88, 169-173.
- Nakajima I., Oe M., Ojima K., Muroya S., Shibata M., Chikuni K. (2011). Cellularity of developing subcutaneous adipose tissue in Landrace and Meishan pigs: adipocyte size differences between two breeds. *AnimSciJ*. 82(1):144–149.
- Neuenschwander S., Rettenberger G., Meijerink E., Jorg H., Stranzinger G. (1996): Partial characterization of porcine obesity gene (OBS) and its localization to chromosome 18 by somatic cell hybrids. *Anim Genet* 27(4):275–278.
- Nezer C., Moreau L., Brouwers B., Coppieters W., Detilleux J., Hanset R., Karim L., Kvasz A., Leroy P., Georges M. (1999): An imprinted QTL with major effect on muscle mass and fat deposition maps to the IGF2 locus in pigs. *Nat. Genet*. 21(2), 155-156.
- Nezer C., Moreau L., Wagenaar D., Georges M. (2002): Results of a whole genome scan targeting QTL for growth and carcass traits in a Piétrain x Large White intercross. *Genet Sel Evol*. 34(3):371-87.
- Nielsen R., Paul J.S., Albrechtsen A., Song Y.S. (2011): Genotype and SNP calling from next-generation sequencing data. *Nat Rev Genet*. 12(6):443-51.
- Ning Z., Cox A.J. & Mullikin J.C. (2001). SSAHA: a fast search method for large DNA databases. *Genome Research* 11, 1725–9.
- Oczkiewicz M., Walinowicz K., Rozycki M., Rejduch B., (2009): Known mutation (A3072G) in intron 3 of the IGF2 gene is associated with growth and carcass composition in Polish pig breeds. *J Appl Genet*. 50(3), 257-259
- Okonechnikov K., Conesa A., García-Alcalde F. (2016). Qualimap 2: advanced multi-

- sample quality control for high-throughput sequencing data. *Bioinformatics*. 32:292–4.
- Ollivier L. (2009). European pig genetic diversity: a minireview. *Animal*. 3:915–24.
- Óvilo C., Clop A., Noguera J.L. et al. (2002): Quantitative trait locus mapping for meat quality traits in an Iberian × Landrace F2 pig population. *J Anim Sci* 80:2801–2808.
- Óvilo C., Fernández A., Noguera J.L., Barragán C., Letón R., Rodríguez C., et al. (2005): Fine mapping of porcine chromosome 6 QTL and LEPR effects on body composition in multiple generations of an Iberian by Landrace intercross. *Genet Res*. 85:57–67.
- Paternoster L., Evans D.M., Nohr E.A., Holst C., Gaborieau V., Brennan P., Gjesing A.P., Grarup N., Witte D.R., Jørgensen T., Linneberg A., Lauritzen T., et al. (2011): Genome-wide population-based association study of extremely overweight young adults - the GOYA study. *PLoS One* 6:e24303.
- Pena R.N., Gallardo D., Guardia M.D., Reixach J., Arnau J., Amills M., Quintanilla R. (2013): Appearance, flavor, and texture attributes of pig dry-cured hams have a complex polygenic genomic architecture. *Journal of Animal Science*, 91, pp. 1051-1058.
- Pennisi E. (2017): Sequencing all life captivates biologists. *Science* 355:894–895.
- Pérez- Montarelo D., Fernández A., Folch J. M. , Pena R. N., Óvilo C., Rodríguez C., Silió L. & Fernández A. I. (2012): Joint effects of porcine leptin and leptin receptor polymorphisms on productivity and quality traits. *Animal Genetics*, 43, 805–809.
- Pettersson E., Lundeberg J., Ahmadian A. (2009): Generations of sequencing technologies. *Genomics*, Volume 93, Issue 2, Pages 105-111.
- Pfender W. F., Saha M. C., Johnson E. A. & Slabaugh M. B. (2011): Mapping with RAD (restriction-site associated DNA) markers to rapidly identify QTL for stem rust resistance in *Lolium perenne*. *Theor. Appl. Genet.* 122, 1467–1480.
- Prudente S., Scarpelli D., Chandalia M., Zhang Y.Y., Morini E., Del Guerra S., et al. (2009): The TRIB3 Q84R polymorphism and risk of early-onset type 2 diabetes. *Journal of Clinical Endocrinology and Metabolism*, 94, 190-196.
- Pugliese C., Calagna G., Chiofalo V., Moretti M., Margiotta S., Franci O., & Gandini G. (2004a). Comparison of the performances of Nero Siciliano pigs reared indoors and outdoors: Joints composition, meat and fat traits. *Meat Science*, 69, 523–528.
- Pugliese C., Calagna G., Chiofalo V., Moretti V.M., Margiotta S., Franci O., Gandini G. (2004b): Comparison of the performances of Nero Siciliano pigs reared indoors and outdoors: 2. Joints composition, meat and fat traits- *Meat Science*, 68(4).
- Pugliese C., Bozzi R., Campodoni G., Acciaioli A., Franci O., & Gandini G. (2005): Performance of Cinta Senese pigs reared outdoors and indoors. Meat and subcutaneous fat characteristics. *Meat Science*, 69, 459–464.



- Pugliese C. & Sirtori F. (2011): Quality of meat and meat products produced from southern European pig breeds. *Meat Science* 90, 511–518.
- Qi L., Shen H., Larson I., Schaefer E.J., Greenberg A.S., Tregouet D.A (2004): Gender-specific association of a perilipin gene haplotype with obesity risk in a white population. *Obes. Res.*, 12, 1758–1765.
- Quail M. A., Smith M., Coupland P. et al. (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and IlluminaMiSeq sequencers. *BMC Genomics*, vol. 13, no. 1, article 341.
- Quinlan A.R., Hall I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinforma. Appl. NOTE*. 26:841–84210.
- Raineri E., Ferretti L., Esteve-Codina A., Nevado B., Heath S., & Pérez-Enciso M. (2012): SNP calling by sequencing pooled samples. *BMC Bioinformatics*, vol. 13, no. 1, article 239.
- Ramos A.M., Crooijmans R.P.M.A., Amaral A.J., Archibald A.L. et al. (2009): Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. *PLoS ONE* 4(8): e6524.
- Raymond M., & Rousset F. (1995). GENEPOP Version 1.2: Population Genetics Software for Exact Tests and Ecumenicism.” *J. Hered.* 86: 248–49.
- Ren J., Knorr C., Guo Y.-M., Ding N.-S., Ai H.-S., Brenig B. & Huang L.-S. (2004): Characterization of five single nucleotide polymorphisms in the porcine stearoyl-CoA desaturase (*SCD*) gene. *Animal Genetics*, 35: 255–257.
- Rhoads A. & Au K.F., (2015): PacBio Sequencing and Its Applications. *Genomics, Proteomics & Bioinformatics*. Volume 13, Issue 5, Pages 278-289.
- Ribani A., Bertolini F., Schiavo G., Scotti E., Utzeri V.J., Dall’Olio S., et al. (2017): Next generation semiconductor based sequencing of bitter taste receptor genes in different pig populations and association analysis using a selective DNA pool-seq approach. *Anim. Genet.* 48:97–102.
- Robertson G. et al. (2007): ChIP-Seq techniques. *Nat. Methods* 4, 651–657.
- Rodriguez J.M., Maietta P., Ezkurdia I., Pietrelli A., Wesselink J-J, Lopez G., et al. (2013). APPRIS: annotation of principal and alternative splice isoforms. *Nucleic Acids Res.* 41:D110–7.
- Rogel-Gaillard C., Bourgeaux N., Billault A., Vaiman M., Chardon P. (1999): Construction of a swine BAC library: application to the characterization and mapping of porcine type C endoviral elements. *Cytogenet Cell Genet.* 85:205-211.
- Rohrer G.A., Keele J.W. (1998a). Identification of quantitative trait loci affecting carcass composition in swine: Muscling and wholesale product yield traits. *J Anim Sci.* 76(9):2255-62.

- Rohrer G.A., Keele J.W. (1998b). Identification of quantitative trait loci affecting carcass composition in swine: Fat deposition traits. *J Anim Sci.* 76(9):2247-54.
- Ros-Freixedes R., Gol S., Pena R.N., Tor M., Ibáñez-Escriche N., Dekkers J.C.M., et al. (2016): Genome-Wide Association Study Singles Out SCD and LEPR as the Two Main Loci Influencing Intramuscular Fat Content and Fatty Acid Composition in Duroc Pigs. *PLoS ONE* 11(3): e0152496.
- Rothberg J. M., Hinz W., Rearick T. M. et al. (2011): An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, vol. 475, no. 7356, pp. 348–352.
- Rowe H.C., Renaut S., Guggisberg A. (2011): RAD in the realm of next-generation sequencing technologies. *Mol Ecol.* 20(17):3499-502.
- Rubin C-J, Megens H-J, Barrio A.M., Maqbool K., Sayyab S., Schwochow D., et al. (2012): Strong signatures of selection in the domestic pig genome. *Proc. Natl. Acad. Sci.* 109:19529–19536.
- Ruiz J.R., Larrarte E., Margareto J., Ares R., Alkorta P., Labayen I. (2011): Preliminary findings on the role of PLIN1 polymorphisms on body composition and energy metabolism response to energy restriction in obese women. *Br. J. Nutr.*, 106, 486–490.
- Rusch D.B., Halpern A.L., Sutton G., Heidelberg K.B., Williamson S., Yooseph S., et al. (2007). The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific, *PLoS Biol.* 5, e77.
- Russo V., Fontanesi L., Davoli R., Nanni Costa L., Cagnazzo M., Buttazzoni L., Virgili R. & Yerle M. (2002). Investigation of candidate genes for meat quality in dry-cured ham production: the porcine cathepsin B (CTSB) and cystatin B (CSTB) genes. *International Society for Animal Genetics, Animal Genetics*, 33, 123–131.
- Russo V., Fontanesi L., Davoli, R., Chiofalo L., Liotta L., Zumbo A. (2004): Analysis of single nucleotide polymorphisms in major and candidate genes for production traits in Nero Siciliano pig breed. *Italian Journal Animal Science*, 3, 2004, p. 19-29.
- Russo V., Fontanesi L., Scotti E., Beretti F., Davoli, Nanni Costa L., Virgili R., & Buttazzoni L. (2008): Single nucleotide polymorphisms in several porcine cathepsin genes are associated with growth, carcass, and production traits in Italian Large White pigs. *J. Anim. Sci.* 86:3300–3314.
- Russo V. & Nanni Costa L. (1995): Suitability of pig meat for salting and the production of quality processed products. *Pig News and Information* 16, 17N–26N.
- Saha P.K., Kojima H., Martinez-Botas J., Sunehag A.L., Chan A.L. (2004). Metabolic adaptations in the absence of perilipin: increased beta-oxidation and decreased hepatic glucose production associated with peripheral insulin resistance but normal glucose tolerance in perilipin-null mice. *J. Biol. Chem.*, 279, 35150–35158.
- Sánchez C. et al. (2009): Single nucleotide polymorphism discovery in rainbow trout by deep sequencing of a reduced representation library. *BMC Genomics* 10, 559.

Sanger F., Nicklen S., & Coulson A. R. (1977a): DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, vol. 74, no. 12, pp. 5463–5467.

Sanger F., Air G. M., Barrell B. G. et al., (1977b): Nucleotide sequence of bacteriophage phiX174 DNA. *Nature*, vol. 265, no. 5596, pp. 687–695.

Sanseverino W., Hénaff E., Vives C., Pinosio S., Burgos-Paz W., Morgante M., et al. (2015): Transposon Insertions, Structural Variations, and SNPs Contribute to the Evolution of the Melon Genome. *Mol. Biol. Evol.* 32:2760–74.

SAS Institute Inc. (1998) User's Guide. Statistical Analysis System. SAS® Inst. Inc., Cary, NC, USA.

Saxena R., Voight B.F., Lyssenko V., Burtt N.P., et al., (2007): Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, Novartis Institutes of BioMedical Research. Science* 316:1331–1336.

Schook L.B., Beever J.E., Rogers J., Humphray S., Archibald A., Chardon P., Milan D., Rohrer G., Eversole K. (2005). Swine Genome Sequencing Consortium (SGSC): a strategic roadmap for sequencing the pig genome. *Compar Funct Genom* 6:251–255.

Seeley R.J., Yagaloff K.A, Fisher S.L., Burn P., Thiele T.E. et al. (1997): Melanocortin receptors in leptin effects. *Nature* 390, 349.

Serra X., Gil F., Pérez-Enciso M., Oliver M. A., Vázquez J. M., Gispert M., Dýaz I., Moreno F., Latorre R., Noguera, J. L. (1998): A comparison of carcass, meat quality and histochemical characteristics of Iberian (Guadyerbas line) and Landrace pigs. *Livestock Production Science*, vol. 56, p 215-223.

Sha B.Y., Yang T.L., Zhao L.J., et al. (2009): Genome-wide association study suggested copy number variation may be associated with body mass index in the Chinese population. *JHum Genet.* 54(4):199–202.

Sham P., Bader J.S., Craig I., O'Donovan M., Owen M. (2002). DNA Pooling: a tool for large-scale association studies. *Nat Rev Genet.* 3(11):862-71.

Sharma V., Sarkar I.N. (2013): Leveraging biodiversity knowledge for potential phyto-therapeutic applications. *J Am Med Inform Assoc* 20:668–679.

Simonella S. (2011): Sistemi di allevamento e aspetti tecnologici funzionali alla produzione del “prosciutto crudo di suino Nero dei Nebrodi DOP”. Tesi di dottorato in “Sanità Pubblica, Igiene Veterinaria e delle Produzioni Animali”, XXIV CICLO, non pubblicata. Università degli Studi di Messina.

Sonah H., Bastien M., Iquira E., Tardivel A., Légaré G., Boyle B., Normandeau É., Laroche J., Larose S., Jean M., Belzile F. (2013): An improved genotyping by sequencing

(GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. *PLoS One*. 8(1):e54603.

Stachowiak M., Szczerbal I., Switonski M. (2016): Genetics of Adiposity in Large Animal Models for Human Obesity—Studies on Pigs and Dogs. *Prog Mol Biol Transl Sci*. 140:233-70.

Stone S., Abkevich V., Russell D.L., Riley R., et al., (2006): TBC1D1 is a candidate for a severe obesity gene and evidence for a gene/gene interaction in obesity predisposition, *Human Molecular Genetics*, Volume 15, Issue 18, Pages 2709–2720.

Sultan M. et al. (2008): A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* 321, 956–960.

Switonski M., Stachowiak M., Cieslak J., Bartz M., Grzes M. (2010). Genetics of fat tissue accumulation in pigs: a comparative approach. *J Appl Genet* 51(2), pp. 153–168.

Szabo A., Viski A., Egyhazi Z., Hazas Z., Horn P., Romvari R. (2010). Comparison of Mangalica and Hungarian Large White pigs at identical bodyweight: 1 Backfat histology. *ArchTierz*. 53:141–146. 27.

Tapio I., Värvi S., Bennewitz J., Maleviciute J., Fimland E., Grislis Z., et al. (2006): Prioritization for conservation of northern European cattle breeds based on analysis of microsatellite data. *Conserv. Biol*. 20:1768–79.

The Second Report on the State of the World's Animal Genetic Resources for Food and Agriculture (2015). FAO Commission on Genetic Resources for Food and Agriculture Assessments.

Thompson J.D., Gibson T.J., Plewniak F., Jeanmougin F., Higgins D.G. (1997). The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*. 25(24):4876-82.

Thorvaldsdóttir H., Robinson J. T., Mesirov J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, Volume 14, Issue 2, Pages 178–192.

Tomás A., Casellas J., Ramírez O., Muñoz G., Noguera J., Sánchez A. (2006): High amino acid variation in the intracellular domain of the pig prolactin receptor (PRLR) and its relation to ovulation rate and piglet. High amino acid variation in the intracellular domain of the pig prolactin receptor (PRLR) and its relation to ovula. *J. Anim. Sci*. 84:1991–8.

Turnbaugh, P.J. et al. (2007). The human microbiome project. *Nature* 449, 804–810.

Van den Maagdenberg K., Stinckens A., Claeys E., Seynaeve M., Clinquart A., Georges M., et al. (2007): The Asp298Asn missense mutation in the porcine melanocortin-4 receptor (MC4R) gene can be used to affect growth and carcass traits without an effect on meat quality. *Animal*, 1(08), 1089–1098.

- Van Emon J.M. (2016): The Omics Revolution in Agricultural Research. *J. Agric. Food Chem.*, 64 (1), pp 36–44.
- Van Gestel S., Houwing-Duistermaat J.J., Adolfsson R., van Duijn C.M., Van Broeckhoven C. (2000): Power of selective genotyping in genetic association analyses of quantitative traits. *Behav Genet*, 30:141-146.
- van Orsouw N.J., Hogers R.C.J., Janssen A., Yalcin F., Snoeijers S., et al (2007). Complexity Reduction of Polymorphic Sequences (CRoPS™): A Novel Approach for Large-Scale Polymorphism Discovery in Complex Genomes. *PLoS ONE* 2(11): e1172.
- Van Rens B.T.T.M., Evans G.J., Van Der Lende T. (2003): Components of litter size in gilts with different prolactin receptor genotypes. *Theriogenology*. 59:915–26.
- van Tassell C. P. *et al.* (2008): SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nature Methods* 5, 247–252.
- Ventanas S., Ruiz J., Garcia C., & Ventanas J. (2007): Preference and juiciness of Iberian dry-cured loin as affected by intramuscular fat content, crossbreeding and rearing system. *Meat Science*, 77, 324–330.
- Venter J.C., Remington K., Heidelberg J.F., Halpern A.L., Rusch D., et al. (2004): Environmental genome shotgun sequencing of the Sargasso Sea, *Science* 304; 66–74.
- Vincent A.L., Tuggle C.K., Rothschild M.F., Evans G., Short T.H., Southwood O.I., et al. (1997): The Prolactin Receptor Gene is Associated with Increased Litter Size In Pigs. *Swine Res. Rep.* 11.
- Virgili R., Parolari G., Schivazappa C., Soresi Bordini C. & Volta R. (1995): Effects of raw material on proteolysis and texture of typical Parma ham. *Industria Conserve* 70, 21–31.
- Virgili R., Parolari G., Schivazappa C., Soresi Bondini C. & Volta R. (1998): Proteases in fresh pork muscle and their influence on bitter taste formation in dry-cured ham. *Journal of Food Biochemistry* 22, 53–63.
- Virgili R., Schivazappa C. (2002): Muscle traits for long matured dried meats. *Meat Science* 62; 331–343.
- Vykoukalová Z., Knoll A., Cepica S. (2009). Porcine perilipin (PLIN) gene: structure, polymorphism and association study in large white pigs. *Czech J. Anim. Sci.*, 54, 359–364.
- Waguri T., Goda T., Kasezawa N., Yamakawa-Kobayashi K. (2013): The combined effects of genetic variations in the GPR120 gene and dietary fat intake on obesity risk. *Biomed Res*; 34:69–74.
- Wakil S.J. (1989): Fatty acid synthase, a proficient multifunctional enzyme *Biochemistry*, 28, pp. 4523-4530.

- Wan Laere A.S., Nguyen N., Braunschwig M., Nezer C., Collette C., Moreau L., Archibald A.L., Haley C.S., Buys N., Tally M., Andersson G., Georges M., Andersson L., (2003): A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature*. 425, 832-836.
- Wang D.G. et al., (1998): Large-Scale Identification, Mapping, and Genotyping of Single-Nucleotide Polymorphisms in the Human Genome. *Science* 280, 1077.
- Wang C., Wang H., Zhang Y., Tang Z., Liu B. (2015): Genome-wide analysis reveals artificial selection on coat colour and reproductive traits in Chinese domestic pigs. *Mol Ecol Resour.* 15:414–24.
- Warr A., Robert C., Hume D., Archibald A.L., Deeb N., Watson M. (2015): Identification of low-confidence regions in the pig reference genome (Sscrofa10.2). *Front. Genet.* 6:1–8.
- Wiedmann R. T., Smith T. P. & Nonneman D. J. (2008): SNP discovery in swine by reduced representation and high throughput pyrosequencing. *BMC Genet.* 9, 81.
- Wilkinson S., Lu Z.H., Megens H.J., Archibald A.L., Haley C., Jackson I.J., et al. (2013): Signatures of Diversifying Selection in European Pig Breeds. *PLoS Genet.* 9.
- Wilson E.O. (1999): *The Diversity of Life* (Norton, New York).
- World Wildlife Fund (2016) *Living Planet Report 2016: Risk and resilience in a new era* (World Wildlife Fund, Gland, Switzerland).
- Xie W. et al. (2010): Parent-independent genotyping for constructing an ultrahigh-density linkage map based on population sequencing. *Proc. Natl Acad. Sci. USA* 107, 10578–10583.
- Ye K., Schulz M.H., Long Q., Apweiler R., Ning Z. (2009). Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinforma.* 25:2865–287110.
- Yeo Z.X., Chan M., Yap Y.S., Ang P., Rozen S., et al. (2012): Improving indel detection specificity of the Ion Torrent PGM benchtop sequencer. *Plos One* 7: e45798.
- Yooseph S., Sutton G., Rusch D.B., Halpern A.L., Williamson S.J., Remington K., et al. (2007). The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families, *PLoS Biol.* 5, e16.
- Zambonelli P., Gaffo E., Zappaterra M., Bortoluzzi S., Davoli R. (2016): Transcriptional profiling of subcutaneous adipose tissue in Italian Large White pigs divergent for backfat thickness. *Animal Genetics*.
- Zappaterra M., Deserti M., Mazza R., Braglia S., Zambonelli P., Davoli R. (2016): A gene protein expression study on four porcine genes related to intramuscular fat deposition. *Meat Science*.

- Zeggini E., Weedon M.N., Lindgren C.M., Frayling T.M., Elliott K.S., Lango H., Timpson N.J., Perry J.R., Rayner N.W., Freathy R.M., Barrett J.C., Shields B., Morris A.P., Ellard S, Groves C.J., Harries L.W., Marchini J.L., Owen K.R., Knight B., Cardon L.R., Walker M., Hitman G.A., Morris A.D., Doney A.S., Wellcome Trust Case Control Consortium (WTCCC), McCarthy M.I., Hattersley A.T. (2007): Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 316:1336–1341.
- Zeng F., Jiang R., Chen T. (2013). PyroHMMvar: a sensitive and accurate method to call short indels and SNPs for Ion Torrent and 454 data. *Bioinformatics*.
- Zerbino D.R., Achuthan P., Akanni W., Amode M.R., et al., (2018): Ensembl 2018, *Nucleic Acids Research*, Volume 46, Issue D1, Pages D754–D761.
- Zhang G., Nebert D.W., Chakraborty R., Jin L. (2006): Statistical power of association using the extreme discordant phenotype design. *Pharmacogenet Genomics*. 16:401-413.
- Zhu P., He L., Li Y., Huang W., Xi F., Lin L., Zhi Q., Zhang W., Tang Y.T., Geng C., Lu Z., Xu X., (2014). OTG-snpcaller: An Optimized Pipeline Based on TMAP and GATK for SNP Calling from Ion Torrent Data. *PLoS ONE* 9(5): e97507.
- Zumbo A., Chiofalo B., Liotta L., & Chiofalo L. (2002): Physical characteristics of the meat of “Nero Siciliano” pigs living in extensive condition. In 48<sup>th</sup> international congress of meat science and technology, (pp. 738–739). Roma, Italy (2002, August 25–30).
- Zumbo A., Di Rosa A. R., Lo Presti V., Pruiti V., & Piccolo D. (2007a): Analysis of the volatile composition of fresh and seasoned lard in “Nero Siciliano” pigs in relation of fattening diet. Proceedings of the 6th International Symposium on the Mediterranean Pig, Messina, (pp. 194–197). Capo d'Orlando, Italy (2007, October 11–13).
- Zumbo A., Lo Presti V., Di Rosa A. R., Pruiti V., Di Marco V., & Piccolo D. (2007b): Fatty acid profile of intramuscular fat of “Nero Siciliano” fattening pigs fed with different diets. Proceedings of the 6th International Symposium on the Mediterranean Pig, Messina, (pp. 224–227). Capo d'Orlando, Italy (2007, October 11–13).

## 8. Sitografia

[ftp.ncbi.nih.gov/snp/organisms/pig\\_9823/](ftp.ncbi.nih.gov/snp/organisms/pig_9823/)

<http://www.genome.jp/kegg/>

[www.ANAS.it](http://www.ANAS.it)

<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

<https://github.com/iontorrent/TS/tree/master/Analysis/TMAP>

[www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)

[www.thermofisher.com](http://www.thermofisher.com)

<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>

<http://www.fao.org/home/en/>

<http://www.un.org/>

[www.sequentiabiotech.com](http://www.sequentiabiotech.com)

<https://www.ensembl.org/>

<https://www.animalgenome.org/QTLdb/>

<https://treasure.kis.si/>

<https://genome10k.soe.ucsc.edu/>



## **8. Ringraziamenti**

Si ringrazia l'*Unità di Produzione Animale* del Dipartimento di Scienze Veterinarie dell'Università degli Studi di Messina ed il *Consorzio di Ricerca Filiera Carni* per il supporto tecnico-scientifico.