SARDEGNA RICERCHE

Progetto Cluster
TIPIZZAZIONE GENETICA DELLA POPOLAZIONE SARDA

30 Dicembre 2009

A. INQUADRAMENTO E DATI SALIENTI

PREMESSA

Sardegna Ricerche ha avviato la realizzazione e organizzazione di un'insieme di piattaforme tecnologiche, ossia una rete di laboratori e infrastrutture che hanno come fine il potenziamento delle dotazioni tecnologiche a disposizione delle imprese, del parco tecnologico, delle Università e dei Centri di Ricerca delle Sardegna.

Due delle piattaforme piu' rilevanti realizzate sono quelle di genotyping ed expression profiling e di bioinformatica che si inseriscono nel Progetto ICT SIAI101 "Creazione di un polo di eccellenza delle tecnologie bio-informatiche applicate alla medicina personalizzata".

Questa rete di piattaforme è una risorsa a cui devono poter accedere tutte le realtà che svolgono nell'isola attività di ricerca, innovazione e trasferimento tecnologico, e per questo motivo Sardegna Ricerche si occupa della raccolta di informazioni volte alla organizzazione ed implementazione delle attività e a evidenziarne il potenziale applicativo tecnologico e scientifico.

È in quest'ottica che si inseriscono gli obbiettivi generali dei programmi in fase operativa che intendono promuovere e potenziare il trasferimento delle conoscenze della ricerca di base in patrimonio applicativo per la comunità imprenditoriale operante all'interno del distretto tecnologico regionale della biomedicina: ed in particolare favorire le collaborazioni e sinergie tra i soggetti operanti nel settore della biomedicina e promuovere l'accesso alle piattaforme tecnologiche di Sardegna Ricerche.

Due delle piattaforme piu' rilevanti realizzate sono quelle di genotyping ed expression profiling e di bioinformatica che si inseriscono nel Progetto ICT SIAI101 "Creazione di un polo di eccellenza delle tecnologie bio-informatiche applicate alla medicina personalizzata".

Il progetto cluster "Tipizzazione genetica della popolazione Sarda", realizzato mediante l'utilizzo delle strutture e del personale della piattaforma di Genotyping e Expression profiling, prevede appunto un'azione specifica per la valorizzazione della struttura e l'interazione ed integrazione con le università e le imprese. Il progetto prevede la determinazione dell'assetto genetico individuale di un campione rappresentativo della popolazione dell'isola di individui anonimizzati di origine sarda esenti da malattie in corso. Questa considerevole quantità di profili genetici, che verranno processati ed archiviati con procedure bioinfomatiche dedicate, costituirà una buona base di riferimento per lo studio dei profili genetici della popolazione generale sarda e sarà una base di confronto per le aziende/università partecipanti e per altri soggetti potenzialmente interessati per lo studio di patologie a forte componente genetica caratteristiche dell'isola e/o ubiquitarie.

OBIETTIVI GENERALI

Gli obbiettivi del progetto erano effettuare attività sperimentali, in collaborazione e con il massimo coinvolgimento delle imprese e degli altri soggetti (Università, etc), con l'organizzazione e la realizzazione delle seguenti attività:

- training e assistenza nella realizzazione delle attività sperimentali per il personale esterno;
- determinazione dei profili genetici di un campione rappresentativo sano della popolazione sarda;
- analisi di qualità, generazione e gestione dei genotipi tramite analisi bioinformatiche;
- analisi genetiche preliminari delle caratteristiche del campione analizzato sia dal punto di vista di genetica di popolazione che della creazione di un campione di confronto utilizzato come riferimento nello studio di patologie genetiche complesse.

SOGGETTO DESTINATARI

Imprese localizzate al Parco Tecnologico, soggetti industriali localizzati nella Regione Sardegna, centri di ricerca pubblici e universitari Sardi.

SOGGETTO ATTUATORE

Sardegna Ricerche in collaborazione con il gruppo di ricerca delle imprese e centri di ricerca partecipanti al progetto. I referenti del progetto sono per le responsabilità amministrative il Dott. Luca Contini (Polaris – Edificio 2, Loc. Piscinamanna 09010 Pula-CA) e per le competenze tecnico-scientifiche il Dott. Andrea Angius (Lab. Genomica, Polaris – Edificio 3, Loc. Piscinamanna 09010 Pula-CA).

LOCALIZZAZIONE

Il progetto e' stato svolto utilizzando le competenze e le piattaforme tecnologiche disponibili presso il Parco Scientifico della Sardegna – Pula, ed in particolare il Laboratorio di Genomica.

COPERTURA FINANZIARIA

La copertura finanziaria è stata reperita attraverso l'azione 4.e del Progetto ICT SIAI101 "Creazione di un polo di eccellenza delle tecnologie bio-informatiche applicate alla medicina personalizzata", Regione Autonoma della Sardegna, APQ Società dell'Informazione.

DURATA

Il progetto e' stato avviato nel gennaio 2009 ed e' stato concluso a dicembre dello stesso anno.

B. DESCRIZIONE PROGETTO

Svolgimento del progetto

Il primo obbiettivo in ordine temporale svolto dal progetto è stato l'organizzazione di una serie di incontri con i soggetti aderenti al programma proposto in cui sono state spiegate in dettaglio le modalità di adesione al progetto, la modulistica di partecipazione ed è stato esposto il piano tecnico scientifico previsto dalla proposta.

L'obbiettivo principale delle prime riunioni è stato la definizione delle caratteristiche dei campioni biologici da analizzare in modo da creare un campione biologico adatto allo scopo prefissato e che fosse facilmente reperibili, dove possibile, all'interno delle bio-banche già in possesso dei partecipanti sia per motivi di abbattimento dei costi per il reperimento e l'estrazione degli acidi nucleici sia per motivi di tempi di realizzazione visto la durata relativamente limitata dell'intero progetto che non avrebbe consentito di svolgere appieno tutte le attività previste. Il ruolo dei soggetti partecipanti in questa fase consisteva nel contribuire alla definizione dei contenuti del progetto, apportare informazioni, conoscenze ed esigenze specifiche di ognuno in base alle precedenti esperienze e indirizzare le attività verso la migliore soluzione in termini di esigenze condivise.

Hanno aderito i seguenti partecipanti:

BCS Biotech SpA (Cagliari);

Shardna SpA (Pula);

Inforesearch (Pula);

Laboratori di Informatica Applicata di G. Capasso (Cagliari);

Dip. Biologia sperimentale, Università di Cagliari;

Associazione Sindrome di Crisponi e malattie rare Onlus (Oristano);

Dip. Ingegneria elettrica ed elettronica, Università di Cagliari.

La scelta delle caratteristiche dei campioni è stata concordata con alcune riunioni in cui si sono stabilite le informazioni delle quali erano necessario essere in possesso per includere i soggetti nel gruppo da analizzare (provenienza geografica, età, genealogia sia anamnestica che documentata, caratteristiche cliniche finalizzate se possibile all'assenza di una lista di malattie più ampia possibile ed esclusione di legami di parentela stretta). Sono stati infine decisi i seguenti criteri per la selezione del campione: soggetti sani ed in particolare senza familiarità per malattie cardiovascolari, tiroidee, diabete, etc, non imparentati sino alla quarta generazione e residenti nello stesso paese da almeno 3 generazioni. Sono stati inclusi nello studio un totale di 325 individui descritti in dettaglio nella tabella sottostante.

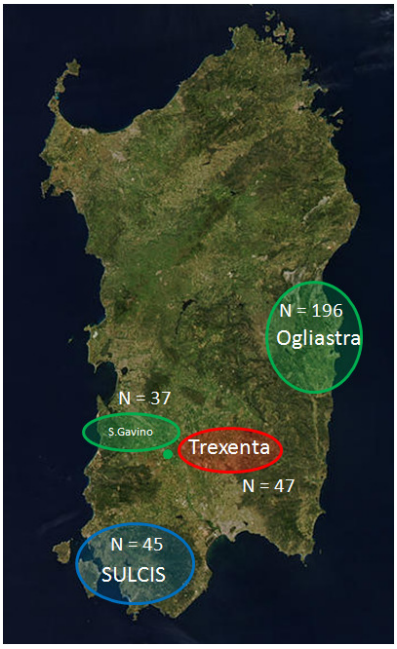

FIGURA 1.

Localizzazione geografica e numero di soggetti inclusi nello studio.

Paese	Tot	Fem	Mas	Paese	Tot	Fem	Mas
Arixi	2	2	0	*Guasila*	3	2	1
Barrali	1	1	0	*Mandas*	4	3	1
Gesico	3	2	1	*Ortacesus*	1	0	1
Guamaggiore	7	3	4	*San Basilio*	6	2	4
Selegas	3	3	0	*Siurgus Donigala*	4	3	1
Calasetta	3	1	2	*Suelli*	3	1	2
Carbonia	7	5	2	*Senorbì*	11	7	4
Domus De Maria	4	3	1	*S.G. Suergiu*	2	1	1
Giba	2	1	1	*Teulada*	4	3	1
Portoscuso	3	2	1	*Barisardo*	27	15	12
Sant'Antioco	19	10	9	*Jerzu*	40	18	22
Ulassai	41	27	14	*Tertenia*	60	22	38
Lanusei	2	2	0	*Villagrande*	26	9	17
San Gavino	37	17	20				
				Tot	**325**	**165**	**160**

Attività di training

La prima delle attività di training ha previsto una giornata di incontri con il Dott. Alessandro Davassi, Field Application Specialist (FAS Italy) e del Dott. Stefano Romanello (account manager per l'Italia), in rappresentanza della ditta Affymetrix, con sede a 3380 Central Expressway - Santa Clara, California (sito internet: www.affymetrix.com), per una serie di presentazioni e descrizioni tecnico scientifiche dei prodotti e delle metodiche da utilizzare nella genotipizzazione dei campioni selezionati a cui hanno partecipato sia i rappresentanti dei vari gruppi di ricerca sia il personale proposto dalle singole unità come collaboratore tecnico per le fasi sperimentali.

Ogni unità del progetto ha messo a disposizione 3 biologi con adeguate conoscenze di base che sono stati opportunamente seguiti e formati durante tutte le fasi del protocollo di genotipizzazione mediante chip dal personale del laboratorio di Genomica, Piattaforma di genotyping e expression profiling (Dott. Andrea Angius, Dott. Marco Marcelli e Dott.ssa Francesca Urru). Sono stati potenziati sia una serie di strumentazioni del laboratorio sia una serie di piccole attrezzature ed il laboratorio ha creato le condizioni ottimali dal punto di vista logistico affinché i gruppi partecipanti

potessero inserirsi in maniera agevole all'interno delle attività di altri progetti che erano già in corso in laboratorio contemporaneamente al progetto cluster. Questa pianificazione ha permesso di operare in condizioni ottimali ai gruppi di 3 operatori di ogni azienda con una processività adatta a rispettare i tempi stabili dal progetto.
La fase di training ha previsto per ogni gruppo 2/3 settimane di affiancamento in laboratorio in cui sono stati genotipizzati una serie di individui di controllo da ogni unità operativa. Il training ha previsto sia attività prettamente sperimentali sia attività di carattere teorico sui principi e il funzionamento della metodica sia una serie di lezioni sui database internazionali da cui sono stati estratte le informazioni necessarie per la realizzazione del chip. Inoltre ogni unità è stata ampiamente istruita nell'utilizzo dei software di gestione delle macchine da laboratorio e nei software di generazione dei genotipi e di controllo di qualità dei dati generati.

Attività sperimentali

L'attività di genotyping svolta presso il laboratorio di Genomica, Piattaforma di genotyping e expression profiling si avvale di strumentazioni tra le più moderne e performanti attualmente presenti sul mercato. Tale strumentazione è stata concepita con una tecnologia ad alte prestazioni che permette di processare un elevato numero di campioni in tempi brevi. Attualmente i nostri laboratori hanno notevolmente aumentato la produttività riducendo tempi e costi. Questo è stato possibile grazie al perfezionamento e messa a punto delle metodologie che hanno consentito di perfezionare i nostri standard lavorativi sino al raggiungimento di una processività paragonabile a quella di laboratori internazionali universalmente riconosciuti.
La piattaforma tecnologica utilizzata nel presente progetto, prevede l'utilizzo di chip ad alta densità per gli studi di genome wide search tramite marcatori a singolo nucleotide (Single Nucleotide Polymorphisms, SNPs). La tecnica basata su microarray consente di ibridare il DNA di un individuo a sonde di acido nucleico legate a un supporto solido. E' possibile determinare simultaneamente il genotipo di un soggetto per circa 1.000.000 SNPs e circa un altro milione di polimorfismi di variazione di numero (Copy Number Variation, CNV) opportunamente scelti in base alla loro informatività e spaziatura all'interno del genoma umano.
La strumentazione necessaria a questa metodica consiste in un sistema composto da: n.2 scanners (Affymetrix Genechip 3000 (7G e 7G Plus) Scanner with Autoloader), 2 workstations, n. 3 fornetti di ibridazione (Gene Chip Hybridization Oven 640) e n.5 stazioni fluidiche (Gene Chip Fluidics Station 450) e la di base occorrente ad un laboratorio di biologia molecolare. Inoltre viene utilizzata un'apposita rete interna che consente il flusso diretto dei dati dalle apparecchiature dove vengono generati i dati genetici ai computer dove risiedono i software per le analisi statistiche. I dati generati mediante la piattaforma tecnologica sono gestiti e depositati su un apposito server settato ad hoc per le particolari necessità determinate dalla ingente mole di dati informatici e i backup dei dati prodotti.

Il genotyping è stato eseguito interamente presso il Laboratorio di Genomica, Piattaforma Genotyping ed Expression Profiling localizzato presso il Parco Tecnologico della Sardegna (Pula, Cagliari), utilizzando il chip a DNA Affymetrix Genome-Wide Human SNP Array 6.0, che rappresenta uno dei prodotti d'avanguardia per questo tipo di studi. Le attività di laboratorio sono state eseguite secondo i protocolli forniti dall'azienda produttrice e implementati e migliorati in laboratorio in alcuni passaggi. Inoltre il primo passo fondamentale è stato quello di analizzare la qualità e quantità del DNA di ciascun soggetto in modo da creare le condizioni ottimale per una buona riuscita dell'esperimento.

Analisi di qualità, generazione e gestione dei genotipi tramite analisi bioinformatiche

La fase successiva alla generazione dei dati su chip ha previsto la generazione vera e propria dei genotipi mediante il Genotyping Console™ Software 4.0. Questo software permette in un primo momento di effettuare una analisi di qualità su un numero limitato (circa 30.000) di SNPs opportunamente selezionati che genera una serie di parametri (Contrast Quality Control algorithm) con dei cut off che permettono di stabilire in maniera chiara la qualità dell'esperimento. I controlli di qualità vengono effettuati separatamente sia sui probe che riconoscono uno degli enzimi di restrizione (Nsp) sia l'altro (Sty) sia combinando le analisi. Inoltre la percentuale dei soggetti genotipizzati con successo è di circa il 99 %. I 5 soggetti con problemi sperimentali sono poi stati rigenotipizzati con successo.
I dati generati dai vari gruppi hanno mostrato dei coefficienti di qualità molto alti che in media avevano un valore di Contrast QC di 2.6-2,8: tenendo conto che il limite inferiore di qualità è rappresentato da un valore maggiore di 0.4 possiamo sicuramente affermare di aver ottenuto il massimo di dati genotipici da ogni esperimento (vedi dettagli nella tabella sottostante).

	Num subiects	*Bounds In*	*Bounds Out*	*%*	*Contrast QC*	*Contrast QC (Random)*	*Contrast QC (Nsp)*	*Contrast QC (Nsp/Sty)*	*Contrast QC (Sty)*
BCS	233	228	5	97,85	2,79	2,79	2,77	3,13	3,05
BPU1/BPU2	92	46	0	100,00	2,63	2,63	2,68	2,93	2,18

Inoltre il nuovo algoritmo sviluppato recentemente (Birdseed v2) è in grado di individuare e valutare con precisione un numero più alto di genotipi utili dallo stesso esperimento. La fase successiva prevede la generazione dei dati genotipici in vari formati per vari software in maniera ordinata in modo da utilizzare direttamente per le successive analisi statistiche gli output del programma.

Analisi genetiche preliminari delle caratteristiche del campione analizzato

Il dataset generato ha rappresentato la base su cui eseguire una serie di analisi preliminari per valutare le caratteristiche genetiche del campione prodotto.

Con l'intento di ottenere un set di SNPs che presentano la minima possibilità di errori di genotipizzazione sono stati esclusi dalla successiva analisi quelli che presentavano un P value per l'equilibrio di Hardy–Weinberg[1] (HWE) con valori inferiori a 10^{-5} e una percentuale di genotipizzazione per ogni marcatore inferiore al 95% (SNP call rate). I criteri adottati per minimizzare gli errori dovuti al genotyping hanno messo in rilievo una particolare affidabilità del metodo utilizzato eliminando dall'intero set una percentuale bassissima di genotipi (circa il 3%). Inoltre in modo da selezionare solamente marcatori particolarmente utili per l'analisi genetica e con frequenze alleliche informative abbiamo filtrato il dataset utilizzando una frequenza allelica minima (Minimum Allele Frequency, MAF) maggiore di 0.05. In seguito a questi parametri abbiamo ottenuto una base dati di 620.830 markers. Il numero di SNPs per ogni cromosoma è compreso tra 8.122 del cromosoma 22 e 52.315 del cromosoma 2 (Figura 2).

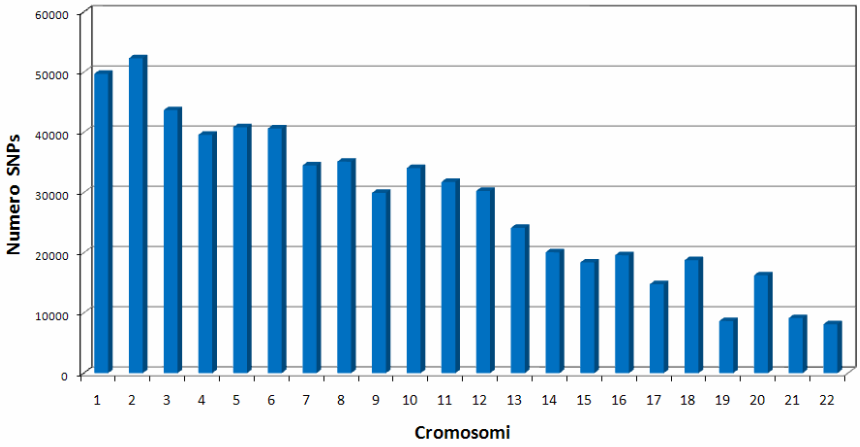

FIGURA 2.

Numero di SNPs informativi (620,830) ottenuti dal dataset iniziale. Sono stati esclusi i polimorfismi che presentavano: P-value Equilibrio di Hardy Weinberg ($P<10^{-5}$, MAF < 0.05 e SNP call rate < 95).

[1] *Legge di Hardy-Weinberg: Nel 1908 G. H. Hardy e W. Weinberg definirono "popolazione in equilibrio" una popolazione all´interno della quale né le frequenze alleliche né la distribuzione dei genotipi mutano col succedersi delle diverse generazioni. Non modificandosi le frequenze degli alleli, non si avrebbe, pertanto, evoluzione. Una popolazione resta in equilibrio solo se in essa si verificano alcune condizioni restrittive: i) non devono verificarsi mutazioni; ii) non deve verificarsi un flusso di geni tra popolazioni, cioè non deve esserci una migrazione netta di alleli verso l´interno della popolazione (immigrazione) o verso l´esterno (emigrazione); iii) la popolazione deve essere ampia (teoricamente infinita); iv) non si deve verificare selezione naturale, vale a dire tutti i genotipi devono possedere le stesse capacità adattative e riproduttive. Soddisfatte queste condizioni, le frequenze alleliche entro una popolazione rimarranno costanti per un periodo di tempo indefinito. Tale equilibrio è espresso dalla seguente equazione: $p + q = 1$ da cui $(p + q) = p + 2pq + q = 1$. In questa equazione la lettera p indica la frequenza di un allele e la lettera q indica la frequenza dell´altro allele; p più q deve sempre essere uguale a 1 (cioè il 100% degli alleli di quel particolare gene nel pool genico). L´espressione p indica la frequenza di individui omozigoti per un allele, q la frequenza di individui omozigoti per l´altro allele e 2pq la frequenza degli eterozigoti.*

Di seguito vengono presentate come esempio alcune analisi preliminari eseguite utilizzando gli SNPs informativi localizzati nei cromosomi 15 e 22, contenenti rispettivamente 18.388 e 8.122 marcatori, selezionati secondo i criteri di MAF, HWE e SNPs call rate descritti precedentemente. L'analisi dei dati è stata eseguita con l'ausilio del software R (pacchetti "POPGEN" e "GENETICS" - http://www.r-project.org/).

I risultati relativi all'equilibrio di Hardy–Weinberg sono mostrati nelle Figure 3A e 3B dove sono stati riportati i valori di P-value per ogni SNP divisi in classi di frequenza. In tutte le popolazioni la classe maggiormente rappresentata (con una frequenza pari al 60%) è quella che comprende i P-value tra 0.5 e 1. L'unica differenza che si riscontra tra le popolazioni esaminate riguarda i campioni dell'Ogliastra (BCS), i quali possiedono una frequenza inferiore della classe 0.5 - 1 rispetto alle altre popolazioni, e una frequenza superiore nelle altre classi. Non si rilevano differenze sostanziali tra i due cromosomi esaminati.

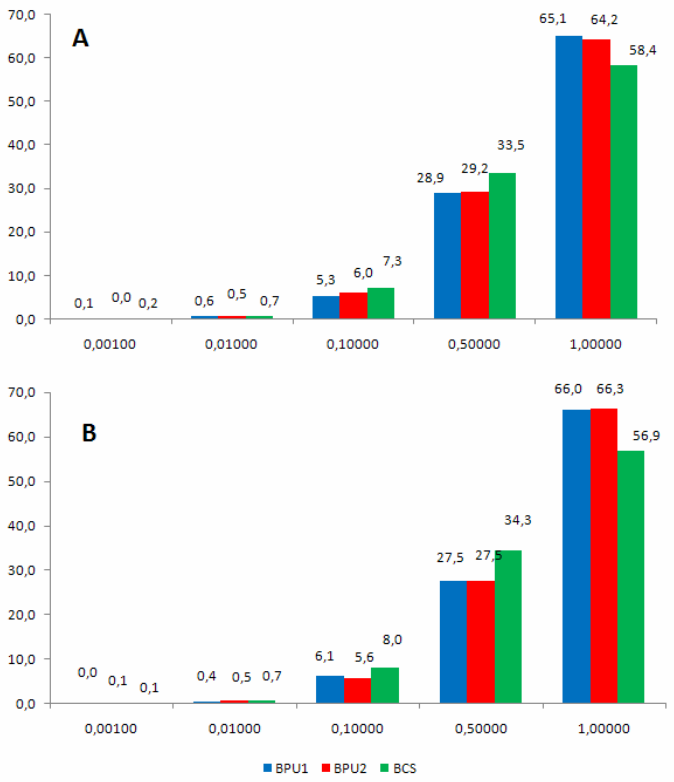

FIGURA 3.

Distribuzione dei P-value relativi al test di Hardy – Weinberg per i cromosomi 15 (A) e 22 (B). In ascissa sono indicate le classi di frequenza, in ordinata le frequenze percentuali per le diverse popolazioni.

Per quanto riguarda l'eterozigosità [2], i risultati sono riportati nelle figure 4A e 4B. Sono stati riscontrati valori di eterozigosità significativamente inferiori per la popolazione Ogliastrina (BCS) rispetto alle altre due popolazioni. Nella popolazione del Sulcis (BPU1) è stata individuata una differenza significativa tra il cromosoma 22 e il 15, con un valore superiore nel 15. Considerando le classi di frequenza, quella maggiormente rappresentata comprende i valori 0.4 – 0.5 di eterozigosità. La popolazione nella quale questa classe è meno rappresentata è, in accordo con il valore medio di eterozigosità, quella Ogliastrina (BCS).

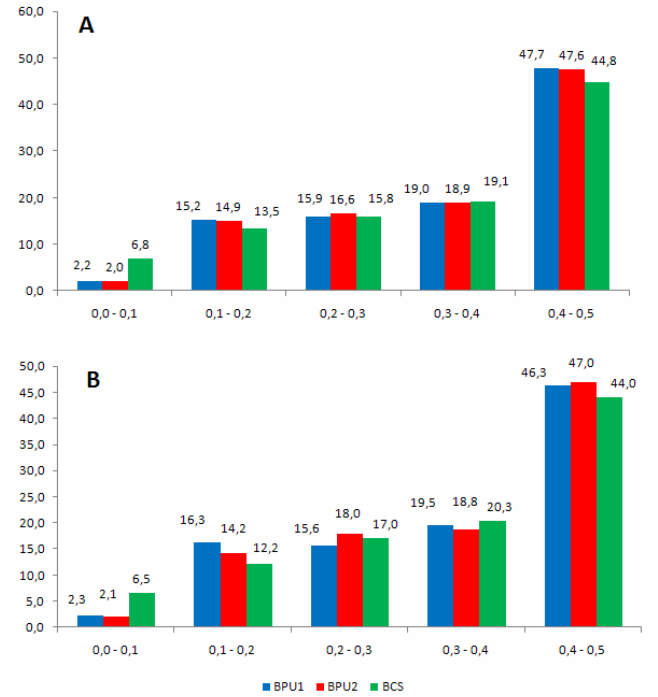

FIGURE 4.

Distribuzione dei valori di eterozigosità per 5 classi di frequenza nel cromosoma 15 e nel cromosoma 22. In ascissa sono riportate le classi di frequenza, in ordinate le percentuali per ogni popolazione.

I risultati relativi al PIC (Polymorphic information content), che rappresenta un parametro che misura il grado di informatività di un marcatore per la mappatura genetica, sono riportati nelle figure 5A e 5B. La maggior parte dei polimorfismi possiede un valore di PIC compreso tra 0.4 e 0.5 rivelando una elevata infromatività dei marcatori. L'unica differenza riscontrabile tra le diverse

[2] *L'eterozigosità (H) corrisponde alla quota delle localizzazioni genetiche (locus) di un individuo che contengono un marcatore che possiede più di una forma (allele). Supponiamo di aver stabilito il genotipo per il locus A di individui tratti da una popolazione e di aver trovato che la frequenza degli eterozigoti è 0,09: l'eterozigosità per questo locus sarebbe allora pari a 0,09. Per ottenere una stima dell'eterozigosità della popolazione si fa una media di questa eterozigosità con quella riscontrata a carico di altri loci.*

popolazioni è la presenza di una percentuale lievemente minore nella classe 0.4-0.5 individuabile nella popolazione dell'Ogliastra (BCS) rispetto alle altre due popolazioni.

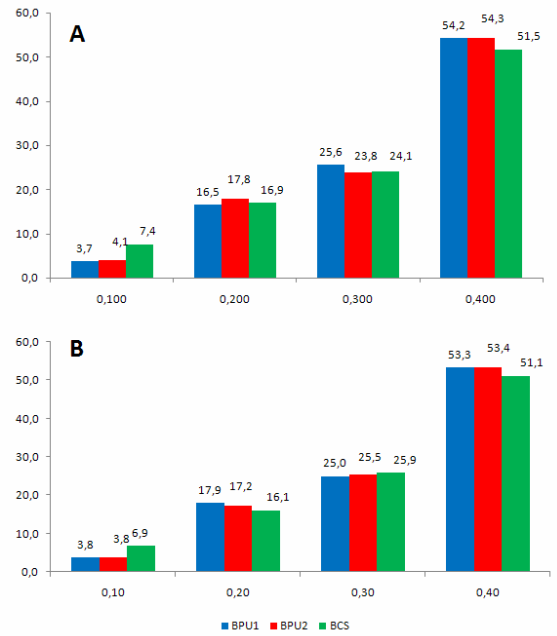

FIGURE 5.

Distribuzione dei valori di PIC per il cromosoma 15 (A) e per il cromosoma 22 (B). In ascissa sono riportate le classi di frequenza e in ordinata le percentuali per ogni popolazione.

Gli ultimi parametri calcolati sono quelli relativi all'analisi del Linkage Disequilibrium. Il linkage disequilibrium (LD) è un termine che indica la presenza di associazione statistica tra specifici alleli relativi a due o più loci, che costituiscono di solito un particolare combinazione ancestrale, diffusa nella popolazione in cui è rilevata perché trasmessa lungo la discendenza da un comune progenitore. Il linkage disequilibrium è un importante strumento per individuare regioni cromosomiche di limitata ampiezza in cui si collocano i geni per una data malattia (mappaggio ad alta risoluzione) e si avvale dell'analisi molecolare di varianti alleliche (per lo più di SNPs o STRs). Infatti è prevedibile che pazienti che hanno ereditato lo stesso segmento cromosomicoabbiano ereditato anche la stessa mutazione in esso contenuto.

Il D' e r^2, che rappresentano due classiche misure per la valutazione del LD, sono stati calcolati dividendo i cromosomi in regioni di 500 kilobasi ciascuna, (overlapping 250 kb). Le medie delle matrici di D' e r^2 , calcolate per ogni regione, sono riportate nel grafico in funzione della distanza di mappa.

I risultati relativi ai valori di Linkage Disequilibrium (LD) sono mostrati in figura 5. In particolare, sono stati messi a confronto i risultati relativi al cromosoma 22 (A) e al cromosoma 15 (B). Il

confronto dei grafici ci permette di osservare che gli andamenti di questi parametri tra le diverse popolazioni non presentano differenze sostanziali, benché si possano individuare delle variazioni in alcune regioni ristrette. Per quanto riguarda i valori medi di D' e r^2, la popolazione proveniente dall'Ogliastra (BCS) ha mostrato dei valori leggermente superiori rispetto alle altre due popolazioni. Considerando i valori di D' e r^2 tra le diverse popolazioni, per il cromosoma 15 risulta che BCS>BPU2>BPU1 mentre per il cromosoma 22 risulta che BCS>BPU1>BPU2. Infine, nelle stesse popolazioni tra i due cromosomi non sono state rilevate differenze sostanziali tra i valori di D' e r^2.

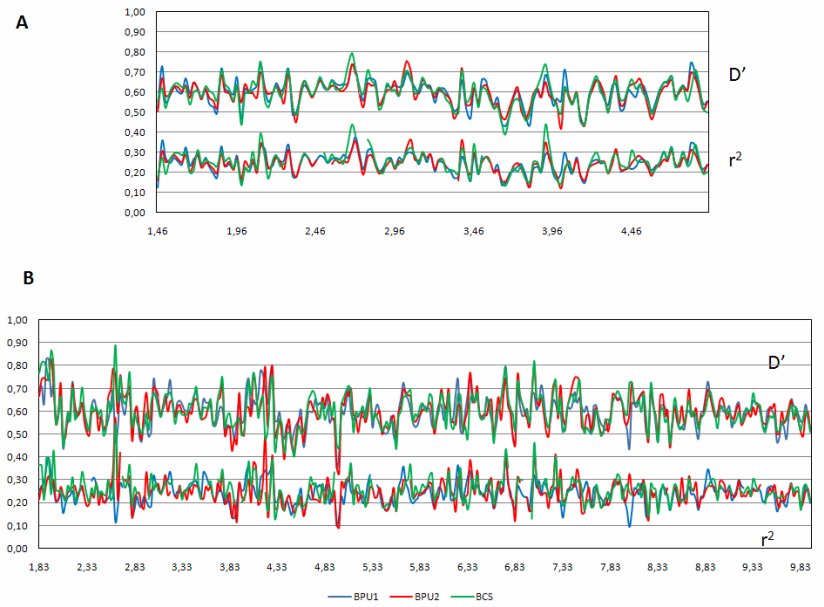

FIGURA 5.

Valori di D' e r^2 nei cromosomi 15 (A) e 22 (B). I cromosomi sono stati divisi in regioni di 500 kilobasi (overlap 250 kilobasi). In ascissa è rappresentata la distanza di mappa in CentiMorgan e in ordinata i valori di D' e r^2 per ogni regione.

In conclusione è emersa, come rilevato da articoli scientifici internazionali, una sub struttura della popolazione sarda con particolare riferimento alla zona dell'Ogliastra che presenta una maggiore omogeneità genetica (vedi ad es. i valori di eterozigosità, etc) verso una variabilità genetica maggiore nei due campioni provenienti dal Sulcis e dalla Trexenta.

L'obbiettivo primario del progetto di creare un campione di confronto di soggetti sani da utilizzare in studi di associazione per malattie comuni e/o complesse è stato raggiunto, almeno per ciò che riguarda la parte del sud dell'isola, e i dati delle analisi effettuate suggeriscono meglio quali debbano essere le caratteristiche territoriali più adatte per selezionare i campioni di affetti da confrontare. Inoltre danno una stima di quelle che potrebbero essere le differenze create dalla localizzazione geografica dei gruppi di pazienti da analizzare e consentono una selezione più mirata per evitare la possibilità di individuare dei segnali statistici spuri. In dipendenza di quali tipi

di analisi e quali malattie e loro frequenze vogliano essere indagate sarebbe necessario incrementare il numero di campioni sani da utilizzare come confronto ma il lavoro effettuato sinora rappresenta sicuramente un buon campione di partenza per effettuare delle analisi preliminari.