Università degli studi di Sassari
Dipartimento di Scienze Biomediche
Viale San Pietro 43/B
07100, Sassari (SS)

Titolo del protocollo:
Studio GENOMico della Malattia di Alzheimer e nelle Demenze associate all'invecchiamento nella popolazione fondatrice sarda: comprensione delle basi eziopatogenetiche e individuazione di bersagli terapeutici - GENOMAD.

Responsabile della Ricerca:
Francesco Cucca
Professore di Genetica Medica presso l'Università degli Studi di Sassari.

Titolo del progetto	***Studio GENOMico della Malattia di Alzheimer e nelle Demenze associate all'invecchiamento nella popolazione fondatrice sarda: comprensione delle basi eziopatogenetiche e individuazione di bersagli terapeutici - GENOMAD***
Centro coordinamento	Università degli Studi di Sassari, Viale San Pietro, Complesso Biologico, Genetica, 07100 Sassari e Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche. *Cittadella Universitaria di Monserrato, S.S. 554 bivio per Sestu Km 4,500 – 09042 Monserrato (CA)*
Responsabile scientifico	*Prof. Francesco Cucca, Università di Sassari*

INDICE

Introduzione e Razionale	3
Obiettivi	4
Obiettivi generali	4
Obiettivi specifici	4
Disegno dello studio	5
Reclutamento pazienti e controlli	5
Criteri di inclusione ed esclusione	6
Raccolta dati e caratterizzazione genetica	7
Calcolo del potere statistico e giustificazione della numerosità campionaria	7
Analisi statistica	7
Anonimato e riservatezza dei dati personali	10
Informazione e consenso informato	10
Conservazione e sicurezza dei dati raccolti e dei risultati della ricerca	10
Bibliografia	10

PROTOCOLLO DI STUDIO

INTRODUZIONE E RAZIONALE

Le peculiari caratteristiche genetiche degli abitanti della Sardegna, legate all'insularità e alla loro storia demografica plurimillenaria, rendono questa popolazione un *unicum* particolarmente adatto alla realizzazione di studi sistematici volti a identificare fattori di rischio per malattie con una base genetica. Una di queste patologie è la malattia di Alzheimer, una malattia multifattoriale causata da molteplici fattori genetici con il concorso di fattori ambientali permissivi ancora in gran parte sconosciuti. La malattia di Alzheimer è la forma più frequente di demenza, il cui rischio aumenta con l'invecchiamento, e in cui la maggior parte dei pazienti ha un'età superiore ai 65 anni. Si tratta di una malattia cronica e invalidante senza cure efficaci che rappresenta un notevole problema da un punto di vista sanitario e sociale in tutto il mondo occidentale, dove l'aspettativa media di vita si sta allungando. Esiste quindi un vasto interesse alla comprensione delle cause di questa malattia e all'individuazione di bersagli terapeutici per la messa a punto di nuove terapie.

Partendo da questi presupposti, è stato concepito il progetto "**Studio GENOMico della Malattia di Alzheimer e nelle Demenze associate all'invecchiamento nella popolazione fondatrice sarda: comprensione delle basi eziopatogenetiche e individuazione di bersagli terapeutici – GENOMAD",** proposto dal **Dipartimento di Scienze Biomediche dell'Università di Sassari** e ammesso al finanziamento dal Ministero della Salute. Il progetto si svolgerà in stretta collaborazione con i principali centri per la diagnosi e cura dei disturbi cognitivi associati all'invecchiamento nel territorio sardo e per la parte scientifica con l'Istituto di Ricerca Genetica e Biomedica (IRGB) del CNR diretto dalla prof.ssa Marcella Devoto, con sede legale a Cagliari e sedi di lavoro a Sassari e Lanusei. In quest'ultima sede, di cui è responsabile il dott. Edoardo Fiorillo, si svolge lo studio SardiNIA/ProgeNIA e saranno condotte alcune attività sperimentali alle quali collaboreranno ricercatori e personale medico e infermieristico dell'IRGB-CNR e dell'Università di Sassari.

Il progetto GENOMAD, che si svolgerà nell'arco di quattro anni dal 13 febbraio 2023 al 12 febbraio 2027, ha come responsabile scientifico il Prof. Francesco Cucca, ordinario di Genetica Medica dell'Università di Sassari, e prevede il coinvolgimento di tutti i centri clinici che si occupano di disturbi cognitivi associarti all'invecchiamento in Sardegna. Il responsabile dello studio prof. Francesco Cucca ha ottenuto il finanziamento per condurre tale ricerca dal Ministero della Salute, Direzione Generale della Ricerca e dell'Innovazione in Sanità, Traiettoria 3 "Medicina rigenerativa, predittiva e personalizzata" - Linea di azione 3.1 "Creazione di un programma di medicina di precisione per la mappatura del genoma umano su scala nazionale" - Piano Sviluppo e Coesione Salute Codice locale progetto T3-MZ-16. Il progetto ha l'obiettivo generale di acquisire dati clinici, genetici e biochimici su 8000 pazienti sardi affetti da malattia di Alzheimer e demenze associate all'invecchiamento, con una forte enfasi su forme iniziali di tali patologie, e altrettanti controlli. I partecipanti allo Studio GENOMAD provengono principalmente dalle strutture regionali (elencate sotto) facenti capo all'Azienda per la Tutela della Salute (ATS) e a partire dal 1° gennaio 2022 alle ricostituite ASL, alle Aziende Ospedaliere Universitarie (AOU) di Sassari e Cagliari e all'Azienda Ospedaliera Brotzu, dove verranno collezionati sia i dati clinici che i campioni biologici di sangue e di siero per effettuare le analisi genetiche e per quantificare molecole chiave nei processi eziopatogenetici della neurodegenerazione. È importante rilevare che il progetto non prevede somministrazione di farmaci.

La Sardegna è un antico isolato genetico, infatti il suo isolamento millenario e la crescita demografica incominciata nel Neolitico in assenza di flussi immigratori rilevanti (in particolare quello proveniente dalla "Steppa" avvenuto circa 3500 anni fa da parte di popolazioni di pastori che arrivavano dall'Asia occidentale e che si sono mescolate con le altre popolazioni locali in Europa diffondendo la lingua

proto-indoeuropea) hanno fatto sì che la popolazione sarda attuale abbia un patrimonio genetico più conservato rispetto alla popolazione iniziale fondatrice in confronto ad altre popolazioni europee. La struttura genetica dei sardi rappresenta quindi meglio che altrove le componenti più antiche, in particolare quella Neolitica, dell'architettura genetica degli europei contemporanei, grazie a numerosi effetti fondatori che hanno determinato il fissarsi di alleli ed aplotipi a media ed elevata frequenza che per contro sono rari o assenti in altre popolazioni [Sidore C et al, Nat Genet. 2015; Chiang CWK et al, Nat Genet. 2018; Marcus JH et al, Nat Commun. 2020]. Le peculiarità genetiche della popolazione sarda sono state osservate in diversi sistemi genetici ed esiste un'elevata probabilità che effetti fondatori coinvolgano varianti predisponenti nei confronti di diverse malattie [Steri M et al, NEJM 2017; Busonero F et al, Br J Haematol. 2020].

Un'altra caratteristica della popolazione sarda è rappresentata da una distribuzione relativamente omogenea della variabilità genetica in diverse subregioni dell'isola che minimizza la presenza di sub-struttura genetica che, in altre popolazioni, può confondere e rendere più difficoltosa la ricerca di geni di suscettibilità attraverso studi di associazione con disegno sperimentale caso-controllo [Francalacci P et al, Science. 2013].

Per tutte queste caratteristiche, la popolazione sarda è altamente informativa per capire, attraverso studi di associazione sistematici sull'intero genoma su malattie multifattoriali comuni nell'isola, il funzionamento dei geni e dei loro prodotti proteici nello stato di salute o per contro il loro malfunzionamento in diverse condizioni patologiche. L'isolamento che fino a ieri è stato considerato uno svantaggio, diventa oggi un formidabile strumento per la ricerca di geni implicati nello sviluppo di malattie, e la popolazione sarda rappresenta una sorta di "laboratorio naturale" per gli studi genetici.

OBIETTIVI

Obiettivi generali

L'obiettivo generale primario della nostra ricerca è rivolto all'identificazione di varianti geniche di suscettibilità ai disturbi cognitivi associati all'invecchiamento, inclusa la malattia di Alzheimer, di particolare interesse per l'aumento della loro frequenza a livello globale e nella popolazione sarda, e mira alla comprensione dei meccanismi influenzati dalle varianti primariamente coinvolte nella manifestazione patologica. L'identificazione delle varianti geniche di suscettibilità e la comprensione del loro ruolo biologico nella suscettibilità alle demenze potrebbero evidenziare molecole target utili allo sviluppo di nuovi farmaci.
Lo studio ha anche un **obiettivo generale secondario**: contribuire al corretto inquadramento diagnostico delle differenti forme di demenza e garantire un'omogeneità dell'approccio diagnostico nei diversi centri che operano nel territorio regionale, promuovendo una stretta collaborazione tra ricercatori e personale del sistema sanitario regionale dedicato allo studio delle demenze.

Obiettivi specifici

L'obiettivo generale primario verrà realizzato attraverso i seguenti obiettivi specifici.
-i) Caratterizzare il genoma di 8000 pazienti con varie forme di disturbi cognitivi associati all'invecchiamento e 8000 controlli, per identificare attraverso uno studio caso-controllo di associazione su tutto il genoma (genome wide association study, o GWAS) le variazioni presenti nella sequenza del DNA (dette varianti genetiche) in grado influenzare la predisposizione o per contro la protezione nei confronti delle demenze.
-ii) Caratterizzare la coorte SardiNIA/ProgeNIA in Ogliastra -- costituita da circa 8000 soggetti della popolazione generale già profilati a livello genomico -- per parametri quantitativi potenzialmente

correlati con la neurodegenerazione cerebrale alla base delle demenze, e la realizzazione di GWAS per ciascuno dei parametri testati.
-iii) Ricerca di sovrapposizioni ("associazioni coincidenti") tra le varianti genetiche associate al rischio di sviluppare demenza identificate nel primo obiettivo intermedio e quelle in grado di influenzare la variabilità delle variabili quantitative identificate nel secondo obiettivo intermedio. Questo permetterà di mettere in evidenza i cosiddetti fenotipi intermedi, ovvero caratteri quantificabili localizzati a livello di punti di controllo chiave (checkpoint) lungo la catena di eventi molecolari che portano alla demenza.
-iv) Individuare retrospettivamente nei volontari della coorte SardiNIA - grazie ad una raccolta longitudinale di materiale biologico e dati medici avvenuta negli ultimi 20 anni - eventuali cambiamenti significativi nei livelli di variabili quantitative (selezionate nell'obiettivo precedente) in un periodo che precede l'insorgenza di deficit cognitivo.
-v) Caratterizzare a livello genomico e raccogliere campioni biologici per analisi immunologiche, proteomiche e trascrittomiche, da almeno di 1000 pazienti (degli 8000 citati nell'obiettivo 1) con deficit cognitivo modesto (Mild Cognitive Impairment o MCI) che saranno monitorati clinicamente nel tempo per effettuare un GWAS longitudinale volto ad identificare le varianti genetiche implicate nella progressione verso la demenza franca e ricercare biomarcatori di regressione cognitiva.
-vi) Ci proponiamo inoltre di studiare periodicamente un sottogruppo di pazienti con MCI da un punto di vista neuro-fisiologico per identificare deviazioni dalle normali traiettorie di declino cognitivo nelle fasi precoci della malattia e individuare e monitorare nel tempo segni prodromici che informino su possibili transizioni dal mild cognitive impairment alla malattia conclamata.

DISEGNO DELLO STUDIO

Reclutamento pazienti e controlli

Lo studio caso-controllo per l'identificazione di associazioni genetiche con le diverse forme di demenza prevede la raccolta di circa 8000 casi di pazienti con malattia di Alzheimer e altre forme di demenze correlate e 8000 controlli sani che saranno reclutati e visitati nell'arco di cinque anni. I partecipanti sono distribuiti nell'intero territorio regionale e il numero è proporzionale al bacino di utenza dei centri clinici coinvolti. Le aree di raccolta principali sono quattro: 1. Sassari/Olbia, 2. Nuoro/Oristano/Lanusei, 3. Carbonia/Iglesias e 4. Cagliari. Il limite superiore della dimensione del campione di 8000 casi e 8000 controlli si basa su numerose riunioni intermedie con i rappresentanti dei principali centri per lo studio delle demenze che hanno evidenziato come nel territorio Regionale esistano già circa 20.000 casi di demenza che rappresentano il 5-8% della popolazione sopra i 65 anni di età. Sulla base dei dati forniti dai centri per lo studio delle demenze, abbiamo stimato che 4 centri di reclutamento principali durante 200 giorni lavorativi all'anno e raccogliendo una media di 5 pazienti e controlli al giorno saranno in grado di campionare facilmente più di 8000 casi e 8000 controlli durante i 4 anni dello studio GENOMAD: (5 pazienti e controlli) x 200 giorni/anno x 4 anni x 4 centri = 16000.
Le visite dei pazienti e dei controlli arruolati nello studio saranno eseguite presso le strutture cliniche appartenenti al servizio sanitario nazionale (o ad esso associate come le Università di Cagliari e Sassari) all'interno delle quattro macroaree già indicate. La scelta di questi luoghi di reclutamento è stata effettuata durante gli incontri di messa a punto dello studio che hanno coinvolto gli esponenti di tutte le strutture sarde impiegate nella clinica delle patologie associate alle demenze.
La fonte principale di "*selection bias*" negli studi di associazione genetica è legata alla possibilità che esista una struttura genetica differente tra casi e controlli dovuta ad una loro differente origine etnico-geografica. È noto, ad esempio, che alcuni alleli sono maggiormente rappresentati in specifiche popolazioni come quella sarda rispetto ad altre popolazioni. Per questo motivo nello studio si arruoleranno esclusivamente individui di origine sarda. A questo scopo saranno richieste informazioni sui luoghi d'origine dei quattro nonni di ciascun partecipante, e il loro arruolamento è

stato pianificato sull'intero territorio regionale per garantire una omogenea rappresentazione dei pazienti e controlli nelle varie aree dell'isola. Per determinare e correggere un'eventuale stratificazione di popolazione residua tra casi e controlli sarà inoltre effettuata un'analisi delle componenti principali (PCA) come descritto più avanti.
Inoltre, negli studi genetici di associazione caso-controllo su tratti binari (come quelli per identificare i fattori di rischio per una determinata malattia), un requisito primario per garantire la validità dei test di associazione è rappresentato dall'indipendenza (assenza di correlazione genetica stretta) tra gli individui comparati. Nel nostro studio caso-controllo sulla malattia di Alzheimer particolare attenzione sarà quindi rivolta ai criteri di selezione dei controlli sani al fine di assicurare l'assenza di correlazione genetica con i casi. Per garantire questo requisito preleveremo preferenzialmente i coniugi o partner dei pazienti che comunemente non hanno rapporti di parentela stretta tra di loro. L'utilizzo dei coniugi come popolazione di controllo preferenziale per uno studio genetico offre anche il vantaggio della condivisione delle abitudini e degli stili di vita tra controlli e malati consentendo quindi di controllare l'effetto di queste variabili sul rischio di malattia. Per assicurare lo stesso numero di controlli e casi compensando situazioni in cui i pazienti non hanno coniugi reclutabili per lo studio, arruoleremo un ulteriore gruppo di controlli distribuiti equamente per età e sesso rispetto ai casi e controllando le variabili succitate. Tutti i controlli comunque selezionati verranno sottoposti a valutazione neuropsicologica analoga a quella effettuata nei malati che ci consentirà di escludere deficit cognitivi di vario tipo.
L'indagine psicometrica dei partecipanti comprende la somministrazione di test globali come il MMSE e test più specifici che consentono di quantificare i deficit cognitivi e la loro evoluzione nel tempo come il test MOCA (*MOntreal Cognitive Assessment*) che comprende una serie di prove di funzioni visuo-spaziali/esecutive. Sarà anche effettuato l'Esame Obiettivo Neurologico completo con valutazione neuropsicologica indirizzata a esaminare domini cognitivi quali memoria, linguaggio, funzioni esecutive, funzioni visuo-spaziali, funzioni prassiche e gnosiche.

In parallelo con il reclutamento di pazienti con demenza franca, intendiamo includere nello studio almeno 1000 pazienti affetti da un deficit cognitivo più modesto, o MCI, che saranno valutati clinicamente ogni anno per cinque anni per effettuare al termine di tale periodo uno studio GWA volto ad identificare le varianti genetiche associate con la progressione verso la demenza manifesta. Dagli stessi individui saranno raccolti campioni biologici per la caratterizzazione di fenotipi immunologici sia cellulari che solubili, e molecolari nel sangue periferico e follow-up longitudinale di soggetti con mild cognitive impairment per la ricerca di fattori in grado di biomarcatori di progressione o mancata progressione.
Recenti evidenze indicano che alterazioni cognitive e alterazioni sensorie-motorie non cognitive, spesso sub-cliniche, possono precedere di diversi anni le manifestazioni cliniche conclamate dell'AD. In un sub-gruppo di 100 pazienti con MCI accertati nelle fasi iniziali del progetto ogni 6 mesi si effettuerà quindi un'accurata valutazione neurofisiologica per individuare e monitorare nel tempo segni prodromici che informino precocemente su possibili transizioni dal MCI alla malattia conclamata. A tal fine sarà effettuato il monitoraggio di numerose funzioni fisiologiche (come le funzioni autonomiche, pattern del sonno, articolazione della parola e linguaggio, movimenti oculari, risposte pupillari, mimica facciale, metabolismo e pattern nutrizionale).

Infine, il disegno sperimentale prevede uno studio retrospettivo per la determinazione di associazioni genetiche con variabili quantitative potenzialmente correlate con la demenza sui campioni biologici già raccolti in circa 8000 volontari della coorte SardiNIA reclutati a partire dal 2001. Poiché questa coorte è stata ripetutamente valutata e campionata nel siero e per test cognitivi, la disponibilità di questi dati offre l'opportunità di effettuare una valutazione retrospettiva dei marcatori degenerativi e di altri tratti quantitativi in un numero di individui che hanno sviluppato successivamente malattia di Alzheimer o demenze associate alla malattia di Alzheimer, fornendo informazioni chiave sulla storia naturale della malattia.

Criteri di inclusione ed esclusione

I partecipanti allo studio sono pazienti affetti da malattia di Alzheimer e altre forme di demenza degenerativa seguiti nei centri di riferimento isolani ospedalieri, universitari e misti. Gli individui con disabilità cognitive tali da non essere in grado di comprendere pienamente gli obiettivi dello studio e di esprimere un consenso informato possono partecipare se il consenso al trattamento è acquisito da chi esercita legalmente la potestà, ovvero da un prossimo congiunto, da un familiare, da un convivente o, in loro assenza, dal responsabile della struttura presso cui dimora l'interessato. (Autorizzazione n. 8/2016 - Autorizzazione generale al trattamento dei dati genetici - 15 dicembre 2016, articolo 6 (Pubblicato sulla Gazzetta Ufficiale n. 303 del 29 dicembre 2016 – Provvedimento recante le prescrizioni relative al trattamento di categorie particolari di dati, ai sensi dell'art. 21, comma 1 del d.lgs. 10 agosto 2018, n. 101, Pubblicato sulla Gazzetta Ufficiale Serie Generale n. 176 del 29 luglio 2019).
La popolazione oggetto di studio comprende esclusivamente soggetti con chiara origine sarda (almeno due generazioni, da accertarsi in sede di anamnesi), e saranno adottati i seguenti criteri di inclusione ed esclusione.

Criteri di inclusione:

1. Malattia di Alzheimer e tutte le altre forme di demenza degenerativa sia sporadiche che familiari.

Inoltre, sono stati precisati i criteri di diagnosi clinica da utilizzare per le principali forme di demenza degenerativa:
- Malattia di Alzheimer Probabile: riferimento alla revisione dei criteri NINCDS-ADRDA del 1984 pubblicata nel maggio 2011 che mantiene il concetto fondamentale di diagnosi clinica, supportato e validato dall'evidenza biochimica e dalle tecniche di neuroimaging e che tiene conto delle varianti atipiche della malattia [McKhann GM et al, Neurology 1984; McKhann GM et al, Alzheimers Dement 2011].
- Demenza Fronto-temporale variante comportamentale (bvFTD) [Rascovsky K et al, Brain 2011].
- Afasia Progressiva Primaria [Gorno-Tempini ML et al, Neurology 2011].
- Malattia a Corpi di Lewy [McKeith IG et al, Neurology 2005].

Criteri di esclusione:

1. Depressione maggiore (escludere le forme attuali e persistenti/croniche) ed altre forme di psicosi maggiore
2. Malattia cerebrovascolare in riferimento alla Demenza Vascolare
3. Lesione cerebrale traumatica
4. Idrocefalo normoteso
5. Patologia neoplastica sia primitiva che secondaria dell'encefalo
6. Vasculiti cerebrali o malattie autoimmuni sistemiche con interessamento encefalico (LES e altre malattie del connettivo)
7. Infezione da HIV
8. Encefalopatia epatica porto-sistemica in corso di cirrosi epatica
9. Malattia di Wilson
10. Insufficienza renale cronica dialitica
11. Etilismo cronico e tossicosi esogene
12. Sindromi carenziali e/o da alterato assorbimento intestinale
13. Insufficienza respiratoria cronica con alterati scambi emogasanalitici, ipossiemiche e/o ipercapniche
14. Gravi condizioni di anemia
15. Scompenso cardiaco grave

16. Alterazioni endocrine in condizioni di scompenso, gravi turbe idroelettrolitiche e dell'equilibrio acido-base.

Raccolta dati e caratterizzazione genetica

La scheda raccolta dati paziente (caratteristiche anagrafiche, demografiche e cliniche), compilata dal medico che arruola il paziente, saranno registrati centralmente presso un database elettronico appositamente predisposto e gestito in collaborazione dal Dipartimento di Scienze Biomediche dell'Università di Sassari e dall'IRGB-CNR. I documenti cartacei saranno conservati in sicurezza in appositi armadi. Ogni mese i campioni raccolti saranno portati presso una biobanca centralizzata localizzata in locali messi a disposizione dei partecipanti allo studio dal comune di Lanusei per la conservazione definitiva a -80 gradi dei campioni di siero e plasma e l'estrazione di DNA dai linfociti e/o RNA da PBMC totali o sottocategorie di linfociti. Tutti i campioni dello studio saranno conservati in condizioni appropriate che ne garantiscono la qualità necessaria per gli studi e per gli esperimenti descritti nel protocollo di studio. Il DNA di tutti i partecipanti allo studio sarà analizzato mediante genotipizzazione su piattaforma Illumina disegnata appositamente per i successivi studi di associazione oggetto dello studio. Specificamente, tutti i campioni verranno caratterizzati mediante l'utilizzo del *Global Screening Array* (GSA), che permette lo studio di 660,000 marcatori predefiniti oltre che 50,000 marcatori *custom* ed eventuali altri *chips* informativi disponibili e utili ai nostri scopi.

Calcolo del potere statistico e giustificazione della numerosità campionaria

Per quanto riguarda gli studi di associazione genetica, la dimensione di questo dataset ci permetterà di identificare sia varianti associate a malattia frequenti in Sardegna ma con effetti relativamente piccoli, che varianti più rare con effetti forti. Più in dettaglio, considerata una soglia di significatività statistica stringente pari a un *p-value* < $5x10^{-8}$, tradizionalmente utilizzata nei GWAS, e un potere statistico >80%, ci aspettiamo di individuare:

- varianti comuni in Sardegna (con frequenza allelica >5%) associate a malattia con effetti relativamente modesti (OR >1.3);
- varianti meno comuni (frequenza allelica compresa tra 2% e 5%), con effetti più forti (OR compresi tra 1.4 e 1.6);
- varianti più rare (frequenza allelica compresa tra 1% e 2%), con effetti forti (OR>1.9);
- varianti rare (frequenza compresa tra 0.5% e 1%) con effetti molto forti (OR >2.3). A un potere statistico del 50%, ci aspettiamo di poter individuare anche varianti rare con OR>2.1.

Lo studio dei parametri quantitativi verrà condotto sulla popolazione generale del progetto SardiNIA, che comprende circa 8000 campioni distribuiti in circa 1500 famiglie, già reclutati e caratterizzati dal punto di vista genetico. Alle stesse soglie di significatività e potere statistico rispetto al caso-controllo (*p-value* < $5x10^{-8}$ e potere statistico >80%), anche in questo caso ci aspettiamo di individuare associazioni con i fenotipi di interesse aventi effetti normalizzati (beta, espressi in numero di deviazioni standard) via via più grandi al decrescere della frequenza allelica:

- beta > 0.03 per varianti con frequenza >5%;
- beta compresi tra 0.4 e 0.5 per varianti con frequenza compresa tra 2% e %5;
- beta compresi tra 0.6 e 0.7 per varianti con frequenza compresa tra 1% e 2%;
- beta > 1.1 per varianti con frequenza tra 0.5% e 1%; in questo intervallo di frequenza, abbiamo un potere statistico del 50% di individuare anche varianti con effetti meno forti (beta >0.9).

Analisi statistica

Analisi genetica dello studio caso-controllo: i dati provenienti dalla genotipizzazione dei casi e dei controlli mediante *Global Screening Array* (GSA) Illumina eseguita presso la sede IRGB di Lanusei saranno sottoposti a stringenti controlli di qualità, sia a livello di campione che a livello di variante genetica (*call rate* >98%, deviazione da equilibrio di Hardy-Weinberg (HWE) *p-value* > 10^{-6} nei controlli). Un'analisi delle componenti principali (PCA) applicata alle varianti con frequenza

maggiore dell'1% consentirà di determinare e correggere eventuale stratificazione di popolazione tra casi e controlli. Le varianti genetiche che avranno superato i controlli di qualità verranno usate come *baseline* per la generazione di una mappa genetica ad alta risoluzione: mediante metodi di inferenza statistica detti di "imputazione", le varianti genotipizzate verranno integrate con quelle caratterizzate con il sequenziamento dell'intero genoma di circa 3500 individui sardi, con lo scopo di ottenere, per ciascun paziente e per ciascun controllo, una sequenza ad altissima risoluzione comprendente circa 20 milioni di varianti di alta qualità. Sulla base della mappa genetica così costruita, verrà condotto un GWAS caso-controllo, opportunamente corretto per fattori confondenti (sesso, età, PCA) e che ci permetterà di individuare, a soglie stringenti di significatività (*p-value* $<5x10^{-8}$), una serie di loci di rischio o protettivi per la patologia [Steri M et al, NEJM 2017]. Ad ogni locus, infine, verrà individuato, mediante analisi di mappaggio fine, l'insieme delle varianti aventi un potenziale ruolo causale nella patologia (*credible set*), da usarsi nelle successive analisi di correlazione genetica con tratti quantitativi [Orrù V et al, Nat Genet. 2020].

Analisi di genetica quantitativa su parametri sierici e cellulari nella coorte SardiNIA: l'impatto della variazione genetica sui tratti fenotipici molecolari e cellulari caratterizzati nella coorte SardiNIA sarà affrontato anche in questo caso attraverso analisi di tipo GWAS. Preliminarmente agli studi genetici, per ogni fenotipo di interesse valuteremo la sua ereditabilità, che rappresenta la quantità di variazione fenotipica dovuta alla variazione genetica, mediante metodi standard basati sulle strutture familiari in cui i tratti saranno quantificati. Il calcolo dell'ereditabilità si basa sulle relazioni familiari tra i campioni oggetto di studio e la coorte ogliastrina, che comprende circa 1500 famiglie verticali, è particolarmente informativa per questo tipo di studio. Il GWAS sarà quindi eseguito sui tratti quantitativi ereditabili misurati ad ogni punto temporale separatamente e sui tratti derivati dalle curve longitudinali individuali utilizzando modelli lineari multilivello a effetti misti. Anche in questo contesto, i GWAS verranno seguiti da analisi di *fine-mapping* per individuare i *credible sets* [Orrù V et al, Nat Genet. 2020].
Utilizzando i dati retrospettivi della coorte, effettueremo inoltre un'analisi longitudinale dei tratti biochimici e immunologici di interesse ed eseguiremo un'analisi esplorativa per scoprire i modelli di una loro variazione sistematica nel tempo, così come gli aspetti di variazione casuale che distinguono gli individui. Per esempio, calcoleremo statistiche riassuntive, come medie e deviazioni standard, per osservare se diversi tratti stanno cambiando in modo simile o diverso. Le differenze tra gruppi e punti di tempo saranno testate con il test t di Student o Anova (per variabili normalmente distribuite) e con i test Mann-Whitney o Kruskal-Wallis (per variabili non distribuite normalmente). Le variabili discrete saranno trattate come conteggi e percentuali, e le differenze tra gruppi e *timepoint* saranno testate con il test del chi-quadrato. Per i tratti distribuiti normalmente, analizzeremo anche la varianza/covarianza (tramite Anova/Ancova) per tenere conto della correlazione tra i campioni diversi e all'interno degli stessi campioni per le misure ripetute; questo determinerà se la natura della variabilità osservata è casuale o sistematica. Utilizzando i dati ottenuti nei punti temporali valutati, applicheremo modelli lineari a effetti misti per valutare il modello temporale medio di ogni tratto fenotipico durante il *follow-up* dello studio. Le analisi descrittive saranno condotte sia nella coorte complessiva che nei maschi e nelle femmine separatamente, per valutare gli effetti sesso-specifici nel tempo, di cruciale importanza per le demenze che hanno una diversa frequenza tra i due sessi.

Analisi di correlazione genetica tra le associazioni provenienti dai GWAS caso-controllo e GWAS sui tratti quantitativi nella coorte SardiNIA. Allo scopo di individuare fenotipi intermedi/biomarcatori per le patologie oggetto di studio, sfrutteremo le statistiche ottenute dagli studi GWAS come fonte di "strumenti genetici" per giungere a conclusioni causali tra il fattore di rischio e la malattia. Tali tecniche includono la ricerca di associazioni coincidenti tratto-patologia mediante analisi di colocalizzazione e la successiva prioritizzazione di associazioni multiple coincidenti e coerenti tra fenotipo e malattia, nonché metodi di randomizzazione mendeliana, che permettono di

effettuare inferenza di causalità combinando molteplici associazioni genetiche in un unico modello [Orrù V et al, Nat Genet. 2020].

Analisi GWAS longitudinale di progressione "*case-only*" per stimare l'influenza della variabilità genetica nella progressione da MCI a malattia di Alzheimer manifesta. Includeremo nello studio non meno di 1000 casi di MCI per valutare al termine del periodo di osservazione le differenze genetiche tra individui che evolvono verso demenza franca e quelli che non evolvono, analogamente a quanto effettuato nello studio caso-controllo sopra descritto. Effettueremo anche un'analisi per stimare l'influenza della variabilità genetica sul tempo dall'inizio dell'osservazione fino alla progressione verso la malattia di Alzheimer. A tal fine eseguiremo un'analisi di sopravvivenza (modello di Cox a rischio proporzionale aggiustato per età all'esordio, sesso, anni di istruzione, e componenti principali della struttura della popolazione). Calcoleremo anche un punteggio di rischio poligenico (o *polygenic risk score*, PRS). Il PRS sarà calcolato come la somma ponderata del numero di alleli di rischio posseduti da un individuo, in cui il peso è rappresentato dal log naturale dell'*odds ratio* associato ad ogni singola variante genetica. Utilizzeremo separatamente le principali varianti note associate alla suscettibilità per la malattia di Alzheimer e quelle che emergeranno dal nostro GWAS di progressione per valutare eventuali sovrapposizioni e differenze tra fattori di rischio per l'Alzheimer e fattori di progressione a partire dal MCI.

Analisi di correlazione fenotipica tra i dati fenotipici e neuro-fisiologici raccolti periodicamente nei pazienti con MCI
I dati quantitativi e discreti rappresentati dai fenotipi immunologici, trascrittomici e neuro-fisiologici misurati nel corso dello studio in un subrgruppo di individui con diagnosi iniziale di MCI verranno inoltre correlati tra si loro per coppie di tratti utilizzando il coefficiente di correlazione di Spearman.

ANONIMATO E RISERVATEZZA DEI DATI PERSONALI

Ai partecipanti verrà garantito l'anonimato mediante l'adozione di diverse procedure:

- tutti i campioni verranno identificati con un codice numerico. Quest'ultimo verrà utilizzato su tutte le provette.
- I codici numerici verranno utilizzati per identificare i dati genetici. Le chiavi di tali codici saranno accessibili solo ai responsabili dello studio.

Verranno adottate misure di sicurezza per il rispetto della riservatezza dei dati che non sono immessi in rete e non sono divulgati all'esterno del laboratorio di ricerca.

INFORMAZIONE E CONSENSO INFORMATO

Tutti i partecipanti saranno informati dello scopo ed obiettivi dello studio e saranno altresì invitati a firmare copia del consenso informato (copie in allegato).
I soggetti partecipanti, qualora lo ritengano opportuno, possono revocare in qualsiasi momento il loro consenso allo studio. Tale opportunità è espressamente indicata nel consenso informato.
Nell'eventualità di un'interruzione prematura dello studio o qualora il soggetto ne faccia richiesta, i campioni di materiale biologico verranno smaltiti con l'utilizzo delle procedure previste per tutti i materiali biologici.
Il soggetto partecipante potrà espressamente richiedere la distruzione dei suoi campioni di materiale biologico utilizzando l'apposito modulo allegato al consenso informato.

CONSERVAZIONE E SICUREZZA DEI DATI RACCOLTI E DEI RISULTATI DELLA RICERCA

I dati raccolti saranno conservati dal gruppo di ricercatori proponenti lo studio, mentre i risultati saranno divulgati presso la comunità scientifica internazionale a mezzo stampa su riviste specializzate.
Dato il valore scientifico dei dati raccolti e l'importanza e lo sforzo necessario alla raccolta, i dati verranno conservati fino alla conclusione della ricerca in quanto non è attualmente prevedibile il tempo necessario alla sconfitta delle malattie oggetto di studio.
I dati sensibili verranno conservati sotto la responsabilità del Prof. Francesco Cucca in formato database elettronico con chiave d'accesso e schede cartacee presso i locali del centro di ricerca di Lanusei.

BIBLIOGRAFIA

1. *Sidore C, et al. Genome sequencing elucidates Sardinian genetic architecture and augments association analyses for lipid and blood inflammatory markers. Nat Genet. 2015 Nov;47(11):1272-1281.*

2. *Chiang CWK, et al. Genomic history of the Sardinian population. Nat Genet. 2018 Oct;50(10):1426-1434.*

3. *Marcus JH, et al. Genetic history from the Middle Neolithic to present on the Mediterranean island of Sardinia. Nat Commun. 2020 Feb 24;11(1):939.*

4. *Steri M, et al. Overexpression of the Cytokine BAFF and Autoimmunity Risk. N Engl J Med. 2017 Apr 27;376(17):1615-1626.*

5. *Busonero F, et al. A Sardinian founder mutation in glycoprotein Ib platelet subunit beta (GP1BB) that impacts thrombocytopenia. Br J Haematol. 2020 Dec;191(5):e124-e128.*

6. *Francalacci P, et al. Low-pass DNA sequencing of 1200 Sardinians reconstructs European Y-chromosome phylogeny. Science. 2013 Aug 2;341(6145):565-9.*

7. *McKhann GM, et al. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology. 1984 Jul;34(7):939-44.*

8. *McKhann GM et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 2011 May;7(3):263-9.*

9. *Rascovsky K et al. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. Brain 2011: 134; 2456–2477.*

10. *Gorno-Tempini ML et al. Classification of primary progressive aphasia and its variants. Neurology 76 March 15, 2011.*

11. *McKeith IG et al. Diagnosis and management of dementia with Lewy bodies: third report of the DLB Consortium. Neurology. 2005 Dec 27;65(12):1863-72.*

12. *Orrù V et al. Complex genetic signatures in immune cells underlie autoimmunity and inform therapy. Nat Genet. 2020 Oct;52(10):1036-1045.*