



**UNIVERSITA' DEGLI STUDI DI CAGLIARI
FACOLTA' DI MEDICINA E CHIRURGIA
Dipartimento di Scienze Biomediche e Biotecnologie**

**Dottorato di Ricerca (XX ciclo)
"Terapia Pediatrica e Farmacologia dello Sviluppo"
Coordinatore Scientifico: Prof. Renzo Galanello
MED 03**

**UTILIZZO DELLA TECNOLOGIA MICROCHIP PER
L'IDENTIFICAZIONE DI GENI CANDIDATI
RESPONSABILI DELL'AUMENTO DI HbF**

Tutor:
Prof. Renzo Galanello

Tesi di dottorato di:
Dott. Franco Anni

ANNO ACCADEMICO 2006-2007

Indice

Indice	1
Introduzione	2
1.1. La β -talassemia.....	2
1.2. GeneChip Human Mapping 500K Array.....	5
1.3. Copy Number Variation.....	7
Scopo della tesi	9
Materiali e Metodi	11
Risultati	19
Conclusioni	23
Bibliografia	24

Introduzione

1.1. La beta-talassemia

La beta-talassemia, uno dei più comuni disordini monogenici, è caratterizzata da ridotta o assente produzione di catene beta globiniche. Queste catene sono codificate da un gene localizzato sul cromosoma 11. I geni Beta codificano per le catene beta che formano, insieme a quelle alfa, l'emoglobina adulta umana HbA, una proteina tetramericata costituita da due catene alfa e due beta. Il gene Beta fa parte di un insieme di altri geni globinici (G-gamma, A-gamma, Epsilon e Delta) del cluster beta-globinico. Durante la vita fetale viene sintetizzata l'emoglobina fetale HbF, costituita da due catene alfa e due gamma. L'HbF viene sostituita, a partire soprattutto dalla 36^a settimana di gestazione, dall'HbA.

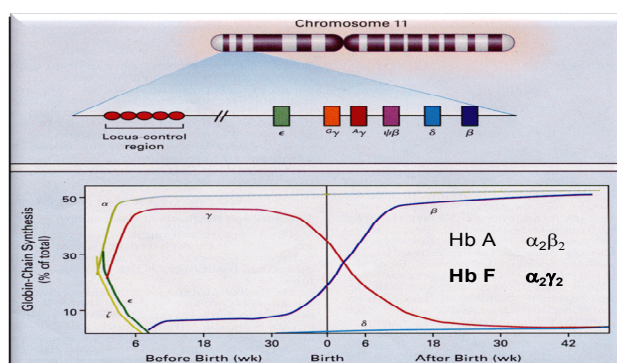


Figura 1. Sintesi delle catene emoglobiniche nei vari stadi di sviluppo

Il quadro clinico delle sindromi beta-talassemiche va dalla forma grave di anemia trasfusione dipendente (talassemia major), dovuta allo stato omozigote o a composti genetici per mutazioni beta talasemiche gravi, ad una relativamente lieve non trasfusione dipendente (talassemia intermedia), eterogenea sia a livello clinico che a livello molecolare. Poiché il maggior determinante della gravità del quadro clinico della beta talassemia è il grado di sbilanciamento della sintesi globinica dovuta ad un eccesso di catene alfa, ogni fattore capace di ridurre tale sbilanciamento sarà in grado di determinare un fenotipo attenuato (Galanello R. e coll. 1998).

I più importanti meccanismi capaci di produrre un quadro clinico di talassemia intermedia sono: la presenza di mutazioni beta talasemiche lievi o silenti con residua produzione di catene beta globiniche; la coereditarietà di alfa talassemia; la coereditarietà di determinanti genetici responsabili di una aumentata produzione di gamma catene. La condizione in cui è presente un aumento della produzione di emoglobina fetale (HbF) nella vita adulta è la cosiddetta Persistenza Ereditaria di Emoglobina Fetale (HPFH). I meccanismi attraverso i quali questo aumento di HbF può essere attivato non sono stati del tutto chiariti (Galanello R. e coll. 1989).

Solo alcuni di questi determinanti sono stati finora identificati: alcuni sono situati all'interno del cluster beta (Delta-Beta talassemia, Hb

Lepore, mutazione -196 C→T 5' nel gene A-gamma globinico, -158 C→T nel gene G-gamma globinico) altri nella Locus Control Region (LCR) responsabile dell'espressione dei geni beta-globinici (Talbot D. e coll. 1990) e altri al di fuori del cromosoma 11 (un locus nel cromosoma X (Xp22.2-22.3) (Dover e coll. 1992) e uno nel cromosoma 6 (6q22.3-23.1) (Craig e coll. 1996, Lapoumeroulie e coll. 1999).

La beta-talassemia, nella popolazione sarda, è quasi esclusivamente del tipo beta^o, con assenza di produzione di catene globiniche beta, dovuta alla mutazione non senso CAG→TAG al codon 39. A livello clinico la grande maggioranza degli omozigoti beta^o manifesta la forma grave di anemia, trasfusione dipendente (talassemia major), mentre circa il 10% presenta talassemia intermedia. I meccanismi responsabili di questo lieve fenotipo, a tutt'oggi, sono stati solo in parte definiti. Infatti solo alcuni dei pazienti con talassemia intermedia coereditano il genotipo -alfa/- alfa che riduce lo sbilanciamento alfa/non alfa e può quindi migliorare il grave quadro clinico della beta^o talassemia omozigote. Nella maggior parte dei pazienti è necessario ricercare altri determinanti responsabili dell'aumento della sintesi di catene gamma e quindi della produzione di HbF perché un suo aumento rende il quadro clinico più attenuato tale da rendere non necessaria la trasfusione.

1.2. GeneChip Human Mapping 500K Array

Il set Mapping 500K Array consiste di due arrays: un array utilizza l'enzima di restrizione NspI per genotipizzare 262,000 SNPs e un secondo array genotipizza altri 238,000 SNPs usando l'enzima Sty I.

Ogni array include più di 6.5 milioni di celle, ognuna contenente più di un milione di copie di probe oligonucleotidiche di 25 bp di una sequenza definita, sintetizzate in parallelo attraverso un processo fotolitografico. Ogni SNP è individuato da una specifica probe alla quale ne vengono aggiunte altre che controllano le basi adiacenti dell'SNP in esame confermando la validità del riconoscimento (Figura 1).

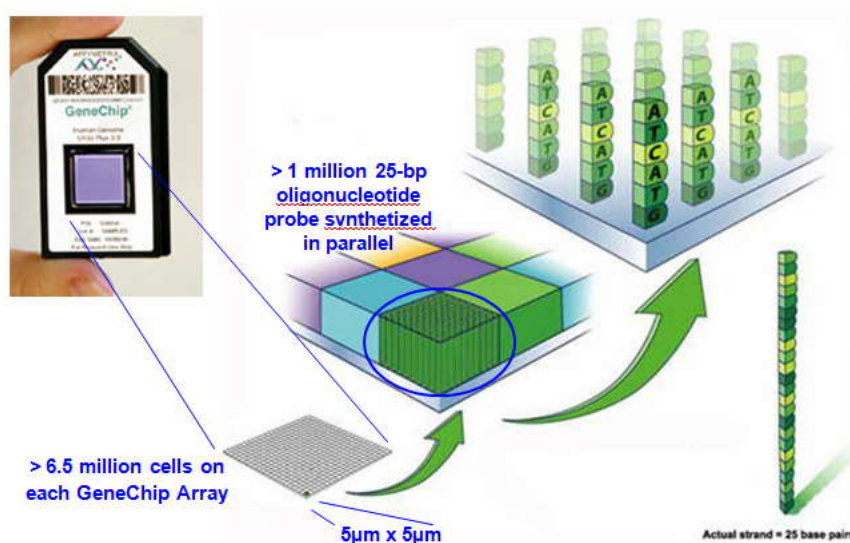


Figura 2. GeneChip Human Mapping 500K Array

Tutti gli SNPs del GeneChip Human Mapping 500K Array provengono da un rigoroso screening e un accurato processo di validazione. Gli SNPs sono stati selezionati in base all'alta riproducibilità,

all'elevata percentuale di rilevazione (call rate) e alle informazioni su linkage disequilibrium relative al progetto International HapMap. Sono inoltre stati validati su più di 400 individui provenienti da varie popolazioni (Caucasici, Asiatici, Afro-americani, Africani e individui dell'America del Sud). La distanza fisica media tra gli SNPs è di circa 2.5 kb, la distanza genetica media tra SNPs è 5.8kb mentre l'eterozigosità media è 0,30. Il sistema 500K permette di avere una copertura del genoma tale per cui è presente almeno 1 SNP per 10 kb. Informazioni addizionali per ogni SNP sono presenti sia in GeneChip Genotyping Analysis Software (GTYPE) che nel NetAffx Analysis Center. Queste annotazioni combinano dati derivanti da sorgenti multiple di pubblico dominio e inserite all'interno di un database. Le annotazioni degli SNPs includono: una loro sigla identificativa (dbSNP ID), i geni che li fiancheggiano, la loro locazione di mappa fisica, le citobande e le frequenze alleliche in diverse popolazioni.

Gli arrays a DNA ad alta densità come il Mapping 500K Array Set permettono di ottenere un elevato numero di informazioni in un singolo esperimento (Fodor e coll. 1991, 1993; Pease e coll. 1994) e vengono utilizzati per una varietà di applicazioni inclusi studi di linkage, studi di associazione e studi sulla cancerogenesi (Klein R.J., e coll. 2005).

1.3. Copy Number Variations (CNVs)

Negli ultimi anni il completamento della sequenza del genoma umano (International Human Genome Sequencing Consortium 2004) ha portato notevoli progressi nella comprensione della complessa architettura del genoma umano, rivelando un grado straordinario di diversità fra individui normali (Fredman e coll. 2004; Iafrate e coll. 2004; Sebat e coll. 2004; Sharp e coll. 2005; Tuzun e coll. 2005; Conrad e coll. 2006; Hinds e coll. 2006; Locke e coll. 2006; McCarroll e coll. 2006).

Le varianti strutturali includono una serie di modificazioni molecolari come inversioni, duplicazioni, delezioni (Figura 3) e differenti dalla diversità della sequenza genetica, rappresentata dai polimorfismi nucleotidici (SNPs) (Feuk e coll. 2006a,b; Freeman e coll. 2006).

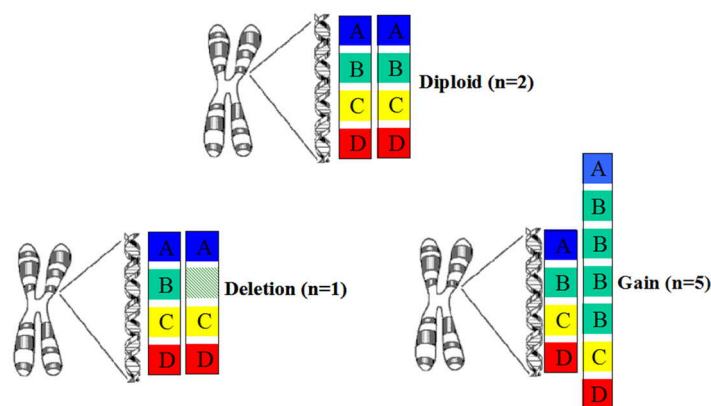


Figura 3

Recenti studi hanno rivelato un ulteriore tipo di variazione nel genoma umano, che include intervalli di DNA genomico relativamente

grandi (approssimativamente 100kb-2.5Mb), chiamati comunemente variazioni di numero di copia (CNVs). Gli ultimi studi effettuati con la tecnologia microarray hanno dimostrato che circa il 12% del genoma umano è variabile in numero di copia (Carter NP, 2007) ma l'estensione di CNVs e la sua frequenza nelle diverse popolazioni è ancora in gran parte sconosciuta (White SJ e coll., 2007).

Una mappa completa che cataloghi ed indicizzi varianti strutturali, in particolare, varianti di numero di copia (CNVs) nel genoma umano è necessaria per capire il ruolo dei CNVs nel contesto del fenotipo normale e del fenotipo affetto (Redon R. e coll., 2006).

A tutt'oggi il ruolo biologico di CNVs non è ancora del tutto conosciuto ma ci sono esempi che mostrano come i CNVs possono influenzare la suscettibilità all'infezione di HIV (Gonzalez e coll. 2005), modulare risposte ai farmaci (Ouahchi e coll. 2006), o contribuire a sindromi dovute a microdelezioni e duplicazioni (Inoue and Lupski 2002). Molte e differenti tecniche molecolari possono essere usate per la rilevazione di CNVs, ma gli esperimenti basati su array a DNA offrono il più efficiente ed economico metodo per scansioni di alta risoluzione per evidenziare queste caratteristiche strutturali del genoma (Sharp e coll. 2005; Speicher and Carter 2005; Hoheisel 2006; Urban e coll. 2006).

Scopo della tesi

Come già detto, uno dei fattori che può ridurre la gravità della beta-talassemia omozigote è l'aumento di emoglobina fetale (HbF), il cui meccanismo molecolare non è stato ancora del tutto chiarito.

Il laboratorio di Ematologia dell'Ospedale Microcitemico da diversi anni conduce uno studio per l'identificazione e caratterizzazione di geni modificatori, che siano responsabili di una sovraespressione di HbF nei pazienti sardi con talassemia intermedia beta⁰³⁹ omozigote non trasfusione dipendente. Le analisi molecolari e statistiche hanno permesso di ottenere risultati preliminari come l'individuazione di diverse zone con elevato LOD score in diversi cromosomi. Considerato l'elevato numero di geni candidati presenti nelle regioni di interesse con elevato LOD score, non potendo incrementare il numero delle famiglie in esame, la ricerca si è spostata sull'utilizzo di un diverso approccio sperimentale basato sulla tecnologia dei microarray, metodo rapido ed efficace, che permette di genotipizzare un numero molto elevato di SNPs distribuiti su tutto il genoma.

La Sardegna, essendo geograficamente isolata, ha una bassa suscettibilità al mescolamento genetico. Pertanto un ipotetico fattore modificatore e i suoi SNPs in linkage segregheranno in un unico aplotipo. Lo scopo della tesi è quindi quello di identificare e

caratterizzare, mediante il GeneChip Human Mapping 500K, l'elemento modificatore responsabile di una sovraespressione di HbF in pazienti sardi con talassemia intermedia beta³⁹ omozigote non trasfusione dipendente (NTD).

Materiali e metodi

In questo studio, condotto in collaborazione con il dott. Paolo Fortina presso la Thomas Jefferson University – Philadelphia, sono stati esaminati 14 pazienti con talassemia intermedia non trasfusione dipendente (NTD), 14 pazienti con talassemia major trasfusione dipendente (TD), entrambi i gruppi con analoghe caratteristiche genotipiche: omozigoti per la mutazione non senso al codon 39 CAG → TAG, aplotipo II/II, XmnI -/-, genotipo alfa-globinico normale (alfa alfa / alfa alfa) , oppure con un gene alfa deletato (- alfa/ alfa alfa), che potrebbe influenzare il fenotipo talassemico, e 10 controlli normali. Il precedente lavoro su questi pazienti ha permesso di escludere:

- mutazioni nel cluster beta-globinico che determinano aumento di HbF (gene gamma triplicato o quadruplicato, delezioni HPFH, delezione 5' beta promoter, delezioni db-talassemia, Hb Lepore),
- che l'aumento di HbF in pazienti con talassemia intermedia sia dovuto alla presenza di particolari polimorfismi nel cluster beta-globinico (HSII-LCR (AT)_xNy(AT)_z; A/G in posizione -1396 G → G/A in posizione -1225 G; IVSII G → (TG)_x(CG)_y; IVSII A (TG)_x(CG)_y (TG)_z; -540 → (AT)_xTy),
- una eredità X-linked,

- loci responsabili di aumento di HbF non associati al cluster beta-globinico descritti in letteratura (Xp22.2-22.3, 6q22.3-23.1, 8q12).

Lo studio invece ha permesso di individuare nei pazienti NTD 8 regioni cromosomiche con un LOD score significativo mediante l'analisi con Genome Wide Search secondo l'approccio della mappatura per omozigosi (1q42.12, 2q14.3, 7q21.3, 9q34.3, 11q23.3, 15q23.2, 16q23.2 e 20q11.21).

Il DNA genomico dei soggetti in esame è stato ottenuto da sangue periferico con estrazione salina ed analizzato mediante il metodo Affymetrix 500K (Fig. 4).

Il DNA di ogni paziente viene sottoposto a due distinte digestioni enzimatiche, una utilizza l'enzima di restrizione NspI e l'altra l'enzima di restrizione StyI. I campioni di DNA così digeriti vengono sottoposti ad ibridazione su due array (NspI e StyI) che insieme sono in grado di genotipizzare 500.000 SNPs distribuiti nell'intero genoma.

La fase iniziale del processo comporta che il DNA genomico totale venga diluito a 50ng/ul in EDTA TE buffer ridotto. Una quantità di DNA pari a 250ng è digerito con un enzima di restrizione, ad esempio NspI, NE Buffer 2 e BSA (100 mg/ml) per 2 ore a 37°C. La successiva reazione di ligation utilizza T4 DNA Ligase (400 U/µL), T4 DNA Ligase Buffer (10X), Adaptor Nsp I (50 µM) per legare un oligonucleotide a doppio filamento (adaptor) a ogni frammento di restrizione prodotto

dalla digestione enzimatica. Il protocollo prevede una reazione di 3 ore a 16°C seguita da una fase a 70°C per 20 minuti. Dopo la reazione di ligation i campioni vengono diluiti con 75ul di acqua AccuGENE.

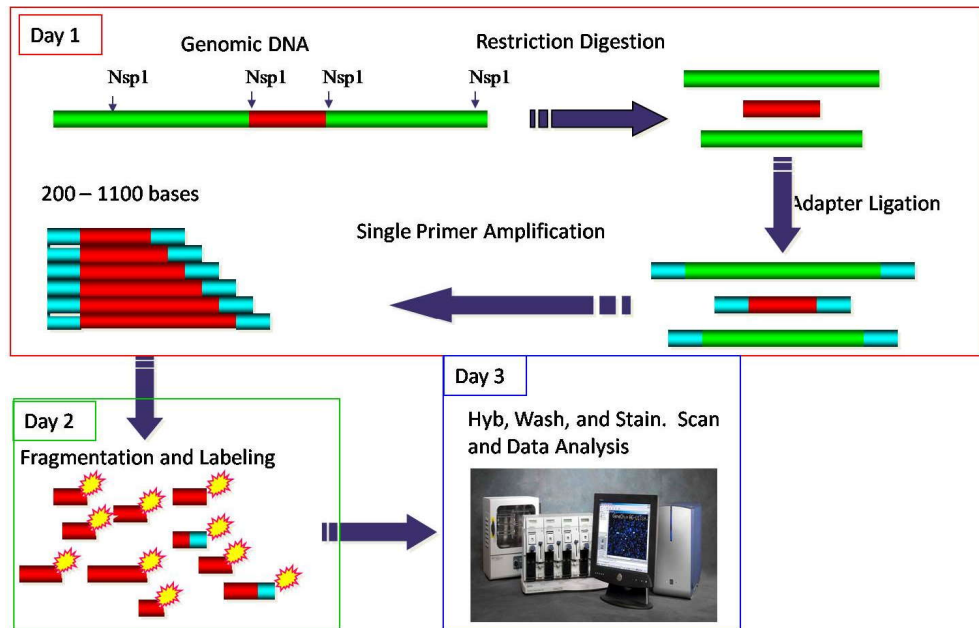


Figura 4. Le fasi del saggio GeneChip Human Mapping 500K

Un generico e unico primer riconosce le sequenze dell'adaptor e amplifica il DNA. Le condizioni di PCR sono ottimizzate per amplificare preferenzialmente frammenti di dimensione compresa tra 200 e 1.000 bp. Il protocollo prevede l'uso di Titanium Taq DNA Polymerase (50X), TITANIUM Taq PCR Buffer (10X), dNTPs (2.5 mM), GC-Melt (5M), PCR Primer 002 (100 μ M). Una parte del prodotto di PCR (3ul) viene caricata su un gel al 2% di agaroso in TBE a 120V per 40 minuti per verificare la presenza dei frammenti desiderati.

Prima di iniziare la fase successiva della frammentazione sono indispensabili 90ug di DNA, pertanto sono necessarie 3 distinte reazioni

di PCR per ogni campione. I prodotti delle 3 reazioni di PCR vengono riuniti e si procede alla purificazione mediante il sistema Clontech DNA Amplification Clean Up Kit che utilizza il vuoto a ~600 mbar. Vengono eseguiti tre lavaggi con 50ul di acqua AccuGENE e il prodotto purificato è risospeso con RB buffer.

La quantificazione della PCR viene fatta mediante Nanodrop per verificare che la quantità del prodotto sia compreso nel range 2.5ug/uL - 3.5 ug/uL.

Il prodotto purificato viene frammentato utilizzando Fragmentation Reagent (DNase I) e Fragmentation Buffer (10X) per 35 minuti a 37°C, 15 minuti di inattivazione a 95°C per ottenere frammenti di dimensioni <180 bp e la loro visualizzazione effettuata su gel al 4% di agaroso in TBE a 120V per 40.

Il prodotto frammentato viene marcato con GeneChip DNA Labeling Reagent (30 mM), Terminal Deoxynucleotidyl Transferase (30 U/μL), Terminal Deoxynucleotidyl Transferase Buffer (5X) a 37°C per 4 ore, 15 minuti di inattivazione a 95°C.

Una quantità del prodotto così ottenuto pari a 70ul viene addizionato alla miscela di ibridazione costituita da MES (12X; 1.25 M), Denhardt's Solution (50X), EDTA (0.5 M), Herring Sperm DNA (10 mg/mL), Oligo Control Reagent, Human Cot-1 DNA (1 mg/mL), Tween-20 (3%), DMSO (100%), Tetramethyl Ammonium Chloride

(TMACL) (5 M) in un volume finale di 190ul e il tutto viene avviato alla denaturazione in un thermal cycler programmato per 95°C poi a 49°C. Il DNA marcato e denaturato è ibridizzato a 49°C a 60 rpm per 16-18 ore su un GeneChip Human Mapping 250K Array. L'ibridazione comporta la formazione di eteroduplex tra il DNA marcato in soluzione e le probe ancorate alla superficie del chip e contenenti gli SNPs. I segnali fluorescenti risultanti dall'ibridazione verranno individuati nella fase di scansione.

Le fasi successive di stabilizzazione e lavaggio sono eseguite nella stazione di lavaggio GeneChip Fluidics Station 450s. La stabilizzazione si effettua utilizzando Array Holding Buffer, R-Phycoerythrin Streptavidin, Anti-streptavidin monoclonal antibody, mentre per il lavaggio si utilizza Wash A (Non-Stringent Wash Buffer: 6X SSPE, 0.01% Tween 20) e Wash B (Stringent Wash Buffer: 0.6X SSPE, 0.01% Tween 20).

I chip sono poi scansionati nel GeneChip Scanner 3000 7G. Dalla scansione sono generati files immagine con estensione .dat (Fig. 4 Punto 1.) che vengono memorizzati in GeneChip Operating Software (GCOS) e trasformati in files .cel (Fig. 4 Punto 2.).

Il GeneChip Genotyping Analysis Software (GTYPE) analizza i dati provenienti da GCOS mediante l'algoritmo Dynamic Model algorithm (DM) (Matsuzaki, e coll. 2004; Di X., e coll. 2005) e fornisce un controllo finale sulla percentuale di rivelazione degli SNPs presenti

nell'array: uno SNPs call uguale o maggiore del 93% con un p-value 0.33 indica che tutte le fasi del processo 500K, dalla digestione del DNA di partenza alla sua marcatura, sono state eseguite dall'operatore in modo ottimale. Verificate queste condizioni si può quindi procedere all'analisi statistica dei dati.

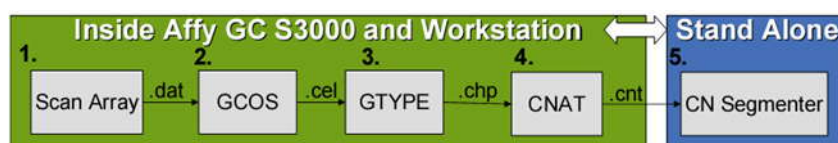


Figura 5. Flusso dei dati generati dalla scansione dei GeneChip attraverso gli applicativi del GeneChip Human Mapping 500K Assay

Recentemente l'algoritmo DM è stato implementato dall'uso del Bayesian Robust Linear Model with Mahalanobis distance classifier (BRLMM), (Nusrat and Terence P., 2005, Affymetrix, Inc., 2006) che compie un'analisi di chip multipla, stimando simultaneamente gli effetti dei segnali delle probe e degli alleli per ogni SNP. Inoltre rivela i genotipi da identificare con una classificazione multipla di genotipi noti.

I dati generati dall'utilizzo degli array ad alta densità come i 500K oltre all'analisi statistica per studi di associazione, possono essere utilizzati per evidenziare eventuali CNVs, segmenti di DNA presenti in numero variabile di copie rispetto a una popolazione di riferimento. Per lo studio dei CNVs, i files con estensione .chp vengono processati con CNAT (Chromosome Copy Number Analysis Tool) (Fig. 4 Punto 3.)

producendo files .cnt (Fig. 4 Punto 4.) che vengono letti dal CN Segmenter (Fig. 4 Punto 5.). Il CN Segmenter visualizza un sommario dei CNVs e rileva la localizzazione fisica lungo ogni cromosoma sulla base della significatività statistica tra i dati provenienti dai campioni in studio e i dati del database di riferimento.

Il software CNAT è provvisto di un'interfaccia grafica che facilita l'identificazione di queste variazioni (Figura 6).

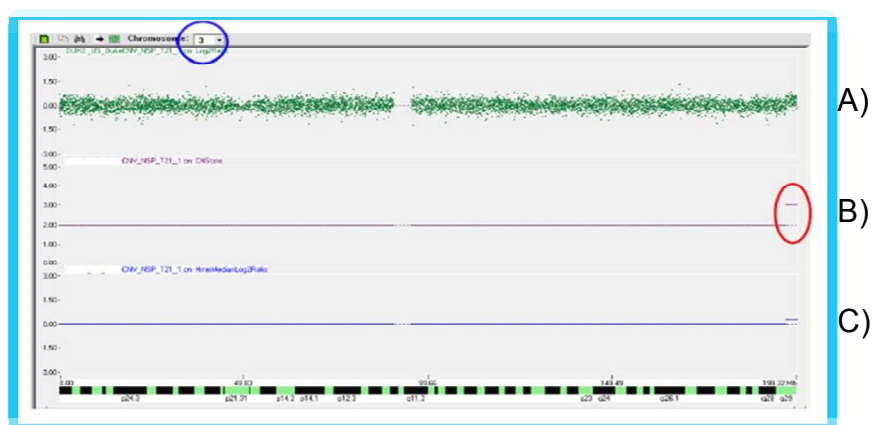


Figura 6. Interfaccia grafica di CNAT. A) Distribuzione degli SNPs, B) Regione duplicata (Gain), C) Assetto diploide 2n

CNAT utilizza ed elabora i dati generati dalla genotipizzazione degli SNPs mediante un algoritmo basato su un modello statistico (Hidden Markov Model) che individua delezioni (Loss) o duplicazioni (Gain) cromosomiche (Figura 6, B) rispetto all'assetto diploide normale (Figura 6, C). Il modello statistico di CNAT genera un report file che mostra una misura di dispersione "l'interquantile range (IQR)" per ogni cromosoma, che deve risultare minore o uguale a 0.35, valore indispensabile per una corretta interpretazione dei dati (Figura 6, A). I

CNVs e i p-values sono calcolati basandosi sull'intensità del segnale di ibridazione degli SNPs dei campioni analizzati in confronto con un set di riferimento contenente 100 individui appartenenti a diverse etnie (Huang e coll. 2004).

Le informazioni provenienti da questa tabella possono poi essere utilizzate per ulteriori ricerche su database presenti in internet attraverso files compatibili con il browser genomico UCSC.

Risultati

Considerata l'impossibilità di incrementare il numero delle famiglie per aumentare l'informatività delle otto regioni di interesse identificate nel lavoro precedente, e che il sequenziamento di tutti i geni candidati presenti in queste regioni richiederebbe un tempo molto lungo, la nostra attenzione si è spostata sull'utilizzo di un diverso approccio sperimentale basato sulla tecnologia dei microarray.

Nella maggior parte dei campioni di DNA analizzati la rivelazione degli SNPs (SNPs call) è stata maggiore del 93% e in alcuni casi superiore al 98%. Ciò indica che tutte le fasi del saggio 500K, dalla digestione del DNA di partenza alla sua marcatura, sono state eseguite in modo ottimale. Si è quindi proceduto all'analisi statistica.

I risultati preliminari dell'analisi statistica hanno permesso di individuare un insieme di SNPs in associazione con alcune zone nella regione di controllo del locus (LCR), HS-111 (-21G>A) e 3' HS1 (+179C>T), che consentono di identificare quattro aplotipi. Il calcolo delle frequenze alleliche in cinque soggetti con talassemia major (TD) e nei controlli normali indicano che la variante allelica 3' HS1 T è significativamente più frequente nei cinque pazienti con talassemia intermedia (NTD) (46%) comparato ai 5 pazienti TD (15%) e nei soggetti normali (16%). La frequenza della variante HS-111 A in cinque

pazienti NTD è risultata del 54.2%, molto simile a quella identificata nei soggetti normali (45.6%). Di conseguenza questa associazione non è statisticamente significativa per i soggetti NTD mentre il dato relativo alla frequenza trovata nei pazienti TD (27.5%) è interessante poiché potrebbe essere utilizzato come marker per la talassemia major.

L'analisi statistica, non ancora del tutto completata e relativa a tutti i campioni in esame, consentirà probabilmente di ottenere un numero di 4-5000 SNPs con $p < 0.01$. Se il p-value risultasse < 0.01 si procederà a un secondo livello di studio basato sull'utilizzo di un numero di SNPs molto ridotto di circa 10-20 SNPs strettamente in linkage con la zona di interesse.

I dati generati dagli arrays ad alta densità, come i GeneChip 500K, possono anche essere elaborati con un diverso approccio statistico che permette di evidenziare la presenza di modifiche molecolari dovute a duplicazioni o delezioni che prendono il nome di "Variazioni del Numero di Copie" (CNVs). La nostra attenzione pertanto si è rivolta allo studio dei comuni segmenti CNVs e di quelli più rari per verificare l'esistenza di una correlazione sulla presenza di tali segmenti con la variazione dell'espressione fenotipica dell' HbF.

L'uso di CNAT ha mostrato, nei pazienti in esame, la presenza di diversi segmenti CNVs, come mostrato nella Tabella 1.

ID	GEN	Cytoband Star Pos	CN_Change Type	CN_State	Size(kb)	#ProbeSet	AvgDistBetween ProbeSets(kB)	%ProbeSets withCNV
I-14	F	4q35.2	Gain	3	166	29	6	76%
I-15	F	4q35.2	Gain	3	771	67	12	1%
I-15	F	4q35.2	Gain	3	809	103	8	57%
I-18	F	4q35.2	Gain	3	129	24	5	63%
I-2	M	4q35.2	Gain	3	323	47	7	47%
M-7	F	4q35.2	Gain	3	140	28	5	79%
M-3	M	10q21.3	Gain	3	208	20	10	40%
M-7	F	10q21.3	Gain	3	567	74	8	30%
M-3	M	13q21.31	Gain	3	1196	135	9	36%
M-7	F	13q21.31	Gain	3	1165	107	11	38%
M-3	M	13q21.32	Gain	3	354	55	6	36%
M-3	M	13q21.33	Gain	3	526	72	7	6%
M-3	M	13q21.33	Gain	3	301	44	7	55%
M-7	F	13q21.33	Gain	3	152	26	6	8%
M-3	M	13q31.1	Gain	3	335	46	7	91%
M-3	M	13q31.1	Gain	3	264	53	5	98%
M-7	F	13q31.1	Gain	3	564	105	5	67%
M-7	F	13q31.1	Gain	3	322	37	9	89%
M-3	M	14q11.2	Loss	1	689	119	6	98%
M-7	F	14q11.2	Loss	1	662	107	6	99%
M-9	F	14q11.2	Loss	1	199	16	12	94%
C-3	F	15q11.2	Gain	3	1662	65	26	100%
M-14	F	16p11.2	Loss	1	345	6	58	17%
I-15	F	17q11.2	Loss	1	1270	40	32	15%
I-15	F	18p11.21	Loss	1	181	8	23	38%
I-16	F	21q21.1	Gain	3	579	69	8	97%
M-16	F	21q21.1	Gain	3	213	27	8	93%
M-3	M	21q21.1	Gain	3	434	60	7	97%
C-4	F	22q11.1	Gain	3	892	6	149	83%
I-13	M	22q11.1	Gain	3	892	6	149	83%
M-3	M	22q11.23	Loss	1	212	16	13	31%
M-7	F	22q11.23	Loss	1	212	16	13	31%
I-1	M	Xp22.33	Loss	1	2716	175	16	15%
I-18	F	Xq22.2	Gain	3	512	11	47	45%
I-18	F	Xq27.1	Gain	3	979	47	21	66%
I-17	F	Xq27.2	Gain	3	275	24	11	46%
M-14	F	Xq27.2	Gain	3	308	27	11	41%

Tabella 1. CNVs in pazienti TD e NTD.

LEGENDA

ID = Pazienti con talassemia intermedia (I) o major (M)

GEN = Sesso

Cytoband Start Pos = Citobanda corrispondente al segmento CNVs deleto o duplicato

CN Change Type = tipo di cambiamento del segmento

CN State = Stato del cambiamento

Size(kb) = Dimensione del segmento deleto o duplicato

#ProbeSet = Numero delle probe che costituiscono il segmento

Avg_Dist Between Probe Sets (kB) = Distanza media tra le probe che costituiscono il segmento

%ProbeSets with CNV = Percentuale delle probe che costituiscono il segmento che si riscontra nel soggetto rispetto alla popolazione di riferimento

Un valore della grandezza %ProbeSet with CNV pari al 100% indica che il corrispondente #ProbeSet è presente nel 100% della popolazione normale di riferimento. Questo significa che una percentuale %ProbeSet minore del 100% è significativa, ma nel nostro studio le percentuali inferiori non sono correlate in modo significativo nelle due popolazioni di soggetti con talassemia major (TD) e con talassemia intermedia (NTD). Dall'analisi dei risultati ottenuti non è quindi emersa una chiara correlazione tra aumento di HbF e presenza di CNVs.

Conclusioni

La Sardegna può essere considerato un isolato genetico (Loudianos G. e coll. 1999, Sahr K.E. e coll. 1993, Pirastu M. e coll. 1984.) con un elevato grado di “omogeneità genetica”. I pazienti sardi da noi esaminati sono tutti omozigoti per la mutazione beta^{o39} e provengono da una zona circoscritta della Sardegna. Questi requisiti sono sufficienti per l’identificazione di un pattern di SNPs strettamente correlato all’aumento di HbF nel gruppo di pazienti NTD. I dati provenienti dai microchip a DNA ad alta densità GeneChip Human Mapping 500K Array, hanno permesso di ottenere informazioni su un numero elevato di SNPs e dai risultati preliminari dell’analisi statistica, non ancora del tutto completata, sono emersi alcuni risultati interessanti.

Nell’immediato futuro ci proponiamo di utilizzare microchip in grado di identificare molecole di microRNA che riteniamo siano coinvolte nella regolazione dell’espressione delle catene gamma globiniche.

Qualora venisse identificato un possibile determinante associato al fenotipo HPFH, sarà possibile definire le basi molecolari che causano la presenza e la sovraespressione di HbF nei pazienti talassemici e

ipotizzare nuove strategie al fine di migliorare il quadro clinico dei pazienti con talassemia major.

Bibliografia:

Affymetrix, Inc., 2006. BRLMM: an improved genotype calling method for the Genechip® Human Mapping 500k Array Set.

Carter NP, 2007. Methods and strategies for analyzing copy number variation using DNA microarrays. *Nat Genet.* 39(7 Suppl):S16-21.

Conrad D.F., Andrews T.D., Carter N.P., Hurles M.E., and Pritchard J.K., 2006. A high-resolution survey of deletion polymorphism in the human genome. *Nat. Genet.* 38: 75–81.

Craig JE., Rochette J., Fischer CA., e coll., 1996. Dissecting the loci controlling fetal haemoglobin production on chromosomes 11p and 6q by the regressive approach. *Nat Genet.* 12(1),58-64.

Di X., Matsuzaki H., Webster T.A., Hubbell E., Liu G., Dong S., Bartell D., Huang J., Chiles R., Yang G., e coll., 2005. Dynamic model based algorithms for screening and genotyping over 100k SNPs on oligonucleotide microarrays. *Bioinformatics* 21(9):1958-63.

Dover G.J., Smith K.D., Chang Y.C., Purvis S., Mays A., Meyers D.A., Sheils C., Serjeant G, 1992. Fetal hemoglobin levels in sickle cell disease and normal individuals are partially controlled by an x-linked gene located at xp22.2. *Blood* 80, 816-824.

Feuk L., Marshall C.R., Wintle R.F., and Scherer S.W., 2006b. Structural variants: changing the landscape of chromosomes and design of disease studies. *Hum. Mol. Genet.* 15: R57–R66.

Fodor S.P., Read J.L., Pirrung M.C., Stryer L., Lu A.T., and Solas D., 1991. Light-directed, spatially addressable parallel chemical synthesis. *Science* 251: 767–773.

Fodor S.P., Rava R.P., Huang X.C., Pease A.C., Holmes C.P., and Adams C.L., 1993. Multiplexed biochemical assays with biological chips. *Nature* 364: 555–556.

Fredman D., White S.J., Potter S., Eichler E.E., Den Dunnen J.T., and Brookes A.J., 2004. Complex snp-related sequence variation in segmental genome duplications. *Nat. Genet.* 36: 861–866.

Freeman J.L., Perry G.H., Feuk L., Redon R., McCarroll S.A., Altshuler D.M., Aburatani H., Jones K.W., Tyler-Smith C., Hurles M.E., e coll., 2006. Copy number variation: new insights in genome diversity. *Genome Res.* 16: 949–961.

Galanello R., Cao A., 1998. Relationship between genotype and phenotype. thalassemia intermedia. *Ann N Y Acad Sci* 850, 325-333.

Galanello R. Dessi E., Melis M.A., Addis M., Sanna M.A. Rosatelli C., Argiolu F., Giagu L., Turco M.P. Cacace E., Pirastu M., Cao A., 1989. Molecular analysis of β^0 -thalassemia intermedia in sardinia. *Blood* 74, 823-827.

Gonzalez E., Kulkarni H., Bolivar H., Mangano A., Sanchez R., Catano G., Nibbs R.J., Freedman B.I., Quinones M.P., Bamshad M.J., e coll., 2005. The influence of ccl3l1 gene-containing segmental duplications on hiv-1/aids susceptibility. *Science* 307:1434–1440.

Gschwend M., Levrin O., Kruglyak L., Ranade K., Verlander P.C. Shen S., Faure S., Weissenbach J., Altay C., Lander E.S., Auerbach A.D., Botstein D., 1996. A

locus for fanconi anemia on 16q determined by homozygosity mapping. *Am J Hum Gen*, 59(2):377-84.

Hinds D.A., Kloek A.P., Jen M., Chen X., and Frazer K.A., 2006. Common deletions and snps are in linkage disequilibrium in the human genome. *Nat. Genet.* 38:82–85.

Hoheisel J.D., 2006. Microarray technology: beyond transcript profiling and genotype analysis. *Nat. Rev. Genet.* 7: 200–210.

Huang, J., e coll., 2004. Whole genome dna copy number changes identified by high density oligonucleotide arrays. *Human Genomics* 1: 287-299.

Iafate A.J., Feuk L., Rivera M.N., Listewnik M.L., Donahoe P.K., Qi, Y., Scherer S.W., and Lee C., 2004. Detection of large-scale variation in the human genome. *Nat. Genet.* 36:949–951.

Inoue K. and Lupski J.R., 2002. Molecular mechanisms for genomic disorders. *Annu. Rev. Genomics Hum. Genet.* 3: 199–242.

International HapMap Consortium, 2005. A haplotype map of the human genome. *Nature* 437:1299–1320.

International Human Genome Sequencing Consortium, 2004. Finishing the euchromatic sequence of the human genome. *Nature* 431: 931–945.

Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J., 2005. Complement factor h polymorphism in age-related macular degeneration. *Science*, 308(5720):385-9

Kruglyak L., 1999. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nature Genetics*, 22 :139-44.

Kruglyak L., Daly M.J., Lander E.S., 1995. Rapid multipoint linkage analysis of recessive traits in nuclear families including homozygosity mapping. *Am J Hum Genet* 56, 519-527.

Jorde LB, Rogers AR, Bamshad M, Watkins WS, Krakowiak P, Sung S, Kere J, Harpending HC., 1997. Microsatellite diversity and the demographic history of modern humans. *Proc Natl Acad Sci U S A*, 94(7):3100-3

Lapoumeroulie C., Castiglia L., Ruberto C., e coll., 1999. Genetic variations in human fetal globin gene microsatellites and their functional relevance. *Hum Genet.* 104(4),307-314.

Locke D.P., Sharp A.J., McCarroll S.A., McGrath S.D., Newman T.L., Cheng Z., Schwartz S., Albertson D.G., Pinkel D., Altshuler D.M., e coll., 2006. Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome. *Am. J. Hum. Genet.* 79: 275–290.

Loudianos G., Dessi V., Lovicu M., Angius A., Figus A., Lilliu F., De Virgiliis S., Nurchi A.M., Deplano A., Moi P., Pirastu M., Cao A, 1999. Molecular characterization of wilson disease in the sardinian population: evidence of a founder effect. *Hum. Mutat* 14, 294-303.

Matsuzaki, e coll., 2004. Genotyping over 100,000 snps on a pair of oligonucleotide arrays. *Nature Methods* 1:109-111.

McCarroll S.A., Hadnott T.N., Perry G.H., Sabeti P.C., Zody M.C., Barrett J.C., Dallaire S., Gabriel S.B., Lee C., Daly M.J., et al., 2006. Common deletion polymorphisms in the human genome. *Nat. Genet.* 38:86–92.

Nusrat Rabbee and Terence P. Speed., 2005. A genotype calling algorithm for affymetrix snp array. *Bioinformatics Advance Access* published online on Nov 2.

Ouahchi K., Lindeman N., and Lee C., 2006. Copy number variants and pharmacogenomics. *Pharmacogenomics* 7: 25–29.

Pease A.C., Solas D., Sullivan E.J., Cronin M.T., Holmes C.P., and Fodor S.P., 1994. Light-generated oligonucleotide arrays for rapid dna sequence analysis. *Proc. Natl. Acad. Sci.* 91: 5022–5026.

Pirastu M., Kan Y.W., Galanello R., Cao A., 1984. Multiple mutations produce delta beta 0 thalassemia in sardinia. *Science* 223, 929-930.

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, González JR, Gratacòs M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zerjal T, Zhang J, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurles ME, 2006. Global variation in copy number in the human genome. *Nature.* 23;444(7118):428-9.

Sahr K.E., Coetzer T.L., Moy L.S., Derick L.H., Chishti A.H., Jarolim P., Lorenzo F., Del Giudice Miraglia E., Iolascon A., Galanello R., Cao A., Delaunay J., Liu S.C., Palek J., 1993. An ala>gly substitution in helix 1 of beta spectrin repeat 17

that severely disrupts the structure and self-association of the erythrocyte spectrin heterodimer. *The Journal of Biological Chemistry* 268, 22656-22662.

Sebat J., Lakshmi B., Troge J., Alexander J., Young J., Lundin P., Maner S., Massa H., Walker M., Chi M., e coll., 2004. Large-scale copy number polymorphism in the human genome. *Science* 305: 525-528.

Sharp A.J., Locke D.P., McGrath S.D., Cheng Z., Bailey J.A., Vallente R.U., Pertz L.M., Clark R.A., Schwartz S., Se Graves R., e coll., 2005. Segmental duplications and copy-number variation in the human genome. *Am. J. Hum. Genet.* 77: 78-88.

Speicher M.R. and Carter N.P., 2005. The new cytogenetics: blurring the boundaries with molecular biology. *Nat. Rev. Genet.* 6: 782-792.

Talbot D., Philipsen S., Fraser P., Grosveld F., 1990. Detailed analysis of the site 3 region of the human β -globin dominant control region. *EMBO* 9, 2169-2177.

Tuzun E., Sharp A.J., Bailey J.A., Kaul R., Morrison V.A., Pertz L.M., Haugen E., Hayden H., Albertson D., Pinkel D., e coll., 2005. Fine-scale structural variation of the human genome. *Nat. Genet.* 37: 727-732.

Urban A.E., Korbel J.O., Selzer R., Richmond T., Hacker A., Popescu G.V., Cubells J.F., Green R., Emanuel B.S., Gerstein M.B., e coll., 2006. High-resolution mapping of dna copy alterations in human chromosome 22 using high-density tiling oligonucleotide arrays. *Proc. Natl. Acad. Sci.* 103: 4534-4539.

White SJ, Vissers LE, Geurts van Kessel A, de Menezes RX, Kalay E, Lehesjoki AE, Giordano PC, van de Vosse E, Breuning MH, Brunner HG, den Dunnen JT, Veltman JA., 2007. Variation of CNV distribution in five different ethnic populations. *Cytogenet Genome Res.*;118(1):19-30.

Zavattari P., Deidda E., Whalen M., Lampis R., Mulargia A., Loddo M., Eaves I., Mastio G., Todd J.A., Cucca F., 2000. Major factors influencing linkage disequilibrium by analysis of different chromosome regions in distinct populations: demography, chromosome recombination frequency and selection. *Hum Mol Genet* 9, 2947-2957.