

**UNIVERSITÀ DEGLI STUDI DI PADOVA**

**FACOLTA' DI INGEGNERIA**

**LAUREA MAGISTRALE IN BIOINGEGNERIA**

**ANALISI DELLA VARIANZA DI DATI DI  
ESPRESSIONE GENICA**

**RELATORE: PROF. SSA BARBARA DI CAMILLO**

**LAUREANDO: GIULIO RIGON**

**ANNO ACCADEMICO 2012-2013**



# Indice

<b>1. Introduzione.....</b>	<b>1</b>
1.1 L'analisi del trascrittoma e la tecnologia dei microarray.....	1
1.2 Microarray di oligonucleotidi Affymetrix.....	2
1.2.1 Tecnologia e sintesi dei microarray di oligonucleotidi Affymetrix.....	2
1.2.2 Principio di funzionamento dei microarray di oligonucleotidi Affymetrix...	6
<b>2. Scopo della tesi.....</b>	<b>9</b>
<b>3. Dati della tesi.....</b>	<b>11</b>
3.1 Descrizioni dei dati.....	11
3.2 Il GeneChip® Human Genome U133 Plus 2.0 Array.....	11
<b>4. Gene Ontology.....</b>	<b>13</b>
4.1 Il NetAffx Analysis Center.....	13
<b>5. La patologia.....</b>	<b>15</b>
5.1 Il diabete mellito di tipo 2.....	15
5.1.1 Sintomi.....	15
5.1.2 Cause.....	15
5.1.3 Fisiopatologia.....	15
5.2 L'intolleranza al glucosio.....	16
<b>6. Pre-processing dei dati Affymetrix.....</b>	<b>17</b>
6.1 Analisi dell'immagine.....	17
6.2 Controllo di qualità.....	17
6.3 L'algoritmo RMA.....	18
6.3.1 Correzione del segnale di ibridazione non specifica.....	18
6.3.2 Normalizzazione.....	20
6.3.3 Calcolo del segnale.....	20
<b>7. I metodi di analisi.....</b>	<b>23</b>
7.1 SAM (Significance Analysis of microarrays).....	23
7.1.1 Descrizione.....	23
7.1.2 Implementazione.....	26
7.2 GSEA (Gene Set Enrichment Analysis).....	26
7.2.1 Descrizione.....	27
7.2.2 Implementazione.....	29
7.3 Analisi della varianza.....	29
<b>8. Risultati.....</b>	<b>33</b>
8.1 Risultati SAM.....	33
8.2 Risultati GSEA.....	35
8.2.1 Analisi dei dati di espressione genica.....	35
8.2.1.1 Normal VS Glucose Intollerant.....	35
8.2.1.2 Normal VS Diabetic.....	36
8.2.1.3 Glucose Intollerant VS Diabetic.....	38

8.2.2 Analisi dei dati della varianza dell'espressione genica.....	38
8.2.2.1 Normal VS Glucose Intollerant.....	38
8.2.2.2 Normal VS Diabetic.....	46
8.2.2.3 Glucose Intollerant VS Diabetic.....	54
8.3 Risultati di Analisi della varianza.....	67
<b>9. Conclusioni e sviluppi futuri.....</b>	<b>97</b>
<b>Appendice: Guida ai risultati GSEA.....</b>	<b>99</b>
<b>Bibliografia.....</b>	<b>105</b>
<b>Web Resource.....</b>	<b>106</b>

# 1. INTRODUZIONE

## 1.1 L'analisi del trascrittoma e la tecnologia dei microarray

Il trascrittoma è l'insieme delle molecole di RNA (o trascritti) presenti in una cellula e la sua espressione varia in dipendenza delle condizioni ambientali della cellula. L'analisi del trascrittoma è utile a comprendere quali geni siano attivati nelle diverse fasi del ciclo cellulare e nello sviluppo di condizioni patologiche. Dall'analisi differenziale dell'espressione genica in soggetti sani e soggetti malati si possono infatti individuare i geni che risultano essere espressi in maniera significativamente diversa fra i due gruppi e quindi evidenziare le differenze che la condizione patologica comporta.

L'analisi del trascrittoma può essere realizzata sfruttando la tecnica dell'ibridazione. Essa si basa sulla proprietà dei nucleotidi di appaiarsi con i loro complementari. Di questa categoria fanno parte i microarray, da anni largamente utilizzati per ottenere informazioni sull'espressione genica. I microarray sono costituiti da un supporto solido a cui sono ancorate delle sonde di DNA, dette *probe*, in numero molto elevato per ogni gene e disposte in posizioni note.

L'RNA estratto dalla cellula viene retrotrascritto, marcato con fluorescente e ibridato con il microarray. L'intensità della fluorescenza è una misura di quante molecole hanno ibridato la *probe*, ovvero di quanto il gene associato alla *probe* è espresso nella cellula.

La diffusione negli ultimi 20 anni dei microarray e delle tecnologie *high-throughput* in generale, ha profondamente modificato l'approccio allo studio delle malattie genetiche. La misura del livello di espressione dei singoli geni, mediante quantificazione delle diverse molecole di mRNA cellulare, consente di indagare i complessi meccanismi di controllo dell'espressione genica e di analizzare le possibili modificazioni dell'attività regolatoria che possono dare luogo all'alterazione dei processi biologici e delle funzioni della cellula.

La vera rivoluzione introdotta dai microarray è data dalla possibilità di monitorare contemporaneamente il livello di espressione di decine di migliaia di geni presenti nell'uomo su una stessa piattaforma, resa possibile dal crescente grado di miniaturizzazione consentito dalle nuove tecnologie. Il monitoraggio dell'intero trascrittoma, senza dover restringere l'analisi a singoli geni o pathway isolati, e la possibilità di considerare il sistema di regolazione genica nella sua globalità consente di studiare malattie complesse caratterizzate dalla presenza di numerose mutazioni e alterazioni funzionali a carico di diversi geni.

Questa tecnica presenta tuttavia anche alcuni limiti, come la necessità di conoscere a priori le sequenze geniche per la progettazione delle *probe* e il limitato range dinamico (cioè il rapporto fra i livelli di massima e minima espressione genica misurabili), dovuto al rumore di fondo e al fenomeno di saturazione del segnale.

Inoltre i metodi che si basano sul confronto dei dati di espressione provenienti da diversi esperimenti con microarray sullo stesso caso biologico producono spesso risultati differenti. Tali difficoltà nella selezione dei geni e la scarsa riproducibilità dei risultati sono riconducibili a caratteristiche intrinseche dei dati di espressione genica e della tecnologia dei microarray :

- il numero di trascritti monitorati è dell'ordine delle decine di migliaia: se da un lato la possibilità di monitorare l'intero trascrittoma consente di considerare il sistema nella sua globalità, dall'altro complica la ricerca delle variabili che caratterizzano la patologia analizzata;

- in genere è possibile monitorare l'intero trascrittoma a fronte di qualche decina o al massimo centinaia di array, sia per motivi tecnici ed etici legati alla raccolta dei campioni che per ragioni di costo: il numero di campioni disponibili per una variabile genica è di gran lunga inferiore al numero di variabili monitorate e il problema risulta essere mal condizionato;
- la variabilità tecnica e biologica che caratterizza gli esperimenti con microarray (dalla strumentazione utilizzata ai protocolli sperimentali) introduce una componente di rumore ai dati di espressione genica che ne complica l'analisi e l'elaborazione e una scarsa riproducibilità delle procedure;

L'analisi dell'espressione genica è diventata uno strumento onnipresente per lo studio di molte malattie dell'uomo. Tipicamente in un'analisi vengono confrontati diversi gruppi fenotipici per identificare i geni che sono differenzialmente espressi sulla base della media dell'espressione genica. Nel presente lavoro viene descritto un approccio innovativo all'analisi dei dati di espressione genica, utilizzando le differenze nella varianza dell'espressione tra i gruppi come parametro informativo sul fenotipo.

A tal proposito studi recenti (*Mar JC, et al., 2011 Variance of Gene Expression Identifies Altered Network Constraints in Neurological Disease*) condotti su malattie quali la schizofrenia e il morbo di Parkinson, hanno evidenziato che geni con differenti profili della varianza dell'espressione genica non sono distribuiti casualmente nella rete di comunicazione della cellula; tale studio mostra come geni con una bassa varianza di espressione sono maggiormente connessi tra loro in relazione ai pathway metabolici della cellula rispetto ai geni che presentano una elevata varianza di espressione. Prima ancora che nella genetica umana, questo tipo di analisi è stata condotta nello studio di *C. elegans* da *Raj et al.*; è stato osservato che la variazione del numero di trascritti espressi in ogni cellula, influenza direttamente lo sviluppo dell'intestino dei vermi. La mutazione dei fattori di trascrizione influenzava non solo l'espressione media dei geni target, ma anche la varianza del loro livello di espressione.

## 1.2 Microarray di oligonucleotidi Affymetrix

### 1.2.1 Tecnologia e sintesi dei microarray di oligonucleotidi Affymetrix

Un microarray di DNA è costituito da un insieme di microscopiche sonde di DNA attaccate ad una superficie solida come vetro, plastica, o chip di silicio formanti un array (o matrice). Tali array permettono di esaminare simultaneamente la presenza di moltissimi geni all'interno di un campione di DNA (che spesso può rappresentare anche tutto il genoma o il trascrittoma di un organismo).

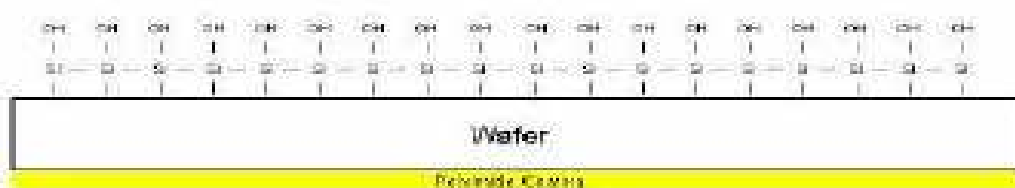
I microarray sfruttano una tecnica di ibridazione inversa, che consiste nel fissare tutti i segmenti di DNA (detti *probe*) su un supporto e nel marcare invece l'acido nucleico che vogliamo identificare (detto *target*).

Per studiare gli mRNA, essi vengono prima estratti dalle cellule, e poi solitamente convertiti in cDNA con l'uso di un enzima chiamato trascrittasi inversa e allo stesso momento marcati con una sonda fluorescente.

Con la tecnologia Affymetrix invece di cDNA vengono utilizzate catene di nucleotidi (oligonucleotidi) sintetizzate in laboratorio lunghe circa 70 basi e già a singola catena. Gli oligonucleotidi vengono sintetizzati direttamente su un supporto in silicio (chip) usando tecniche fotolitografiche. Questa tecnologia permette di sintetizzare per ogni sequenza interrogata (per ogni spot) un numero oligonucleotidi dell'ordine di 109.

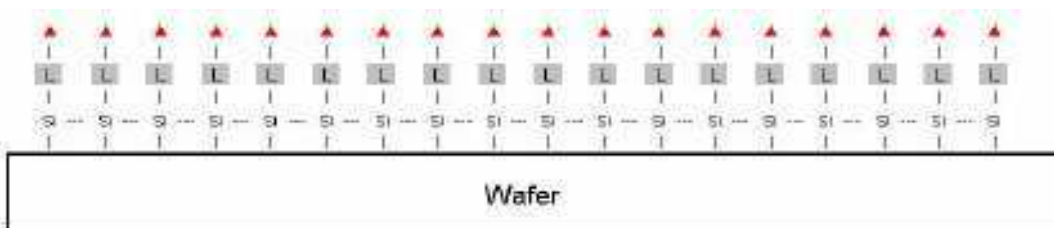
Nei seguenti passi si descrive la tecnica di sintesi degli oligonucleotidi *in situ*:

1. Il wafer di silicio viene rivestito (coating) di molecole di Silano cui sono legati dei gruppi ossidrilici (OH)



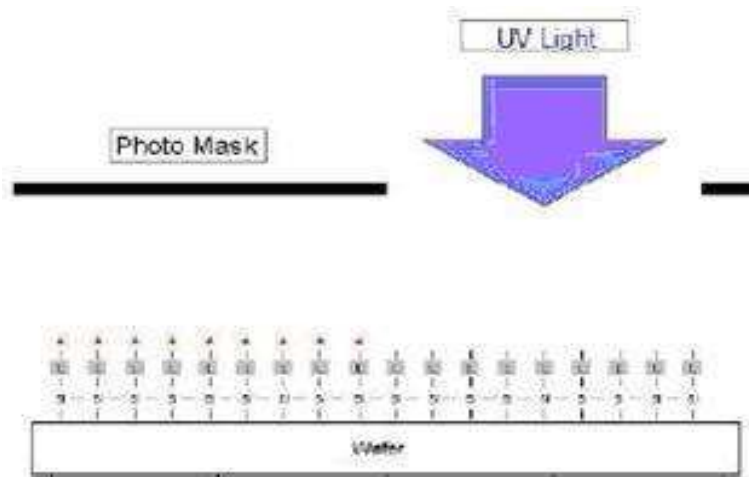
**Figura 1.1: Wafer di silicio rivestito di molecole di Silano e gruppi ossidrilici.**

2. Viene aggiunto un linker che lega le molecole di silano e contiene un gruppo ossidrilico in grado di legare un gruppo fosfato di un nucleotide. Il gruppo ossidrilico è però protetto da una molecola fotosensibile (i triangolini rossi in figura), che può essere eliminata esponendo il wafer a luce ultravioletta



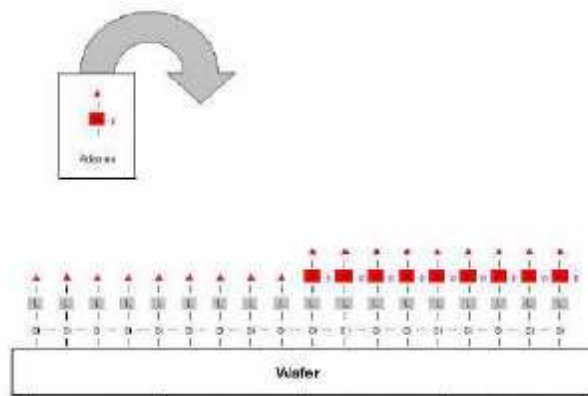
**Figura 1.2 Wafer di silicio con aggiunta dei linker.**

3. Sopra il wafer viene posizionata una maschera in grado di proteggere il wafer dalla luce UV ad eccezione di aree molto piccole ( $5\mu\text{m}\times 5\mu\text{m}$ ) in cui la maschera è forata
4. Il wafer viene esposto a luce UV e i linker perdono le protezioni in corrispondenza ai punti in cui la maschera è forata



**Figura 1.3: Wafer esposto a luce UV.**

- Supponiamo di voler inserire un nucleotide contenente la base adenina. Esso viene messo in contatto con il wafer e aderisce ai gruppi OH liberi dei linker (che funzionano come un'estremità 3'). Il nucleotide adenina appena aggiunto è legato ad una molecola protettiva fotosensibile (i triangolini rossi in figura) in posizione 3', pertanto non è in grado di legare un gruppo fosfato di un altro nucleotide.



**Figura 1.4: Aggiunta di nucleotidi al wafer.**

- A questo punto vengono ripetuti i passi 3 e 4 (sovrapposizione di una maschera opportuna ed esposizione a luce UV per rimuovere la protezione in corrispondenza ai fori nella maschera)



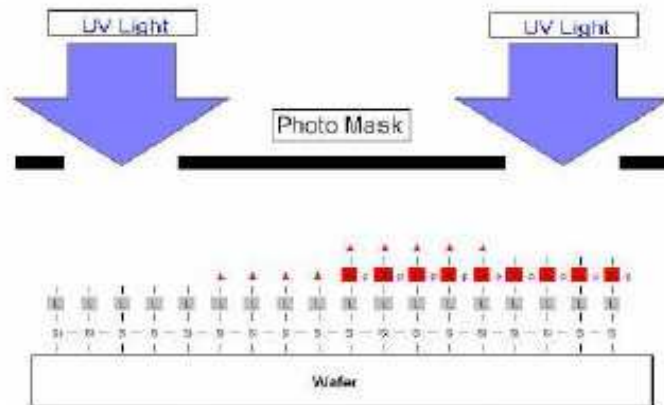


Figura 1.5: Esposizione del wafer a luce UV con ausilio di una maschera.

e il passo 5 (adesione di un nucleotide a cui è legata la molecola di protezione). E così via.

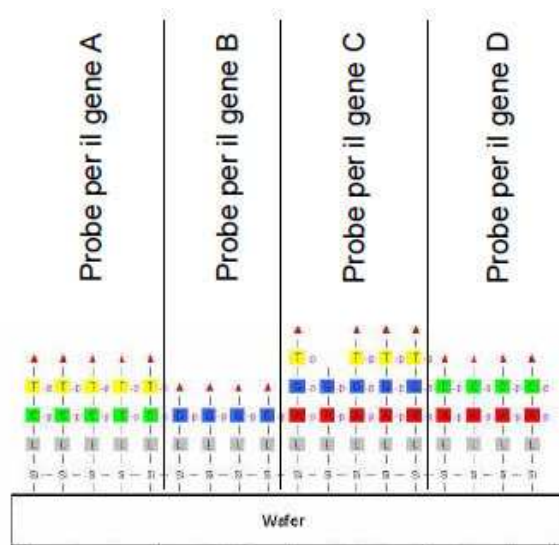


Figura 1.6: Costruzione delle *probe* di oligonucleotidi per più geni sullo stesso wafer.

Progettando opportunamente maschere e strategia di aggiunta delle basi è possibile sintetizzare contemporaneamente tutte le probe specifiche per interrogare migliaia di geni (è possibile arrivare a 1.6 milioni di spot; wafer :  $1.28\text{cm} \times 1.28\text{cm} = \sim 1.6\text{ cm}^2$ )

La densità delle probe in ogni spot permette di evitare, in linea di massima, fenomeni di saturazione. Il numero elevato di spot consente di utilizzare più probe differenti per ogni singolo trascritto rendendo l'insieme delle probe usate per ciascun trascritto molto sensibile.

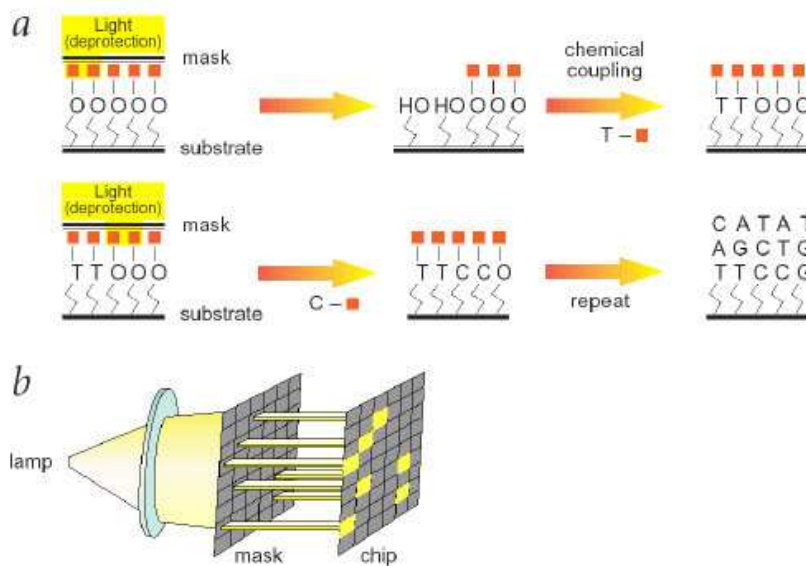


Figura 1.7a: Riassunto del processo di costruzione delle *probe* di oligonucleotidi; Figura 1.7b: Esposizione a luce UV con ausilio della maschera.

## 1.2.2 Principio di funzionamento dei microarray di oligonucleotidi Affymetrix

Per ogni gene da monitorare vengono progettate dalle 11 alle 20 sequenze *probe* (l'insieme delle quali è detto *probe-set*). Ogni *probe* è detta *perfect-match* (PM). Le *probe* sono progettate posizionandosi vicino all'estremità 3' del gene da monitorare. Inoltre, per monitorare fenomeni di cross-ibridazione, per ogni *perfect-match* viene progettata una sequenza *mismatch* (MM) che coincide con il PM tranne che per il nucleotide centrale, che non appaia in maniera specifica con la sequenza del gene. Il nucleotide centrale del MM rende l'appaiamento del MM con la sequenze dell'RNA monitorato meno probabile. La quantità di *probe* MM che legano con le sequenze dell'RNA monitorato (o sequenze simili) è rappresentativa della percentuale di *probe* PM che cross-ibridano con RNA non specifici.

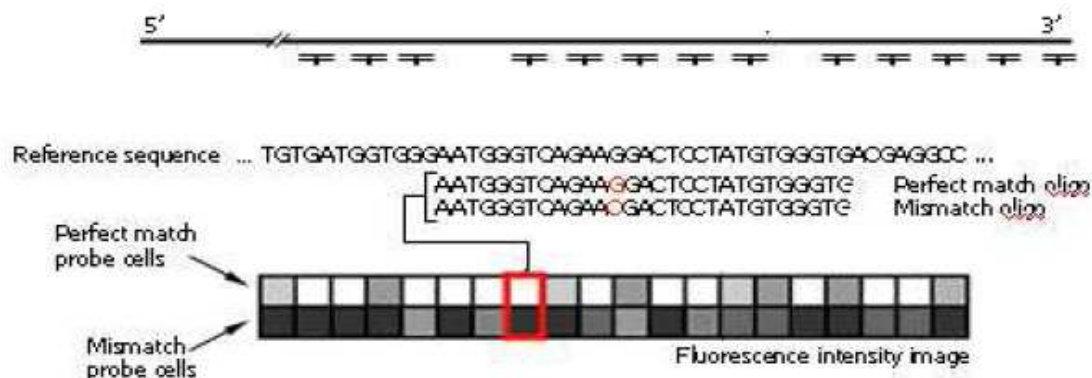
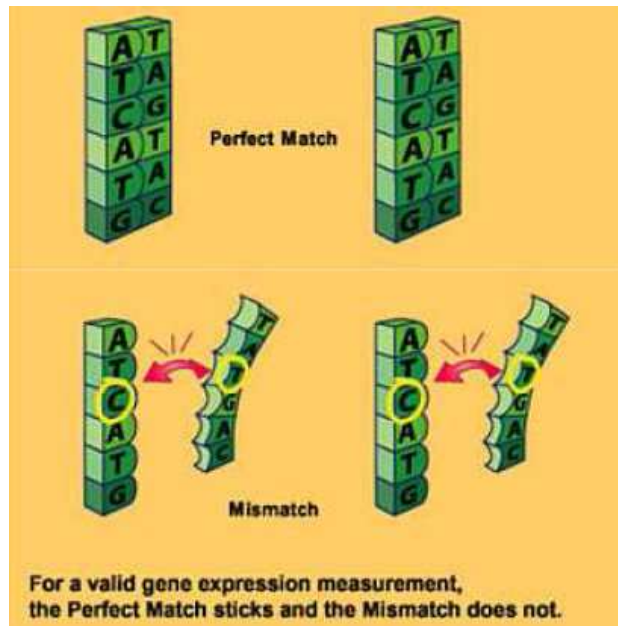
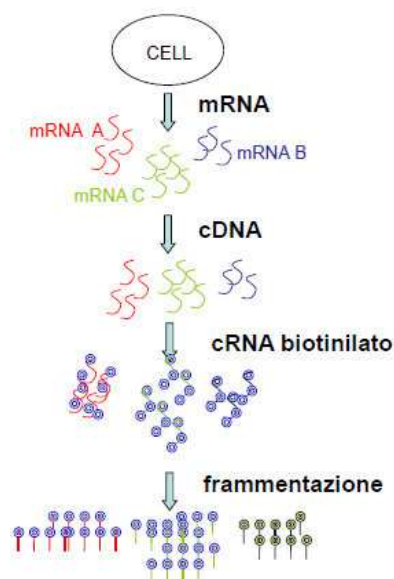


Figura 1.8: Schema costruttivo del microarray di oligonucleotidi Affymetrix.



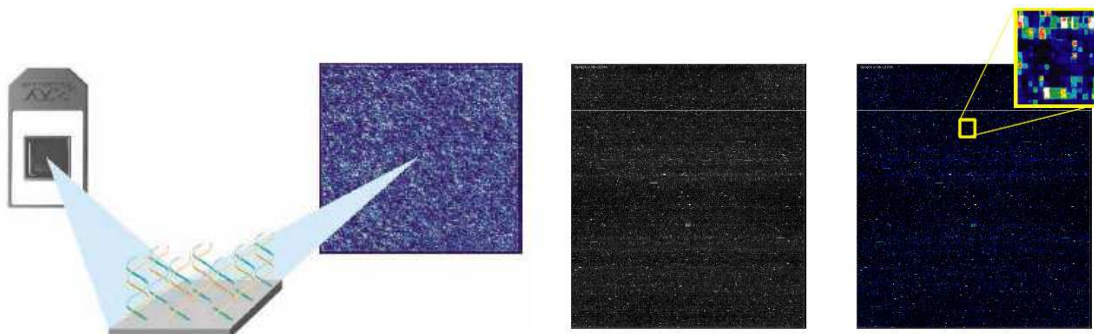
**Figura 1.9:** Rappresentazione a livello molecolare della differenza tra Perfect Match e Mismatch.

Dalle cellule di interesse si estrae l'RNA *target* (genoma effettivamente espresso) da questo si sintetizza il DNA complementare (cDNA) mediante l'uso dell'enzima trascrittasi inversa. Il cDNA è trascritto in RNA complementare (cRNA) contenente molecole di biotina. Il cRNA è frammentato in segmenti di 35-200 basi (devono legare probe di 11-25 basi). Il cRNA frammentato viene messo a contatto con l'array per qualche ora per consentire l'ibridazione specifica e poi viene lavato per eliminare le molecole non ibridate. Solo a questo punto un marcatore fluorescente (che lega la biotina) viene incorporato nei frammenti target. Il fluorescente viene incorporato in un secondo momento perchè non interferisca con il processo di ibridazione.



**Figura 1.10:** Schema riassuntivo del processo di estrazione del RNA e preparazione del cRNA.

Eccitando il fluorescente con una luce laser, esso emetterà onde elettromagnetiche ad una lunghezza d'onda caratteristica. L'intensità dell'emissione per ciascuno spot viene rilevata da uno scanner ad alta sensibilità. Un software analizza l'immagine, sottrae il background, calcola le intensità di ogni spot e infine crea una pseudoimmagine del microarray.



**Figura 1.11: Immagine del microarray elaborata tramite software, da cui si ricava il segnale di intensità di espressione genica.**

## 2. SCOPO DELLA TESI

Nello studio della genomica si è soliti considerare i gruppi fenotipici come gruppi definiti sulla base delle differenze tra le medie di espressione. Questo tipo di analisi si è rivelata molto utile nella comprensione dei sistemi biologici, in quanto molto spesso il loro comportamento è descrivibile dall'interpretazione di queste differenze.

La differenza tra le medie di espressione può tuttavia riflettere la differenza, più in generale, tra le distribuzioni dei valori di espressione nei gruppi fenotipici; le distribuzioni sono definite da due parametri: il valore medio appunto, e la varianza.

In questo contesto la varianza è solitamente utilizzata come misura della significatività di tali differenze medie; dalla sua analisi, tuttavia, si possono avere implicazioni sia sperimentali che biologiche che possono fornire un importante aiuto nel valutare le differenze fenotipiche che possono manifestarsi in una malattia come il diabete mellito, per la quale la distinzione tra i gruppi fenotipici è difficilmente deducibile dalle differenze tra le medie di espressione.

Il presente lavoro di tesi si compone di due parti:

- la prima parte prevede un'analisi standard dei dati di espressione per individuare i geni differenzialmente espressi tramite due diverse procedure (SAM e GSEA)
- nella seconda parte viene analizzata la varianza dei geni nei vari gruppi fenotipici per valutarne l'espressione differenziale, tramite una procedura di selezione dipendente dall'intensità di espressione e GSEA.



## 3. DATI DELLA TESI

### 3.1 Descrizione dei dati

I dati sono forniti da *Gene Expression Omnibus* (GEO), una banca dati che archivia e distribuisce liberamente dati relativi a microarray, *next-generation sequencing* e altre forme di tecnologie *high-throughput* presentati dalla comunità scientifica.

La tecnologia dei microarray è stata utilizzata per definire il profilo di espressione dell'mRNA estratto dalle cellule del muscolo scheletrico di tre classi di soggetti:

- normali
- intolleranti al glucosio
- diabetici (diabete mellito di tipo 2).

La classificazione dei soggetti segue i criteri stabiliti dalla World Health Organization (WHO) e i pazienti diabetici non sono stati sottoposti a terapia anti-ipoglicemica per una settimana prima della biopsia.

L'RNA è stato estratto da cellule prelevate dal vasto laterale, un muscolo che fa parte dei muscoli anteriori della coscia ed uno dei quattro muscoli del quadricipite femorale. L'RNA è stato ibridato all'array HGU133 Plus 2.0 utilizzando le indicazioni del costruttore.

### 3.2 Il GeneChip® Human Genome U133 Plus 2.0 Array

Affymetrix GeneChip® Human Genome U133 Plus 2.0 Array (HG-U133 Plus 2.0) è un microarray costituito da 1 300 000 sequenze di oligonucleotidi che coprono oltre 47000 trascritti che rappresentano approssimativamente 39000 geni umani tra i più frequentemente rilevati. Tutti i probe set sono identicamente numericamente rappresentati sul chip. La maggior parte dei probe set utilizzati nell'array sono selezionati da GenBank®, dbEST, and RefSeq.

Le sonde oligonucleotidiche sono sintetizzate *in situ* complementariamente alle rispettive sequenze di RNA. Undici coppie di sonde (*probe pair*) sono utilizzate per misurare il livello di espressione del trascritto per ogni sequenza oligonucleotidica rappresentata.



**Figura 3.1: Il Gene Chip HGU133 Plus 2.0**

Il *dataset* consiste in una matrice contenente i valori di espressione genica così formata:

- 54675 righe, una per ogni *probe* indagata
- 118 colonne, una per ogni microarray (47 normali, 26 intolleranti al glucosio, 45 diabetici)



## 4. GENE ONTOLOGY

*Gene Ontology* (GO) è un'ontologia per la descrizione dell'annotazione funzionale dei geni e dei loro prodotti genici in generale, a cui si è fatto riferimento per il presente lavoro di tesi.

Il progetto GO ha definito tre diversi tipi di vocabolari di termini biologici unificati:

**PROCESSI BIOLOGICI (BP):** classifica un meccanismo controllato da più prodotti genici che viene portato a termine e completato attraverso un insieme ordinato di funzioni molecolari, spesso coinvolgendo una trasformazione chimica o fisica;

**FUNZIONI MOLECOLARI (MF):** si riferisce all'attività biochimica di prodotto genico (o complesso proteico) di esprimersi in un determinato contesto metabolico

**COMPONENTI CELLULARI (CC):** identifica una zona specifica della cellula dove il prodotto genico è attivo

Ognuno dei tre vocabolari controllati della GO è descritto da una struttura chiamata *Grafo Diretto Aciclico (DAG)* che collega i suddetti termini secondo una relazione padre-figlio: i termini possono avere più di un termine-padre e zero, uno o più termini-figli. Un prodotto genico (rappresentato da un codice identificativo) viene associato al codice del termine GO che descrive la proprietà di quel prodotto genico. Tuttavia spesso un prodotto genico può possedere più proprietà, perciò può essere associato a più termini GO diversi; viceversa, ogni termine GO viene associato a uno o più prodotti genici; a causa della struttura associata all'ontologia, quando un prodotto genico viene annotato ad un termine GO, in realtà è annotato anche a tutti i relativi termini-padre. I termini GO sono collegati tra loro principalmente da due tipi di relazioni: *is\_a* e *part\_of*; i nodi più vicini al nodo radice del grafo sono legati a concetti più generali, mentre i più distanti rappresentano concetti più specifici.

### 4.1 Il NetAffx Analysis Center

Affymetrix fornisce per tutti i microarray che produce le rispettive annotazioni. Per ogni *probeset* dato in ingresso (caricando il file contenente l'identificativo delle probe nella sezione *Batch Query* del *NetAffx Analysis Center* del sito <http://www.affymetrix.com>), *NetAffx* elenca le *probe* e le corrispondenti sequenze geniche interrogate dalle *probe* (specificandone *gene symbol* e *gene title*); l'annotazione include inoltre i termini *Gene Ontology* (GO): *biological process term*, *molecular function term* e *cellular component term*.



## **5. LA PATOLOGIA**

### **5.1 Il diabete mellito di tipo 2**

Il diabete mellito di tipo 2 (chiamato anche diabete mellito non insulino-dipendente, NIDDM) o diabete dell'adulto, è una malattia metabolica caratterizzata da glicemia alta in un contesto di insulino-resistenza e insulino-deficienza relativa. Si differenzia dal diabete mellito di tipo 1 che presenta una carenza assoluta di insulina a causa della distruzione delle Isole di Langerhans del pancreas. Il diabete di tipo 2 rappresenta circa il 90% dei casi di diabete con il restante 10% dovuto principalmente al diabete mellito tipo 1 e al diabete gestazionale.

#### **5.1.1 Sintomi**

I sintomi classici che presentano i pazienti diabetici sono minzione frequente, aumento della fame e della sete e perdita di peso. Altri sintomi comunemente presenti sono: vista offuscata, prurito, neuropatia periferica. Spesso tali sintomi non si presentano durante i primi anni e l'eventuale diagnosi del diabete viene fatta tramite test di routine. Nei casi estremi le persone con diabete mellito tipo 2 possono presentarsi con coma iperosmolare-iper-glicemico non chetotico, dovuto ad una presenza di glucosio nel sangue molto elevata, associata ad una diminuzione del livello di coscienza e ipotensione.

#### **5.1.2 Cause**

Lo sviluppo del diabete di tipo 2 è causato da una combinazione tra lo stile di vita e fattori genetici. Alcuni fattori, come la dieta e l'obesità, sono sotto controllo del paziente, mentre l'invecchiamento, il sesso femminile e la genetica non lo sono. Altre cause che influiscono sullo sviluppo del diabete di tipo 2 sono:

- la mancanza di sonno che si ritiene agisca attraverso il suo effetto sul metabolismo
- lo stato nutrizionale della madre durante lo sviluppo fetale può svolgere un ruolo nella malattia nel nascituro con un meccanismo proposto riguardante l'alterazione della metilazione del DNA

#### **5.1.3 Fisiopatologia**

Il diabete di tipo 2 è dovuto alla insufficiente produzione di insulina dalle cellule beta del pancreas in un quadro di insulino-resistenza. La resistenza all'insulina, che è l'incapacità delle cellule di rispondere adeguatamente a livelli normali di insulina, si verifica

principalmente nei muscoli, nel fegato e nel tessuto adiposo. Nel fegato, l'insulina sopprime normalmente il rilascio di glucosio. Tuttavia, nella condizione di insulino-resistenza, il fegato rilascia impropriamente glucosio nel sangue. La percentuale delle cellule beta non funzionanti differenzia la gravità della malattia.

Altri meccanismi, potenzialmente importanti, associati al diabete di tipo 2 ed alla insulino-resistenza sono: una maggiore ripartizione dei lipidi nelle cellule adipose, la mancanza di incretine, elevati livelli di glucagone nel sangue, l'aumento della ritenzione di acqua e sale dai reni e la regolamentazione inadeguata di metabolismo da parte del sistema nervoso centrale. Tuttavia, non tutte le persone con diabete sviluppano insulino-resistenza, dal momento che è necessaria anche una riduzione della secrezione di insulina dalle cellule beta pancreatiche.

Il diabete di tipo 2 viene trattato inizialmente con l'aumento dell'esercizio fisico e con modifiche nella dieta. Se, tramite queste misure, i livelli di glucosio nel sangue non vengono adeguatamente controllati, può rendersi necessaria la somministrazione di farmaci come la metformina o l'insulina. Nei pazienti che necessitano di insulina, solitamente, vi è l'obbligo di controllare regolarmente i livelli di zucchero nel sangue.

## **5.2 L'intolleranza al glucosio**

L'intolleranza al glucosio è una condizione di elevazione del glucosio nel sangue: la glicemia supera i valori normali, ma non è più elevata di un certo limite (180 mg/100 ml) due ore dopo l'assunzione di una dose standard di glucosio, e non si accompagna a perdita di glucosio con le urine. Nel diabete invece la glicemia supera la soglia dei 180 mg/100 ml, dopo la dose standard di glucosio, e può comparire glucosio nelle urine. Il diabete manifesto viene trattato con insulina. Per diagnosticare precocemente sia l'intolleranza al glucosio che il diabete si esegue un "test di tolleranza al glucosio" che consiste nel far assumere per via orale una quantità standard di glucosio, e nel misurare la glicemia più volte nel tempo: prima della somministrazione e dopo 30, 60, 120 minuti dalla somministrazione.

## 6. PRE-PROCESSING DATI AFFYMETRIX

Il Data Pre-processing è l'insieme delle procedure che permette di acquisire il segnale delle *probe* dai microarray e di passare, attraverso l'uso di opportuni algoritmi, al segnale di espressione del gene.

Il Pre-processing è costituito da quattro passi:

- Analisi dell'immagine
- Controllo di qualità
- Trasformazione dei dati
- Normalizzazione

### 6.1 Analisi dell'immagine

E' costituito da tre passi:

1. *Gridding*: identificazione della posizione di ogni singolo spot
2. *Segmentation*: per ogni spot definisce i pixel appartenenti al *foreground* e quelli appartenenti al *background*
3. Estrazione dell'intensità: ad esempio calcolando la media (o la mediana) dei pixel *foreground* e sottraendo la media (mediana) dei pixel *background*.

### 6.2 Controllo di qualità

Artefatti o difetti nell'immagine (polvere, strisci sul vetrino) possono determinare una scarsa affidabilità del segnale per certi spot o per certe aree dell'array.

Esistono algoritmi opportuni (<http://www.bioconductor.org>) per determinare quali e quanti spot abbiano un segnale non affidabile.

Se il loro numero è limitato, assieme alla misura si assegna un flag allo spot, di cui poter tenere conto nell'analisi successiva; se il loro numero è elevato, si può scegliere di non considerare l'array nelle analisi successive.

I passi 3. e 4. del pre-processing per i microarray Affymetrix sono spiegati nelle seguenti procedure.

Lo scanner Affymerix produce un file immagine .DAT che contiene i valori di intensità dell'immagine pixel per pixel. L'obiettivo è passare dall'intensità dei pixel ad un indicatore generale dell'intensità delle singole *probe cell* (salvate in un file .CEL) e da queste ad un indicatore generale del segnale del *probe set*, che rappresenta la concentrazione di trascritto misurata.

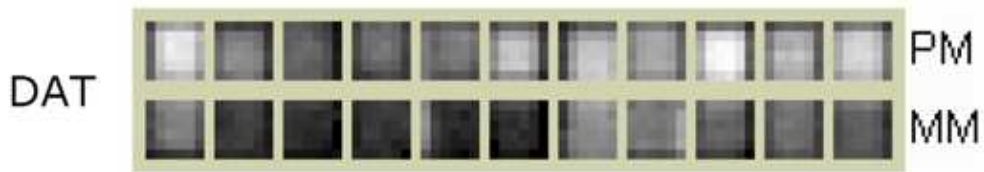


Figura 6.1: Immagine .DAT da cui si ricavano i valori di intensità della *probe pair* pixel per pixel.

## 6.3 L' algoritmo RMA

L' algoritmo Robust Multi-array Analysis (RMA) prevede tre passi successivi:

1. Correzione del segnale di ibridazione non specifica
2. Normalizzazione quantile sulle *probe cell*
3. Regressione robusta di un modello additivo in scala logaritmica (effetto del *probe* + effetto del campione)

### 6.3.1 Correzione del segnale di ibridazione non specifica

Si assume un modello a rumore additivo per la correzione del *background*. Vengono considerati solo i valori di intensità del PM e si assume che ciascun array  $j$  ( $j=1, \dots, M$ ) abbia un unico parametro di *background*  $b_j$

$$I_{PM\_ij} = 2^{H_{ij}} + b_j \quad (6.1)$$

$$H_{ij} = \log_2(I_{PM\_ij} - b_j) \quad (6.2)$$

$i$  indica la *probe pair*,  $j$  indica l'array;

$I_{PM\_ij}$ : segnale di intensità misurato

$H_{ij}$ : segnale di intensità "vero"

$b_j$ : background array specifico.

In una prima versione di RMA, tutti i MM di un array venivano usati per stimare  $b_j$  come

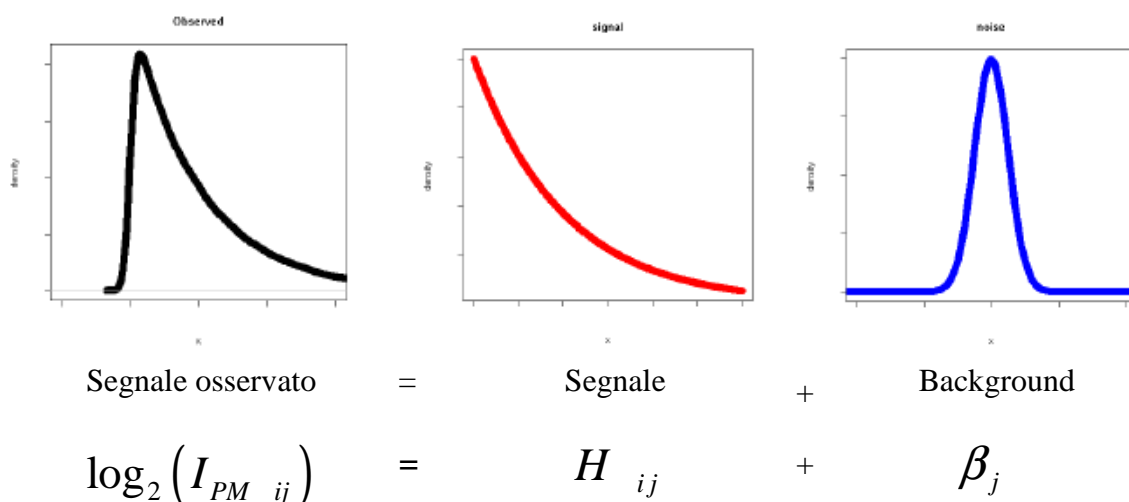
$\log_2(b_j) = (\text{moda del } \log_2(\text{MM}))$ . Tuttavia per una piccola percentuale dei PM,  $I_{PM\_ij} \leq b_j$

e non era quindi possibile valutare il logaritmo della differenza:

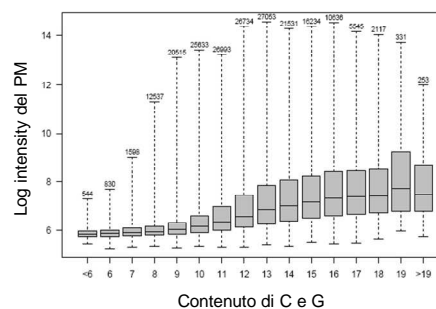
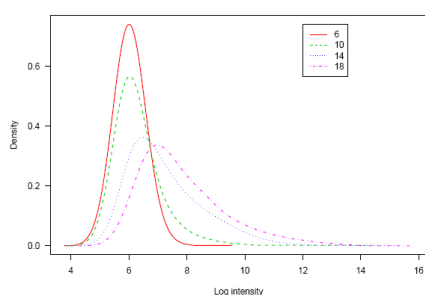
$$H_{ij} = \log_2(I_{PM\_ij} - b_j) \quad (6.3)$$

Si può porre  $H_{ij}=0$  in tutti i casi in cui  $I_{PM\_ij} - b_j \leq 1$ , quindi sia quando il logaritmo non è definito  $I_{PM\_ij} - b_j \leq 0$  sia quando il logaritmo ha risultato negativo ( $0 < I_{PM\_ij} - b_j \leq 1$ ), dato che valori negativi dell'intensità non hanno molto senso.

Nella versione attuale di RMA si utilizza un modello che ipotizza, per ogni array, una distribuzione esponenziale di  $H_{ij}$  e Gaussiana del background  $b_j$ .  $H_{ij}$  viene identificato utilizzando uno stimatore Bayesiano.



È possibile utilizzare un'opzione della correzione del background che tiene conto del contenuto di guanina (G) e citosina (C): le ibridizzazioni non specifiche sono causate da omologia parziale tra due strand con complementarità imperfetta ed il problema è più pronunciato per sequenze con alto contenuto di C e G. È possibile sottrarre un segnale alle intensità dei PM che tiene conto di questa dipendenza della cross ibridazione dal contenuto G e C delle probe (GCRMA).



**Figura 6.2:** Segnale di log intensità del PM, tenendo conto del contenuto di C e G; **Figura 6.3:** Numero di probe cell con il contenuto di C e G indicato in ascissa

### 6.3.2 Normalizzazione

In RMA la normalizzazione *between array* viene eseguita direttamente sulle *probe cell* (non sui *probe set*). In particolare viene eseguita una normalizzazione quantile sui soli PM (il segnale MM non viene utilizzato per il calcolo del segnale). L'obiettivo è quello di rendere uguali le distribuzioni dell'intensità dei PM di ciascun array .

L'ipotesi di base è che la maggior parte dei geni, e quindi dei PM, non è differenzialmente espressa nei diversi array. Dato che in ciascun array vado a monitorare un numero molto elevato di geni e quindi ho un numero ancor più elevato di PM, la distribuzione delle intensità dei PM deve essere circa la stessa nei diversi array.

Data una matrice di dati di dimensioni  $N \times M$  dove ciascun array è rappresentato da una colonna, la normalizzazione quantile prevede i seguenti passi:

1. si ordina ciascuna colonna della matrice in ordine crescente
2. si tiene memoria dell'ordinamento
3. si costruisca un vettore (Medie) delle medie riga per riga
4. la matrice dei valori delle intensità normalizzate è formata dai valori di Medie riarrangiati per ciascuna colonna secondo l'ordine iniziale della matrice dei dati.

### 6.3.3 Calcolo del segnale

Per il calcolo del segnale viene proposto un modello per ogni singolo *probe-set* (per ogni singolo trascritto): l'intensità log trasformata dei PM, (corretta dell'effetto del "background" e normalizzata) segue un modello additivo lineare

$$H'_{ij} = \theta_j + \varphi_i + \varepsilon_{ij} \quad (6.4)$$

$H_{ij}$ : intensità log trasformata dei PM corretta dell'effetto del "background" e normalizzata

$\theta_j$ : espressione del trascritto ibridato all'array j

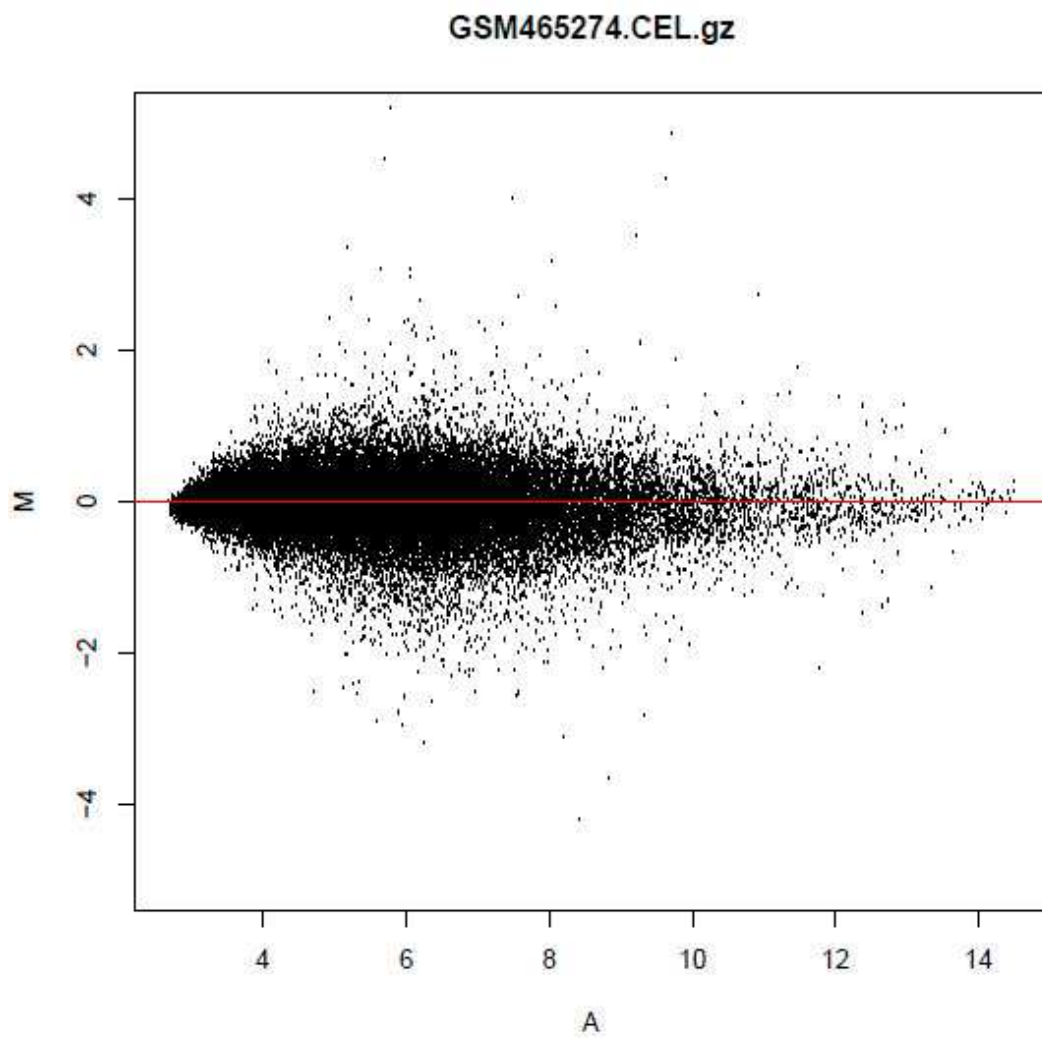
$\varphi_i$ : effetto *probe*-specifico della PP i-esima

$\varepsilon_{ij}$ : errore gaussiano a media nulla ,  $\varepsilon_{ij} \sim N(0, \sigma^2)$

Per identificare i parametri  $\theta_j$  e  $\varphi_i$  ( $\theta_j$  rappresenta il valore di interesse) si esegue un fit del modello utilizzando un metodo robusto agli *outliers* detto Median Polish

$$\begin{bmatrix} H_{1,1} & H_{1,2} & \dots & H_{1,M} \\ H_{2,1} & H_{2,2} & \dots & H_{2,M} \\ \dots & \dots & \dots & \dots \\ H_{16,1} & H_{16,2} & \dots & H_{16,M} \end{bmatrix} = \begin{bmatrix} \Phi_1 & \Phi_1 & \dots & \Phi_1 \\ \Phi_2 & \Phi_2 & \dots & \Phi_2 \\ \dots & \dots & \dots & \dots \\ \Phi_{16} & \Phi_{16} & \dots & \Phi_{16} \end{bmatrix} + \begin{bmatrix} \theta_1 & \theta_2 & \dots & \theta_M \\ \theta_1 & \theta_2 & \dots & \theta_M \\ \dots & \dots & \dots & \dots \\ \theta_1 & \theta_2 & \dots & \theta_M \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,1} & \varepsilon_{1,2} & \dots & \varepsilon_{1,M} \\ \varepsilon_{2,1} & \varepsilon_{2,2} & \dots & \varepsilon_{2,M} \\ \dots & \dots & \dots & \dots \\ \varepsilon_{16,1} & \varepsilon_{16,2} & \dots & \varepsilon_{16,M} \end{bmatrix} \quad (6.5)$$





**Figura 6.4:** MA-plot del segnale di espressione dei geni nel chip HGU133 Plus 2.0



# 7. I METODI DI ANALISI

## 7.1 SAM (Significance analysis of microarrays)

### 7.1.1 Descrizione

*Significance analysis of microarrays* (Tusher V., Tibshirani R., Chu G., 2001) è un metodo statistico per determinare quali geni in un microarray sono differenzialmente espressi.

Il test SAM rappresenta una variante del t-test, che consente di minimizzare l'effetto della sottostima della varianza  $\sigma$  e si basa sull'assunzione che le osservazione sui due gruppi sono campioni indipendenti.

Il test permette, definendo una opportuna variabile  $t_{sam}$ , di selezionare i geni differenzialmente espressi senza fare alcuna ipotesi a priori sulla distribuzione dei dati. In particolare la distribuzione di  $t_{sam}$  in ipotesi nulla viene stimata a partire dai dati stessi mediante permutazioni dei campioni di espressione genica. Permutando i campioni, ovvero riassegnando i valori sulle colonne della matrice dei dati in maniera casuale ai gruppi, vengono generate classi di soggetti miste e bilanciate. I valori assunti dai  $t_{i,sam}$  in questa particolare situazione permettono di determinare una stima della distribuzione di  $t_{sam}$  in ipotesi nulla.

La distribuzione dei dati in ciascun gruppo ha media  $\mu_k$  e varianza  $\sigma_k$ , mentre il valore medio e la varianza osservati sono rispettivamente  $m_k$  e  $s_k$ . L'ipotesi nulla sostiene che i geni non siano differenzialmente espressi in un gruppo piuttosto che in un altro.

Dati  $n$  campioni che appartengono a  $K$  gruppi (nel nostro caso  $K = 3$ : normali, intolleranti al glucosio e diabetici), sia  $C(k)$  l'insieme dei campioni appartenenti al  $k$ -esimo gruppo, contenente  $n_k$  campioni, l'espressione genica media in ogni gruppo è:

$$\bar{x}_k(i) = \sum_{j \in C(k)} \frac{x_j(i)}{n_k} \quad (7.1)$$

mentre l'espressione media per tutti gli  $n$  campioni del microarray è:

$$\bar{x}(i) = \sum_j \frac{x_j(i)}{n} \quad (7.2)$$

Si definisce la varianza tra i  $K$  gruppi  $r(i)$ , come:

$$r(i) = \left\{ \frac{\sum_k n_k}{\prod_k n_k} \cdot \sum_k n_k \cdot [\bar{x}_k(i) - \bar{x}(i)]^2 \right\}^{1/2} \quad (7.3)$$

e la varianza tra i campioni di ogni gruppo  $s(i)$ , come:

$$s(i) = \left\{ \frac{\sum_k 1/n_k}{\sum_k (n_k - 1)} \cdot \sum_k \sum_{j \in C(k)} [\bar{x}_j(i) - \bar{x}_k(i)]^2 \right\}^{1/2} \quad (7.4)$$

La variabile che permette di valutare  $H_0$  è:

$$t_{sam} = \frac{r(i)}{s(i) + s_0} \quad (7.5)$$

in cui  $s_0$  è una costante positiva calcolata sulla base delle seguenti considerazioni.

Poiché negli esperimenti con i microarray il numero di soggetti  $n_k$  è basso,  $s_k$  non rappresenta una stima robusta di  $\sigma_k$ , perciò il valore di  $t_{sam}$  tende a dipendere molto da  $s_k$ . Per  $s_0 = 0$  la variabile  $t_{sam}$  è la stessa del t-test e si può osservare che:

-se  $\sigma_k$  è sottostimata,  $t_{sam}$  risulta alto e perciò vengono selezionati molti geni e conseguentemente più falsi positivi

-se  $\sigma_k$  è sovrastimata,  $t_{sam}$  risulta basso e perciò viene selezionato un numero minore di geni e conseguentemente aumenta il numero di falsi negativi.

La presenza di  $s_0$  al denominatore di  $t_{sam}$  permette di minimizzare l'effetto della sottostima di  $\sigma_k$  e quindi di avere un minor numero di falsi positivi.

Per fissare  $s_0$  si sfrutta la grande quantità di geni analizzati nei microarray e si suddivide il range di  $s_k$  in  $L$  intervalli equidistanti. Quindi si ipotizza che:

- in ogni intervallo ci sono più geni  $\in G_0$  che  $\in G_1$ , dove  $G_0$  è l'insieme dei geni che verificano nella realtà l'ipotesi nulla e  $G_1$  il suo complementare
- $t_{sam}$  dei geni  $\in G_0$  deve avere valori simili nei vari intervalli

Ora per i diversi valori osservati di  $s_0 = c$  ( $c_1, \dots, c_q, \dots, c_Q$ ) che vanno dal minimo al massimo valore osservato per  $s_k$ :

1. Calcolo  $t_{sam}$  per ciascun gene
2. Calcolo  $MAD(t_{sam})^{(1)}$  in ognuno degli  $L$  intervalli in cui è stato diviso il range di  $s_k$  (ottenendo  $L$  valori  $v_1, \dots, v_L$ )
3. Calcolo  $w_q = CV(v_1, \dots, v_L)^{(2)}$

Si ottengono  $Q$  valori di  $w_q$  (uno per ogni  $c_q$ ) e si sceglie il  $c_q$  che minimizza  $w_q$ .

$$c = c_q \leftarrow \operatorname{argmin}(w_q) \quad (7.6)$$

L'assunzione come vera dell'ipotesi nulla non avviene direttamente sul confronto del p-value con un livello di significatività, ma trattandosi di test multipli si ricorre alla correzione tramite FDR (Storey, Tibshirani, 2001), al fine di controllare il numero di falsi positivi tra i geni che verrebbero selezionati come differenzialmente espressi. Il controllo diretto della FP-rate del singolo test statistico non è infatti un criterio adeguato nel caso di test statistici multipli. Si supponga di essere disposti a commettere errori di tipo falso positivo con probabilità  $\alpha$  e si scelga come soglia di confidenza  $\theta$  che per i singoli test corrisponde al livello di significatività  $\alpha$ .

Sia  $G$  il numero di geni per i quali si esegue il test e si supponga di conoscere il numero vero  $G_0$  di geni non differenzialmente espressi. Considerando i test statistici come  $G$  prove ripetute e indipendenti, il numero atteso di errori di tipo falso positivo commessi complessivamente su tutti i test svolti è dato da  $E[\#FP] = G_0 \cdot \alpha$ .

Perciò la probabilità di commettere errori di tipo falso positivo sui  $G$  test aumenta notevolmente rispetto al valore desiderato  $\alpha$ : per questo motivo al fine di controllare il numero di falsi positivi commessi globalmente sui  $G$  test, si preferisce valutare il valore della FDR, che consente di scegliere il livello di significatività  $\alpha$  in modo tale che la percentuale di falsi positivi non superi una certa soglia.

Per il calcolo della *false discovery rate* è necessario conoscere  $G_0$ . Poichè non si dispone di questa informazione è possibile approssimare  $G_0$  con il numero totale di geni  $G$ : si tratta di una scelta molto conservativa che tende a sovrastimare  $G_0$  e, di conseguenza, il numero atteso di errori di tipo falso positivo.

In alternativa è possibile stimare  $G_0$  a partire dai dati. I geni selezionati  $S_\alpha$ , caratterizzati da un livello di espressione genica superiore al valore di soglia  $\alpha$ , possono essere pensati come la somma di falsi positivi, geni selezionati ma non differenzialmente espressi nella realtà, e di veri positivi, geni selezionati e realmente differenzialmente espressi. Il numero di falsi positivi è approssimabile con il valore atteso  $G_0 \cdot \lambda$ , mentre il numero di veri positivi corrisponde ad una frazione  $k \approx 1$  dei  $G_1$  geni realmente differenzialmente espressi. In formule:

$$S_\lambda = \#FP + \#VP \approx G_0 \cdot \lambda + k \cdot G_1 \approx G_0 \cdot \lambda + G_1 \quad (7.7)$$

dove  $\alpha = \lambda$ . Sostituendo  $G_1$  con  $G - G_0$  è infine possibile stimare  $G_0$  come:

$$G_0 \cong \frac{G - S_\lambda}{1 - \lambda} \quad (7.8)$$

La scelta di  $\lambda$  avviene calcolando  $G_0$  per diversi valori di  $\lambda$  e scegliendo  $\lambda$  che minimizza l'errore quadratico medio sulla stima di  $G_0$ .

<sup>(1)</sup> MAD (*Mean Absolute Deviation*) è la mediana dello scostamento della mediana:  $MAD(x) = \operatorname{mediana}(|x - \operatorname{mediana}(x)|)$

<sup>(2)</sup> CV è il coefficiente di variazione o deviazione standard relativa. Dato un campione il CV è il rapporto tra la sua deviazione standard e il valore assoluto della sua media aritmetica.

La *false Discovery Rate* risulta perciò definita come:

$$FDR = E [\#FP]/\#S_\alpha \quad \text{per } S_\alpha > 0 \quad (7.9)$$

$$0 \quad \text{per } S_\alpha = 0$$

### 7.1.2 Implementazione

L'algoritmo SAM è stato implementato servendosi del linguaggio di programmazione R. Si è utilizzata la funzione *sam()* del pacchetto *siggenes*, che fornisce la procedura per vari tipi di analisi (analisi di una o due classi mediante t-test modificato o test di Wilcoxon, analisi di più classi mediante F test modificato).

*sam()* riceve in ingresso la matrice (*data*) contenente i valori di espressione e il vettore *cl* contenente le *labels* (una per ogni colonna della matrice *data*) che indicano la classe di appartenenza dei campioni del dataset, specificando inoltre il numero di permutazioni usate per stimare la distribuzione in ipotesi nulla.

*sam()* non fa uso in maniera esplicita di parametri standard come il livello di significatività  $\alpha$  o la corrispondente soglia di confidenza  $\theta$ , ma di un parametro  $\Delta$  ad essa correlato.

Per avere un controllo diretto sulla FDR si è calcolato il q-value come:

$$FDR = \frac{\alpha \cdot N}{S_\alpha} \cdot p0 \quad (7.10)$$

dove  $N$  rappresenta il numero di geni nell'array e  $p0$  è un valore calcolato tramite la funzione *p0.global()* come:

$$p0 = \frac{1 - S_\alpha / N}{1 - \lambda} \quad (7.11)$$

## 7.2 GSEA (Gene Set Enrichment Analysis)

*Gene Set Enrichment Analysis* (GSEA) è un metodo computazionale che determina se un set di geni definito a priori risulta essere statisticamente significativo, confrontandone le differenze di espressione tra due stati biologici (*e.g.* fenotipo).

GSEA nasce dalla necessità di identificare gruppi funzionali di geni caratterizzati da una, seppur bassa, coerente espressione differenziale. Dato infatti l'elevato numero di geni  $N$ , molto maggiore del numero di campioni  $M$ , e la conseguente correzione per i test multipli risulta difficile che i singoli geni raggiungano la soglia della significatività statistica, in

quanto le differenze biologiche risultano essere modeste in relazione al rumore introdotto dalla tecnologia di preparazione ed elaborazione dei microarray.

Per ovviare a queste difficoltà, GSEA valuta i dati dei microarray a livello di gene set che sono liste di geni definite sulla base di conoscenze biologiche *a priori*. Ad ogni gene della lista viene assegnato un rango in base all'espressione differenziale tra le classi analizzate.

Lo scopo di GSEA è verificare la posizione dei geni nella lista per valutare se il set di geni risulta correlato alla distinzione tra le classi fenotipiche.

### 7.2.1 Descrizione

GSEA analizza i campioni appartenenti a due classi distinte. Dato un set di geni definito a priori (e.g. geni appartenenti ad una stessa categoria GO, geni con la stessa funzione citogenetica, ecc.), sono tre i passi chiave del metodo GSEA:

- **Step 1: Calcolo dell'Enrichment Score.**

Dopo aver ordinato gli  $N$  geni secondo una statistica a scelta ( di default è utilizzato il rapporto segnale rumore:  $SNR = \frac{\mu_a - \mu_b}{\sigma_a - \sigma_b}$  ) dato il set di geni  $S$  contenente  $N_S$  geni, si definisce l'ES come la somma cumulativa (running sum RS) al variare del rango  $k$ :

$$RS(k) = \sum_{g_i \in S; i \leq k} \frac{|y_i|^p}{\sum_{g_j \in S} |y_j|^p} - \sum_{g_i \notin S; i \leq k} \frac{1}{N - N_S} \quad (7.12)$$

Il default per  $p$  è 1.

La somma cumulativa aumenta quando si incontra un gene appartenente a  $S$ , mentre decresce nel caso di un gene che non appartiene ad  $S$ . La grandezza dell'incremento dipende dalla correlazione del gene con il fenotipo.

L'*Enrichment Score* rappresenta la massima deviazione dallo zero di  $RS(k)$  ed è tanto più alto quanto più i geni di  $S$  sono associati al fenotipo. Perciò i geni appartenenti a  $S$  che stanno in cima alla lista relativa ai ranghi assegnati (*ranked list*) fino al raggiungimento dell'ES, sono proprio quelli che contribuiscono a definire l'associazione di  $S$  al fenotipo.

- **Step 2: Stima del livello di significatività dell'ES**

La significatività statistica dell'ES viene valutata utilizzando un tipo di test basato sulle permutazioni. Nello specifico vengono riassegnati i soggetti alle due classi, permutando le etichette (*label*) degli array B volte e ricalcolando l'*Enrichment Score* per i dati permutati.

Il p-value viene quindi calcolato relativamente alla distribuzione ottenuta per i B *Enrichment Score*.

- **Step 3: Correzione per test multipli**

Poiché GSEA è pensato per valutare interi database di gene set, il livello di significatività viene aggiustato con una correzione per test multipli. Il primo passo consiste nella normalizzazione dell'ES per ogni gene set, tenendo conto della dimensione del set di geni, e ottenendo il corrispondente valore del *Normalized Enrichment Score (NES)*:

$$NES = \frac{ES \text{ attuale}}{\text{media}(ES \text{ di tutte le permutazioni del dataset})} \quad (7.13)$$

Pertanto un cambiamento del metodo di permutazione, o del numero di permutazioni o delle dimensioni del dataset, determina un cambiamento del valore del NES.

Successivamente viene controllata la proporzione di falsi positivi calcolando la *False Discovery Rate (FDR)* per ogni NES.

### **Leading edge subset**

Non tutti i membri di un gene set partecipano in egual modo ad uno specifico processo biologico; per questo è utile estrarre dal set quei geni che contribuiscono a definire l'*Enrichment Score*. La *Leading Edge Analysis* è un'analisi a posteriori sui gene set selezionati impostando una soglia sulla FDR; essa valuta le sovrapposizioni tra i *leading edge subset* dei gene set selezionati.

Si definisce come *leading edge subset* l'insieme dei geni, per il gene set in analisi, che si trovano nella *ranked list* prima che lo score assuma il valore di picco; sono perciò quei geni che contribuiscono a definire l'ES. Nel caso di ES negativi il *leading edge subset* è dato dall'insieme dei geni che si trovano dopo il valore di picco dello *score*.



## 7.2.2 Implementazione

La *Gene Set Enrichment Analysis* è stata realizzata servendosi del software GSEA, un'applicazione Java *open source*, scaricabile dal sito <http://www.broadinstitute.org>, che implementa l'algoritmo GSEA e che provvede oltre all'analisi anche all'annotazione e all'interpretazione dei risultati conseguiti.

L'applicazione richiede in ingresso tre tipi di dati:

- un dataset contenente i valori di espressione genica di vari campioni relativa ai geni testati in un microarray (come i microarray Affymetrix)
- dei *phenotype labels*: sono delle etichette che associano ogni campione/soggetto presente nel dataset ad un gruppo fenotipico
- uno o più gene sets, che raggruppi i geni o le relative *probe* oggetto dell'analisi

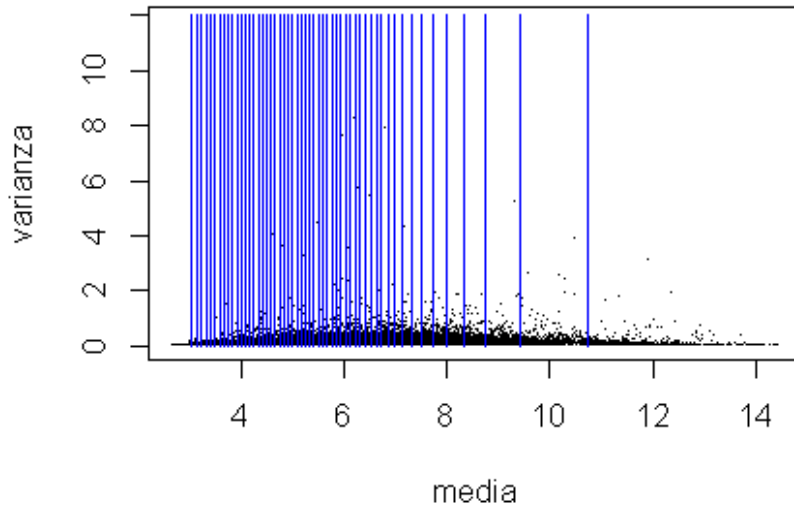
GSEA provvede anche a fornire una collezione di genesets annotati che possono essere esportati dal *Molecular Signature Database* (MSigDb). L'MSigDb contiene 8513 gene sets ed è suddiviso in 6 tipi di raggruppamenti; per il seguente lavoro di analisi si è fatto utilizzo dei gene sets appartenenti al gruppo *c5*, contenente 1454 gene sets che derivano dal vocabolario controllato del *Gene Ontology (GO) project*. I suddetti dati vengono caricati dall'applicazione assieme al file contenente l'annotazione relativa al chip analizzato (HG-U133 Plus 2.0 nel nostro caso); devono essere poi specificate il numero ed il tipo di permutazioni per la stima dell'*Enrichment Score* e se far corrispondere ad ogni probe del dataset il simbolo del gene corrispondente (*HUGO gene symbol*), nel qual caso è necessario specificare il file con l'annotazione relativa al chip (che altrimenti è opzionale).

## 7.3 ANALISI DELLA VARIANZA

A partire dai dati di espressione sono stati calcolati:

- i valori che la varianza assume nei tre gruppi fenotipici per ogni gene
- i valori medi di espressione nei tre gruppi per ogni gene.

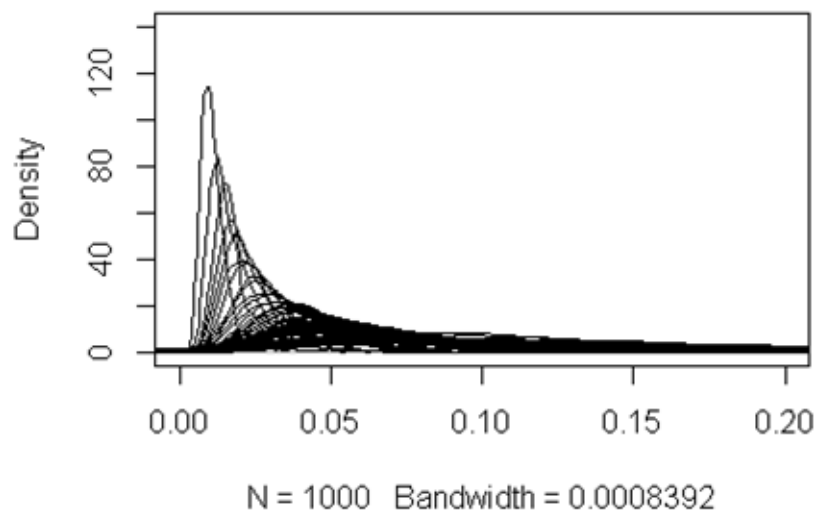
Relativamente all'espressione nei soggetti normali i valori medi nei tre gruppi sono stati ordinati in ordine crescente e suddivisi poi in gruppi (*bin*) di mille geni ciascuno in base al valore medio di espressione.



**Figura 7.1:** Grafico rappresentante la varianza di espressione dei geni al variare del loro valore medio e la suddivisione in *bin*

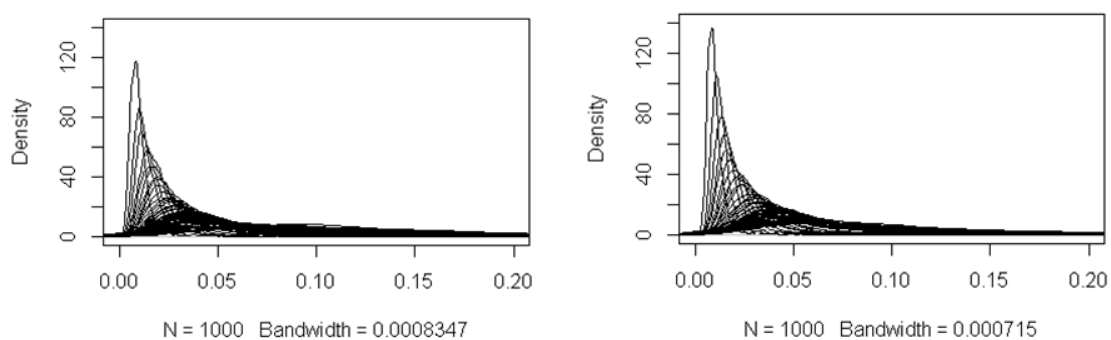
Da tale suddivisione risulta che la distribuzione dei dati della varianza in ciascun *bin* è approssimabile ad una distribuzione gaussiana, che al variare dell'intensità di espressione presenta le seguenti caratteristiche:

- al crescere del valor medio di espressione, diminuisce il valore di picco della gaussiana
- al crescere del valor medio di espressione, aumenta la varianza della gaussiana



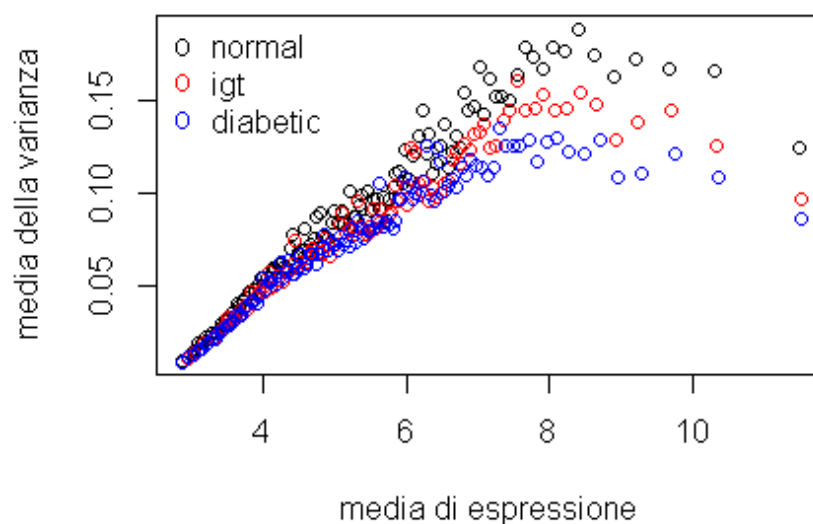
**Figura 7.2:** Density plot della varianza dell'espressione genica per i soggetti normali

Le stesse considerazioni valgono anche per la distribuzione della varianza negli altri due gruppi fenotipici, come risulta evidente in *Figura 7.3*.



**Figura 7.3:** Density plot della varianza dell'espressione genica per soggetti intolleranti al glucosio e diabetici

Dal confronto tra le varianze dei gruppi fenotipici consegue che, a parità del livello di espressione, la varianza dei soggetti normali è mediamente più alta che nei soggetti diabetici e intolleranti al glucosio.



**Figura 7.4:** Plot del valore medio della varianza per ogni *bin* al variare della media di espressione.

Per confermare tale ipotesi si è seguita la seguente procedura:

- per ogni gene  $g_i$  si valuta il  $bin$   $j$  in cui è situato il valore medio di espressione relativo ai gruppi “diabetici” e “intolleranti al glucosio”
- viene calcolato il p-value per il gene  $g_i$  relativo ai gruppi “diabetici” e “intolleranti al glucosio” come:

$$p_i = \frac{\# \text{geni} \in \text{bin}_j^{NORMALI} \mid \text{var}(g)^{NORMALI} > \text{var}(g_i)^{DIABETICI||IGT}}{\# \text{geni} \in \text{bin}_j^{NORMALI}} \quad (7.14)$$

- si ricorre alla correzione per i test multipli tramite *False Discovery Rate* (presentata nel paragrafo 7.1.1) per selezionare i geni differenzialmente espressi.

## 8. RISULTATI

### 8.1 Risultati SAM

I geni differenzialmente espressi vengono valutati sulla base del confronto della FDR con un valore di soglia pari a 0.25. In *Tabella 1* vengono riportati per ogni gene:

- p-value
- q-value

in ordine crescente di FDR, a partire dal valore minimo.

GENE NAME	p-value	FDR
223986_x_at	0.00460791952446273	0.748350224469609
211000_s_at	0.00460574302697759	0.750449198257963
236287_at	0.0046000548696845	0.751987917117898
214157_at	0.00465289437585734	0.75319296059481
207330_at	0.004599890260631	0.754442727367388
200033_at	0.00459257430269776	0.755736995188305
238918_at	0.00473704618198445	0.756952562893568
208624_s_at	0.00458534979423868	0.757054961377029
200884_at	0.00487760402377686	0.757491904892548
239735_at	0.00486443529949703	0.75781497380503
1559075_s_at	0.00489556470050297	0.757912720735813
213075_at	0.00473227251943301	0.758629081050785
228370_at	0.0045493369913123	

		0.758670641343142
204341_at	0.00471992684042067	0.759098654568109
212295_s_at	0.00458275262917238	0.759148248864503
241004_at	0.00454094192958391	0.759820370816842
220172_at	0.00478666666666667	0.75999420660277
1560543_at	0.00457287608596251	0.760045652066866
237465_at	0.00486412437128486	0.760149448916267
236764_at	0.00471328760859625	0.760492016223375

**Tabella 1: Risultati Sam test, FDR e relativo p-value.**

Il minimo valore osservato per la FDR è 0.74835, largamente superiore al valore di soglia imposto per classificare i geni di interesse; pertanto il metodo SAM non conduce a selezionare alcun gene come differenzialmente espresso.

## 8.2 Risultati GSEA

I risultati vengono divisi in 3 diverse categorie sulla base dei gene sets utilizzati nell'analisi: si tratta di GO gene sets che raggruppano geni annotati sulla base dello stesso termine GO:

- CC: cellular component
- BP: biological process
- MF: molecular function

Per ogni tipologia di gene sets si hanno tre tipi di confronti del fenotipo:

1. normali VS intolleranti al glucosio
2. normali VS diabetici
3. intolleranti al glucosio VS diabetici

L'analisi è stata condotta su due dataset contenenti due diversi tipi di dati:

- il primo contiene i dati di espressione genica di tutti i campioni per ogni gene
- il secondo contiene la varianza dei tre gruppi per ogni gene

Per ogni analisi, nelle seguenti tabelle, vengono riportati i GO gene set significativi, cioè con  $FDR < 25\%$ , il numero di geni nel gene set, il valore della FDR e i rispettivi geni appartenenti al *leading edge subset*.

### 8.2.1 Analisi dei dati di espressione genica

#### 8.2.1.1 Normal VS. Glucose intollerant

GENE SET NAME	SIZE	FDR	GENE IN LEADING EDGE SUBSET
BASOLATERAL_PLASMA_MEMBRANE	30	0.030049337	SLC4A11, STX2, ATP7B, TRIP6, ACTN1, LIMA1, VCL, MUC20, LAYN, PTPRC, NRAP, ACTN3
ADHERENS_JUNCTION	21	0.04763847	TRIP6, ACTN1, CDH1, CTNNA3, LIMA1, VCL, LAYN, PTPRC, NRAP, ACTN3
CELL_MATRIX_JUNCTION	16	0.079121195	LIMA1, VCL, LAYN, PTPRC, NRAP, ACTN3
CELL_PROJECTION	98	0.099840745	

			FGD4, S100A6, FGD6, ACTN1, TGFB2, DTNBP1, MYO6, PPT1, OXCT2, EFHC1, CYFIP1, ALS2, FGD5, CDH23, WASF2, CD2AP, CDK5R1, SLC22A12, ACTN4, APBB1, MYO3A, SPAG6, ITGB1BP1, ATP6V0A4, FAP, LAYN, MYO5A, DOCK4, TLN1, ITLN1, CLIC4, CLIC1, MYH9, APBB2, NF1, ACTN3
ACTIN_FILAMENT	16	0.19417734	MYO6, CD2AP, GAS2, IQGAP1, MYO3A, ACTC1, ACTN3
ACTIN_CYTOSKELETON	121	0.21422967	AMPH, SPTBN4, CDH1, CASK, EPB41L2, ARPC5, KALRN, FLNB, WASF2, CD2AP, SORBS2, FSCN1, LIMA1, GAS2, IQGAP1, DAPK1, MYO3A, IPP, ANLN, IQGAP2, FLNA, CLIC4, MYH9, ACTC1, VASP, MYL9, ACTN3
CELL_CORTEX_PART	23	0.2300069	TRPC4, SPTAN1, SPTBN4, EPB41L2, ANLN, PRKCI, MYH9, DLG4
LAMELLIPODIUM	22	0.23072802	FGD4, FGD6, CYFIP1, ALS2, FGD5, WASF2, APBB1, ITGB1BP1, FAP, APBB2
CELL_JUNCTION	75	0.23824097	TRIP6, ACTN1, CLDN15, PNN, CDC42BPA, EPB41L3, ASH1L, CLDN6, CDH1, CTNNA3, CLDN5, MLLT4, LIMA1, VCL, CLDN23, LAYN, GJA1, PTPRC, NRAP, ACTN3

**Tabella 2: Analisi per Cellular Component: gene set associati al fenotipo Normal.**

L'analisi per CC non presenta gene set significativi associati al fenotipo *Glucose intollerant*. Le analisi per BP e MF non presentano gene set significativi per nessuno dei due fenotipi.

### 8.2.1.2 Normal VS Diabetic

L'analisi per CC non presenta gene set significativi per nessuno dei due fenotipi.



GENE SET NAME	SIZE	FDR	GENE IN LEADING EDGE SUBSET
CELL_SUBSTRATE_ADHESION	37	0.13667038	ACTN3, ECM2, ACTN1, ITGB1BP1, SGCE, NF1, PTEN, PKD2, FBLN5, PARVG
CELL_MATRIX_ADHESION	36	0.20739712	ACTN3, ECM2, ACTN1, ITGB1BP1, SGCE, NF1, PTEN, PKD2, FBLN5, PARVG

**Tabella 3: Analisi per Biological Process: gene set associati al fenotipo Diabetic**

L'analisi per BP non presenta gene set associati al fenotipo *Normal*.

GENE SET NAME	SIZE	FDR	GENE IN LEADING EDGE SUBSET
STRUCTURAL MOLECULE ACTIVITY	226	0.14556448	RPS4Y1, ACTN3, SORBS2, MYL9, MGP, MFAP5, ANXA1, CDC42SE2, MYH11, FBN1, MYOC, VIM, MRPS12, PLP1, CCDC6, LAMA4, PLS1, TFPI2, MPZL1, LAMA2, PRELP, MYBPC2, SMTN, COL4A2, ARPC3, KRT2, MYL6, FBLN1, RPL3, LMNB1, PPL, MSN
INTRAMOLECULAR OXIDOREDUCTASE ACTIVITY	18	0.16613477	MPI, PDIA3, HSD3B2, DDT, HSD3B1, TPI1, PTGIS
DI TRI VALENT INORGANIC CATION TRANSMEMBRANE TRANSPORTER_ACTIVITY	22	0.20256186	ATP2A1, ATP2C1, SLC11A2, SLC31A2, SLC31A1, FXN, SLC30A3, SLC39A2, ATP2B4
ELECTRON_CARRIER_ACTIVITY	74	0.21110788	NOX4, P4HA2, SRD5A1, UGDH, NQO1, HSD17B6, NCF2, AOC2, DHRS3, ALDH1A2, FDX1, FXN, ADH1B, ADH6, NDUFS2, QDPR, DMGDH, CYP1A2, CYB5R4, MTRR, COX11, ME2, ASPH, TSTA3, NDUFS1, NDUFS8, IDH3B, DLD, KMO, NDUFS3, GSR, SDHD, NDUFS7, ADH5, NDUFA9, DEGS1, ETFDH

**Tabella 4: Analisi per Molecular Function: gene set associati al fenotipo Diabetic.**

L'analisi per MF non presenta gene set associati al fenotipo *Normal*.

### 8.2.1.3 Glucose intollerant VS Diabetic

GENE SET NAME	SIZE	FDR	GENE IN LEADING EDGE SUBSET
DETECTION_OF_CHEMICAL_STIMULUS	15	0.12538992	TAS2R16, TAS1R2, KCNMB4, KCNMB2, RTP3

**Tabella 5: Analisi per Biological Process: gene set associati al fenotipo Diabetic.**

L'analisi per BP non presentano gene set significativi associati ai due fenotipi. Le analisi per CC e MF non presentano gene set significativi per nessuno dei due fenotipi.

## 8.2.2 Analisi dei dati della varianza dell'espressione genica

### 8.2.2.1 Normal VS Glucose intollerant

GENE SET NAME	SIZE	FDR	GENE IN LEADING EDGE SUBSET
MITOCHONDRIAL_ENVELOPE	92	0.013426261	SDHD, SLC25A3, NDUFS4, ATP5D, NDUFS3, COX6B2, HSD3B2, ATP5B, UQCRC1, ATP5E, MTX2, NDUFAB1, MRPL32, ATP5G2, VDAC2, ACN9, TIMM17A, VDAC1, NDUFA9, TIMM23, VDAC3, ATP5O, TIMM50, NDUFV1, ATP5G3, NDUFA13, MPV17, NDUFS2, COX18, SLC25A11, NDUFS7, SDHA, RHOT1, NDUFS8, TIMM8B, PHB, NNT, NDUFS1, GATM, COX15, ABCB8, PMPCA, TOMM22, OGDH, PHB2, NDUFA1, TIMM13
MITOCHONDRIAL_MEMBRANE_PART	51	0.01543754	NDUFS4, ATP5D, NDUFS3, COX6B2, ATP5B, UQCRC1, ATP5E, NDUFAB1, ATP5G2, TIMM17A, NDUFA9, TIMM23, ATP5O, TIMM50, NDUFV1, ATP5G3, NDUFA13, NDUFS2, COX18, NDUFS7, SDHA, RHOT1, NDUFS8, TIMM8B
MITOCHONDRIAL_MEMBRANE	82	0.017573144	SDHD, SLC25A3, NDUFS4, ATP5D, NDUFS3, COX6B2, HSD3B2, ATP5B, UQCRC1, ATP5E, MTX2, NDUFAB1, MRPL32, ATP5G2, VDAC2,

			TIMM17A, VDAC1, NDUFA9, TIMM23, VDAC3, ATP5O, TIMM50, NDUFV1, ATP5G3, NDUFA13, MPV17, NDUFS2, COX18, SLC25A11, NDUFS7, SDHA, RHOT1, NDUFS8, TIMM8B, PHB, NNT, NDUFS1, COX15, PMPCA, TOMM22, OGDH, PHB2, NDUFA1, TIMM13
MITOCHONDRIAL_MATRIX	46	0.020011308	ACADM, CASQ1, MRPS24, TFB2M, MRPS35, CS, NDUFAB1, MRPS15, MRPS36, ETFA, ETFB, NFS1, DNAJA3, ATP5F1, MRPS10, MRPL52, BCKDK, MARS2, MRPS28, MRPL55, MRPS22, MRPS18C, MRPS16, MRPS21, MRPS11, MRPL42, DBT, MRPL41
MITOCHONDRIAL_LUMEN	46	0.020636518	ACADM, CASQ1, MRPS24, TFB2M, MRPS35, CS, NDUFAB1, MRPS15, MRPS36, ETFA, ETFB, NFS1, DNAJA3, ATP5F1, MRPS10, MRPL52, BCKDK, MARS2, MRPS28, MRPL55, MRPS22, MRPS18C, MRPS16, MRPS21, MRPS11, MRPL42, DBT, MRPL41
MITOCHONDRIAL_PART	137	0.023912437	SDHD, SLC25A3, ACADM, NDUFS4, ATP5D, NDUFS3, COX6B2, CASQ1, HSD3B2, MRPS24, TFB2M, ATP5B, UQCRC1, MRPS35, ATP5E, CS, MTX2, NDUFAB1, MRPS15, MRPS36, MRPL32, ATP5G2, ETFA, ETFB, VDAC2, ACN9, TIMM17A, VDAC1, NDUFA9, NFS1, TIMM23, VDAC3, ATP5O, TIMM50, DNAJA3, NDUFV1, ATP5G3, NDUFA13, MPV17, NDUFS2, COX18, SLC25A11, NDUFS7, ATP5F1, SDHA, RHOT1, NDUFS8, TIMM8B, MRPS10, MRPL52, BCKDK, MARS2, PHB, NNT, NDUFS1, GATM, COX15, ABCB8, PMPCA, MRPS28, TOMM22, OGDH, MRPL55, MRPS22, PHB2, NDUFA1, TIMM13, MRPS18C, MRPS16, MRPS21

MITOCHONDRIAL_RESPIRATORY_CHAIN	23	0.030939348	NDUFS4, NDUFS3, UQCRC1, NDUFAB1, NDUFA9, NDUFV1, NDUFA13, NDUFS2, NDUFS7, SDHA, NDUFS8, NNT, NDUFS1, COX15
ORGANELLE_INNER_MEMBRANE	72	0.032115687	SDHD, SLC25A3, NDUFS4, ATP5D, NDUFS3, COX6B2, HSD3B2, ATP5B, UQCRC1, ATP5E, NDUFAB1, MRPL32, ATP5G2, TIMM17A, NDUFA9, TIMM23, ATP5O, MATR3, TIMM50, NDUFV1, ATP5G3, NDUFA13, MPV17, NDUFS2, COX18, SLC25A11, NDUFS7, SDHA, NDUFS8, TIMM8B
ORGANELLAR_RIBOSOME	22	0.032627005	MRPS24, MRPS35, MRPS15, MRPS36, MRPS10, MRPL52, MRPS28, MRPL55, MRPS22, MRPS18C, MRPS16, MRPS21, MRPS11, MRPL42, MRPL41
MITOCHONDRIAL_RIBOSOME	22	0.035138305	MRPS24, MRPS35, MRPS15, MRPS36, MRPS10, MRPL52, MRPS28, MRPL55, MRPS22, MRPS18C, MRPS16, MRPS21, MRPS11, MRPL42, MRPL41
PROTON_TRANSPORTING_TWO_SECTOR_ATPASE_COMPLEX	15	0.036661036	ATP5D, ATP5B, ATP5E, ATP5G2, ATP5O, ATP5G3
SPINDLE_POLE	17	0.037484955	CETN1, KNTC1, KIF11, NUMA1, BUB1, TPX2
RIBOSOMAL_SUBUNIT	20	0.037900466	MRPS24, MRPS35, MRPS15, MRPS36, MRPL52, MRPS28, MRPL55, MRPS22, MRPS18C, MRPS16, MRPS21, MRPS11, MRPL42, MRPL41
ENVELOPE	158	0.04100194	SDHD, SLC25A3, NDUFS4, ATP5D, NDUFS3, COX6B2, NUP153, HSD3B2, RTN4, ATP5B, UQCRC1, ATP5E, MTX2, NDUFAB1, MRPL32,

			ATP5G2, VDAC2, ACN9, TIMM17A, VDAC1, NDUFA9, GUCY2D, NUP214, TIMM23, VDAC3, ATP5O, MATR3, CBX5, TIMM50, NDUFV1, ATP5G3, DHCR7, NDUFA13, MPV17, NDUFS2, COX18, GNAZ, SLC25A11, NDUFS7, SNUPN, SDHA, GUCY2F, RHOT1, NUTF2, MRPL19, TMPO, HTATIP2, NDUFS8, TIMM8B, KPNA1, RANBP2, PHB, NNT, KPNA3, NDUFS1, GATM, NRM, COX15, NUP88, SREBF1, ABCB8, IPO7, PMPCA, TOMM22, OGDH, DDX19B, PHB2, NDUFA1, TIMM13
ORGANELLE_ENVELOPE	158	0.04118819	SDHD, SLC25A3, NDUFS4, ATP5D, NDUFS3, COX6B2, NUP153, HSD3B2, RTN4, ATP5B, UQCRC1, ATP5E, MTX2, NDUFAB1, MRPL32, ATP5G2, VDAC2, ACN9, TIMM17A, VDAC1, NDUFA9, GUCY2D, NUP214, TIMM23, VDAC3, ATP5O, MATR3, CBX5, TIMM50, NDUFV1, ATP5G3, DHCR7, NDUFA13, MPV17, NDUFS2, COX18, GNAZ, SLC25A11, NDUFS7, SNUPN, SDHA, GUCY2F, RHOT1, NUTF2, MRPL19, TMPO, HTATIP2, NDUFS8, TIMM8B, KPNA1, RANBP2, PHB, NNT, KPNA3, NDUFS1, GATM, NRM, COX15, NUP88, SREBF1, ABCB8, IPO7, PMPCA, TOMM22, OGDH, DDX19B, PHB2, NDUFA1, TIMM13
MITOCHONDRIAL_INNER_MEMBRANE	64	0.042894587	SDHD, SLC25A3, NDUFS4, ATP5D, NDUFS3, COX6B2, HSD3B2, ATP5B, UQCRC1, ATP5E, NDUFAB1, MRPL32, ATP5G2, TIMM17A, NDUFA9, TIMM23, ATP5O, TIMM50, NDUFV1, ATP5G3, NDUFA13, MPV17, NDUFS2, COX18, SLC25A11, NDUFS7, SDHA, NDUFS8, TIMM8B, PHB, NNT, NDUFS1, COX15, PMPCA, PHB2, NDUFA1, TIMM13

SPINDLE	37	0.066691674	CETN1, PRC1, KNTC1, KIF11, NUMA1, BUB1, ARL8B, TPX2, KIF23, CLASP2, NEDD9, CDC16, NDE1, CDC20, LATS1
VOLTAGE_GATED_POTASSIUM CHANNEL COMPLEX	40	0.10247575	KCNJ3, KCNH1, KCNC1, KCNMB3, KCNQ3, KCNA20, KCNE1, KCNQ2, KCNA1, KCNK6, KCNMB4, KCNS1, HCN2, KCNJ4, KCND3, KCNQ1, KCNJ6, KCNE2, KCNN1, KCNJ8, KCNMB2
MITOCHONDRION	311	0.10578152	SDHD, SLC25A3, SUCLA2, CYC1, ACADM, AK3, HSPE1, MTIF2, NDUFS4, ATP5D, NDUFS3, ATPIF1, COX6B2, CASQ1, HSD3B2, MRPS24, CLPX, TFB2M, SDHB, COQ4, ECSIT, ATP5B, FDX1, UQCRC1, MRPS35, GLRX2, ATP5E, CS, TMEM126A, MTX2, NDUFAB1, MRPS15, CPT1B, DLD, MRPS36, MRPL32, ATP5G2, ALDH5A1, ETFA, NDUFA8, COX11, ETFB, MIPEP, TFAM, GOT2, VDAC2, ACN9, TIMM17A, VDAC1, TMEM143, SDHC, NDUFA9, NFS1, SUCLG1, FH, TIMM23, VDAC3, ATP5O, PDK1, MRPS31, PDHA1, DNAJC19, DECR1, AASS, TIMM50, DNAJA3, NDUFV1, 0ATP5G3, HINT2, TRAK1, 0NDUFA13, MPV17, NDUFS2, IDH2, MOSC2, COX18, TUFM, CPOX, SLC25A11, NDUFS7, ECHS1, ATP5F10, HSPD1, SDHA, GLRX5, KIAA0141, ACO2, GHITM, PCCB, GLS2, MTCH1, RHOT1, GBAS, NDUFS8, TIMM8B, MRPS10, MRPL52, OXSM, NAPG, TBRG4, BCKDK, MARS2, PHB, CKMT1A, HADH, NNT, NDUFS1, TP53, GATM, COX15, ABCB8, MTIF3, PMPCA, MRPS28, PTS, ABAT, NDUFA4, TOMM22, OGDH, IFI6, MRPL55, MRPS22

**Tabella 6: Analisi per Cellular Component: gene set associati al fenotipo Normal.**

GENE SET NAME	SIZE	FDR	GENE IN LEADING EDGE SUBSET
RIBONUCLEOPROTEIN COMPLEX_BIOGENESIS AND ASSEMBLY	56	0.00948332	GEMIN5, USP39, EIF1B, EIF2A, PAIP2, PAIP1, EIF2B2, MTIF3, EIF1AX, RPS3A, DICER1, CRNKL1, SF3A3, SNRPD1, MTIF2, EIF5A, EIF2B1, SNRPD2, NOLC1, EIF2B3, GEMIN7, GEMIN6, EIF2S3, EXOSC7, EIF2B4, SLU7, DAZL, TXNL4A, EIF2AK3
PROTEIN_RNA_COMPLEX_ASSEMBLY	44	0.010985684	USP39, EIF1B, EIF2A, PAIP2, PAIP1, EIF2B2, MTIF3, EIF1AX, RPS3A, DICER1, CRNKL1, SF3A3, SNRPD1, MTIF2, EIF5A, EIF2B1, SNRPD2, EIF2B3, EIF2S3, EIF2B4, SLU7, DAZL, TXNL4A, EIF2AK3
RNA_SPLICINGVIA TRANSESTERIFICATION_REACTIONS	21	0.012644427	GEMIN5, USP39, CRNKL1, SF3A3, SNRPD1, SNRPD2, GEMIN7, TRA2A, GEMIN6, SLU7, SNW1, TXNL4A
TRANSLATIONAL_INITIATION	26	0.012982973	EIF1B, PAIP2, PAIP1, EIF2B2, MTIF3, EIF1AX, RPS3A, MTIF2, EIF5A, EIF2B1, EIF2B3, EIF2S3, EIF2B4, DAZL, EIF2AK3
MRNA_PROCESSING_GO_0006397	51	0.08263471	GEMIN5, USP39, PRPF18, KHDRBS1, LSM3, CRNKL1, SF3A3, SNRPD1, SNRPD2, SLBP, GEMIN7, LSM1, DHX15, TRA2A, GEMIN6, PABPC1, NCBP1, SMNDC1, CPSF3, SLU7, SNW1, TXNL4A, NONO
RNA_SPLICING	66	0.08444229	GEMIN5, USP39, SNRPB2, LSM4, PRPF18, SNRPG, CRNKL1, SF3B2, SF3A3, SNRPD1, SNRPD2, NOL3, GEMIN7, LSM1, DDX23, TRA2A, GEMIN6, SNRPC, NCBP1, SMNDC1, SLU7, SNW1, TXNL4A, PRPF3, NONO, SNRPB, DHX8, SNRPD3, SF3B3, ZRANB2, PPAN, LSM6
PROTEIN_FOLDING	55	0.08617144	SEP15, AHSA1, NFYC, CCT4, CLPX, HSP90AA1, CCT6A, PFDN4, BAG2, RP2, ARL2, ST13, STUB1, GLRX2,

			HSPE1, ERP29, PDIA6, MKKS, DNAJA1, FKBP9, DNAJB6, CCT6B, CCT7, APOA2, TBCE, TTC1, DNAJA3
DNA_CATABOLIC_PROCESS	22	0.1358637	UBE2V2, ISG20, NTHL1, ERCC1, XRN2, UBE2N, PPT1, ERCC5, BNIP3, TPD52L1, SOD1
MRNA_METABOLIC_PROCESS	62	0.14037934	GEMIN5, USP39, PAIP1, PRPF18, KHDRBS1, LSM3, CRNKL1, SF3A3, SNRPD1, AUH, SNRPD2, UPF2, SLBP, GEMIN7, LSM1, DHX15, TRA2A, GEMIN6, PABPC1, NCBP1, SMNDC1, CPSF3, SLU7, SNW1, TXNL4A, NONO
CELLULAR_PROTEIN_CATABOLIC_PROCESS	53	0.21469122	PSMD14, UBE2A, PCNP, ARIH1, RNF11, UBE2G1, TSG101, ANAPC10, STUB1, UFD1L, UBE2D1, BTRC, UBB, SMURF2, UBE3A, UBE2E1, UHRF2, ANAPC5, EDEM1, ANAPC4, FBOXO7, CDC23, UBE2I, SELS, PARK2, DERL1, UBE4A
RNA_EXPORT_FROM_NUCLEUS	18	0.22194298	KHDRBS1, EIF5A, UPF2, DDX25, NUP107, NUP133, NUDT4, SMG7, DDX19B, RAE1, SMG1
NITROGEN_COMPOUND_BIOSYNTHETIC_PROCESS	25	0.22239481	HSP90AB1, OAZ1, MAT2B, HSP90AA1, PAH, GLA, GCHFR, PYCR1, BBOX1, GCH1
MICROTUBULE_BASED_MOVEMENT	16	0.22332211	KIF3B, KIF5B, RHOT1, KIF1A, LRPPRC, UXT, KIF4A, KIF13B, OPA1
RESPONSE_TO_TEMPERATURE_STIMULUS	16	0.22821267	EIF2B2, FGF16, EIF2B1, EIF2B3, GCLC, EIF2B4, GMPR, SOD1, DNAJB4
NUCLEOBASENUCLEOSIDENUCLEOTIDE AND NUCLEIC ACID TRANSPORT	28	0.22921985	CKAP5, SLC29A1, KHDRBS1, EIF5A, UPF2, DDX25, NUP107, SLC23A1, NUP133, SLC23A2, NUDT4, SMG7, DDX19B, VDAC3, RAE1, SMG1
REGULATION_OF_TRANSLATIONAL	19	0.23033987	



INITIATION			EIF1B, PAIP2, EIF2B2, MTIF3, MTIF2, EIF5A, DAZL, EIF2AK3
CASPASE_ACTIVATION	24	0.23515597	APAF1, MTCH1, IFNB1, MOAP1, LCK, STAT1, DIABLO
FATTY_ACID_OXIDATION	18	0.24388927	PPARD, ADIPOR1, ADIPOR2, BDH2, PPARA, ECHS1, ACADS, HACL1, CPT1B, ACADM

**Tabella 7: Analisi per Biological Process: gene set associati al fenotipo Normal.**

GENE SET NAME	SIZE	FDR	GENE IN LEADING EDGE SUBSET
TRANSLATION_INITIATION_FACTOR_ACTIVITY	17	0.010877346	EIF2A, EIF2B2, EIF2S2, EIF4E, EIF5A, EIF2B1, EIF2B3, COPS5, EIF2S3, EIF2B4
SINGLE_STRANDED_DNA_BINDING	32	0.013321856	HNRPDL, MLH1, MYT2, TERF2IP, WBP11, RAD51, PCBP1, MSH3, RPA4, IGHMBP2, ERCC1, HMGB2, YBX1, RPA2, ERCC5, XPC, MSH2
STRUCTURE_SPECIFIC_DNA_BINDING	53	0.05266372	XRCC6, HNRPDL, MLH1, MYT2, TERF2IP, WBP11, RAD51, PCBP1, NTHL1, MSH3, RPA4, PNKP, APTX, IGHMBP2, ERCC1, HMGB2, YBX1, RPA2, ERCC5, XPC, MSH2
DOUBLE_STRANDED_DNA_BINDING	31	0.06758366	XRCC6, HNRPDL, MLH1, RAD51, NTHL1, MSH3, PNKP, APTX, HMGB2, YBX1, ERCC5, MSH2
TRANSLATION_REGULATOR_ACTIVITY	31	0.085089825	EIF2A, PAIP2, PAIP1, EIF2B2, EIF2S2, EIF4E, EIF5A, EIF2B1, EIF2B3, COPS5, PABPC1, EIF2S3, EIF2B4, DAZL
CARBON_OXYGEN_LYASE_ACTIVITY	31	0.08793819	APEX1, PTS, FH, NTHL1, ALAD, AUH, CA6, ENO3, EHHADH, CA2, CA7, HADHA, ECHS1, CBS,

			CA5B, ENO1, OGG1
TRANSLATION_FACTOR ACTIVITY NUCLEIC_ACID_BINDING	29	0.09568836	EIF2A, PAIP1, EIF2B2, EIF2S2, EIF4E, EIF5A, EIF2B1, EIF2B3, COPS5, PABPC1, EIF2S3, EIF2B4, DAZL
UNFOLDED_PROTEIN_BINDING	40	0.106366195	TOMM20, CCT4, HSPD1, CCT6A, PTGES3, RP2, PPIC, MKKS, TOR1B, TCP1, CCT6B, CCT7, AFG3L2, TTC1, DNAJA3, DNAJB4, CLGN, AIP, NPM1, CCT3, GRPEL1, HSPA2, CHAF1A, LRPAP1
HYDRO_LYASE_ACTIVITY	27	0.1784426	FH, ALAD, AUH, CA6, ENO3, EHHADH, CA2, CA7, HADHA, ECHS1, CBS, CA5B, ENO1
LIGASE_ACTIVITY	89	0.18651162	FBXO3, RTCD1, ARIH1, UBE2M, SUCLA2, ANAPC10, FARS2, STUB1, MARS, ACACA, TRAF6, KARS, HLCS, GSS, PCCA, DZIP3, GCLC, CAD, GMPS, FPGS, ANAPC11, SMURF2, TRAF7, BRAP, UBE2E1, NARS, PAICS, RNF14, UHRF2, ACSL1, ANAPC5, ANAPC4, FBXO7, ACSL4, QARS, CDC23, PCCB, DDB2, CPS1, ASNS, PARK2
EXONUCLEASE_ACTIVITY	18	0.19333309	APEX1, ISG20, WRN, REXO2, APTX, XRN2, ENPP3, EXOSC7

**Tabella 8: Analisi per Molecular Function: gene set associati al fenotipo Normal.**

Le analisi per CC, BP e MF non presentano gene set significativi associati al fenotipo *Glucose intollerant*.

### 8.2.2.2 Normal VS Diabetic

GENE SET NAME	SIZE	FDR	GENE IN LEADING EDGE SUBSET
MITOCHONDRIAL_MATRIX	46	0.018683963	CASQ1, MRPS18C, MRPS36, MRPS24, MRPS10, TFB2M, MRPL51, MRPS35, ETFB, ETFA,

			MRPL40, MRPL10, NR3C1, MRPS11, MRPS22, NDUFAB1, ATP5F1, MRPS21, MRPS28, MRPS18A, DNAJA3, CS, POLG2, GRPEL1, MRPL42, BCKDHA, SUPV3L1, ACADM, MRPL41
NUCLEAR_MEMBRANE_PART	39	0.020608878	NUP153, SNUPN, IPO7, RANBP2, MATR3, NUTF2, GUCY2F, EIF5A, LEMD3, KPNA1, KPNA3, NUP107, NARF, GUCY2D, DHCR7, NXT1, NUP133, XPO7, SENP2, NUP88, NUP214, PSEN1, DDX19B, RAE1
MITOCHONDRIAL_LUMEN	46	0.021261437	CASQ1, MRPS18C, MRPS36, MRPS24, MRPS10, TFB2M, MRPL51, MRPS35, ETFB, ETFA, MRPL40, MRPL10, NR3C1, MRPS11, MRPS22, NDUFAB1, ATP5F1, MRPS21, MRPS28, MRPS18A, DNAJA3, CS, POLG2, GRPEL1, MRPL42, BCKDHA, SUPV3L1, ACADM, MRPL41
NUCLEAR_MEMBRANE	47	0.022147423	NUP153, MRPL19, SNUPN, IPO7, RANBP2, WTAP, MATR3, NUTF2, GUCY2F, EIF5A, LEMD3, KPNA1, KPNA3, NUP107, NARF, GUCY2D, DHCR7, NXT1, NUP133, XPO7, SENP2, NUP88, NUP214, PSEN1, DDX19B, TRIM27, RAE1
NUCLEAR_PORE	28	0.022507098	NUP153, SNUPN, IPO7, RANBP2, NUTF2, EIF5A, KPNA1, KPNA3, NUP107, NXT1, NUP133, XPO7, SENP2, NUP88, NUP214, DDX19B, RAE1
ER_GOLGI_INTERMEDIATE COMPARTMENT	21	0.02349899	TMED9, P4HB, KDELR1, TMED2, FN1, PDIA6, GOLGB1, NUCB2, RAB37, SEC23IP, YIF1A, ERGIC3
MITOCHONDRIAL_PART	137	0.023700073	CASQ1, COX6B2, VDAC1, PHB2, ATP5O, SDHD, MRPS18C, MRPS36, SLC25A3, NDUFS4, MRPS24, TIMM23, TOMM22, MRPL32,

			MRPS10, MAOB, TFB2M, TIMM9,, OXA1L, MRPL51, HSD3B2, MRPS35, RHOT1, ETFB, MTX2, ETFA, MRPL40, MRPL10, NR3C1, PHB, NDUFS3, MRPS11, ATP5G2, NDUFA1, MRPS22, ATP5B, NDUFAB1, ATP5F1, NDUFS8, PMPCA, BNIP3, MRPS21, MRPS28, VDAC2, NDUFV1, MRPS18A, FIS1, SURF1, SLC25A22, TIMM8B, TIMM17A, DNAJA3, CS, TIMM50, HSD3B1, MSTO1, ABCF2, POLG2, NDUFA13, TIMM17B, TIMM13, NDUFB6, MFN2, COX18, GRPEL1, OPA1, UQCRC1, MRPL42, VDAC3, BCKDHA, SUPV3L1, HCCS, ACADM, MRPL41, TIMM10, NDUFA9, ATP5E, HTRA2, ATP5G3, TIMM44
ORGANELLE_ENVELOPE	158	0.024316777	COX6B2, VDAC1, NUP153, PHB2, ATP5O, MRPL19, SDHD, SNUPN, IPO7, RANBP2, WTAP, SP140, SLC25A3, NDUFS4, TIMM23, MATR3, TOMM22, MRPL32, NUTF2, MAOB, TIMM9, HTATIP2, OXA1L, TMPO, GUCY2F, HSD3B2, EIF5A, RHOT1, LEMD3, MTX2, PHB, KPNA1, NDUFS3, ATP5G2, KPNA3, NDUFA1, ATP5B, NUP107, NDUFAB1, NDUFS8, NARF, PMPCA, BNIP3, VDAC2, NDUFV1, NRM, GUCY2D, FIS1, SURF1, SLC25A22, TIMM8B, TIMM17A, TIMM50, DHCR7, NXT1, HSD3B1, MSTO1, NUP133, ABCF2, EMD, NDUFA13, NEU4, POLA1, TIMM17B, XPO7, TIMM13, NDUFB6, SENP2, MFN2, NUP88, BNIP3L, PARP1, NUP214, COX18, OPA1, UQCRC1, PSEN1, DDX19B, VDAC3, TRIM27, RAE1, HCCS
PORE_COMPLEX	32	0.024592398	VDAC1, NUP153, SNUPN, IPO7, RANBP2, NUTF2, EIF5A, KPNA1, KPNA3, NUP107, NXT1, NUP133, XPO7, SENP2, NUP88, NUP214, DDX19B, RAE1
ORGANELLE_OUTER_MEMBRANE	23	0.025216682	VDAC1, TOMM22, GUCY2F,

			RHOT1, MTX2, BNIP3, VDAC2, GUCY2D, FIS1, DHCR7, MSTO1, MFN2, OPA1, PSEN1, VDAC3, RAE1
OUTER_MEMBRANE	24	0.025274007	VDAC1, TOMM22, GUCY2F, RHOT1, MTX2, BNIP3, VDAC2, GUCY2D, FIS1, DHCR7, MSTO1, MFN2, OPA1, PSEN1, VDAC3, RAE1
MITOCHONDRIAL_MEMBRANE_PART	51	0.025498178	COX6B2, ATP5O, NDUFS4, TIMM23, TOMM22, TIMM9, OXA1L, RHOT1, NDUFS3, ATP5G2, NDUFA1, ATP5B, NDUFAB1, NDUFS8, BNIP3, NDUFV1, FIS1, SURF1, TIMM8B, TIMM17A, TIMM50, NDUFA13, TIMM17B, TIMM13, NDUFB6, MFN2, COX18, OPA1, UQCRC1, TIMM10, NDUFA9, ATP5E, ATP5G3
ENVELOPE	158	0.025594711	COX6B2, VDAC1, NUP153, PHB2, ATP5O, MRPL19, SDHD, SNUPN, IPO7, RANBP2, WTAP, SP140, SLC25A3, NDUFS4, TIMM23, MATR3, TOMM22, MRPL32, NUTF2, MAOB, TIMM9, HTATIP2, OXA1L, TMPO, GUCY2F, HSD3B2, EIF5A, RHOT1, LEMD3, MTX2, PHB, KPNA1, NDUFS3, ATP5G2, KPNA3, NDUFA1, ATP5B, NUP107, NDUFAB1, NDUFS8, NARF, PMPCA, BNIP3, VDAC2, NDUFV1, NRM, GUCY2D, FIS1, SURF1, SLC25A22, TIMM8B, TIMM17A, TIMM50, DHCR7, NXT1, HSD3B1, MSTO1, NUP133, ABCF2, EMD, NDUFA13, NEU4, POLA1, TIMM17B, XPO7, TIMM13, NDUFB6, SENP2, MFN2, NUP88, BNIP3L, PARP1, NUP214, COX18, OPA1, UQCRC1, PSEN1, DDX19B, VDAC3, TRIM27, RAE1, HCCS
PROTEASOME_COMPLEX	22	0.026303034	PSMD14, PSMD8, PSMC2, PSMD12, PSMC6, PSME2, PSMD13, PSMD2, PSME1, PSMD1, PSMC3, PSMD11, PSMD10

NUCLEAR_ENVELOPE	67	0.027314804	NUP153, MRPL19, SNUPN, IPO7, RANBP2, WTAP, SP140, MATR3, NUTF2, HTATIP2, TMPO, GUCY2F, EIF5A, LEMD3, KPNA1, KPNA3, NUP107, NARF, BNIP3, NRM, GUCY2D, DHCR7, NXT1, NUP133, EMD, POLA1, XPO7, SENP2, NUP88, BNIP3L, PARP1, NUP214, PSEN1, DDX19B, TRIM27, RAE1
RIBOSOMAL_SUBUNIT	20	0.0274949	MRPS18C, MRPS36, MRPS24, MRPL51, MRPS35, MRPL10, MRPS11, MRPS22, MRPS21, MRPS28, MRPS18A
RIBOSOME	39	0.02831512	RPS7, APEX1, MRPS18C, MRPS36, MRPS24, MRPS10, MRPL51, SRP68, MRPS35, MRPL40, MRPL10, MRPS11, MRPS22, MRPS21, MRPS28, MRPS18A, RPL7A
SPLICEOSOME	40	0.029044444	SNRPB2, LSM4, PRPF18, SNRPG, SF3B2, SF3A3, SNRPD2, DDX23, DHX15, SNRPC, SMNDC1, SLU7, SNW1, TXNL4A, SNRPB, DHX8, SNRPD3, SF3B3
ORGANELLAR_RIBOSOME	22	0.037853356	MRPS18C, MRPS36, MRPS24, MRPS10, MRPL51, MRPS35, MRPL40, MRPL10, MRPS11, MRPS22, MRPS21, MRPS28, MRPS18A
MITOCHONDRIAL_RIBOSOME	22	0.071536504	MRPS18C, MRPS36, MRPS24, MRPS10, MRPL51, MRPS35, MRPL40, MRPL10, MRPS11, MRPS22, MRPS21, MRPS28, MRPS18A

**Tabella 9: Analisi per Cellular Component: gene set associati al fenotipo Normal.**

GENE SET NAME	SIZE	FDR	GENE IN LEADING EDGE SUBSET
RIBONUCLEOPROTEIN_COMPLEX BIOGENESIS_AND_ASSEMBLY	56	0.008633316	GEMIN5, USP39, EIF1B, EIF2A, PAIP2, PAIP1, EIF2B2, MTIF3, EIF1AX, RPS3A, DICER1, CRNKL1, SF3A3, SNRPD1, MTIF2, EIF5A, EIF2B1, SNRPD2, NOLC1, EIF2B3, GEMIN7, GEMIN6, EIF2S3, EXOSC7, EIF2B4, SLU7, DAZL, TXNL4A, EIF2AK3
RNA_SPLICINGVIA TRANSESTERIFICATION REACTIONS	21	0.008970048	GEMIN5, USP39, CRNKL1, SF3A3, SNRPD1, SNRPD2, GEMIN7, TRA2A, GEMIN6, SLU7, SNW1, TXNL4A
PROTEIN_RNA_COMPLEX_ASSEMBLY	44	0.010980261	USP39, EIF1B, EIF2A, PAIP2, PAIP1, EIF2B2, MTIF3, EIF1AX, RPS3A, DICER1, CRNKL1, SF3A3, SNRPD1, MTIF2, EIF5A, EIF2B1, SNRPD2, EIF2B3, EIF2S3, EIF2B4, SLU7, DAZL, TXNL4A, EIF2AK3
TRANSLATIONAL_INITIATION	26	0.011457299	EIF1B, PAIP2, PAIP1, EIF2B2, MTIF3, EIF1AX, RPS3A, MTIF2, EIF5A, EIF2B1, EIF2B3, EIF2S3, EIF2B4, DAZL, EIF2AK3
PROTEIN_FOLDING	55	0.045276206	SEP15, AHSA1, NFYC, CCT4, CLPX, HSP90AA1, CCT6A, PFDN4, BAG2, RP2, ARL2, ST13, STUB1, GLRX2, HSPE1, ERP29, PDIA6, MKKS, DNAJA1, FKBP9, DNAJB6, CCT6B, CCT7, APOA2, TBCE, TTC1, DNAJA3
RNA_SPLICING	66	0.067493625	GEMIN5, USP39, SNRPB2, LSM4, PRPF18, SNRPG, CRNKL1, SF3B2, SF3A3, SNRPD1, SNRPD2, NOL3, GEMIN7, LSM1, DDX23, TRA2A, GEMIN6, SNRPC, NCBP1, SMNDC1, SLU7, SNW1, TXNL4A, PRPF3, NONO, SNRPB, DHX8, SNRPD3, SF3B3, ZRANB2, PPAN, LSM6
MRNA_PROCESSING_GO_0006397	51	0.06967371	GEMIN5, USP39, PRPF18, KHDRBS1, LSM3, CRNKL1, SF3A3, SNRPD1, SNRPD2,

			SLBP, GEMIN7, LSM1, DHX15, TRA2A, GEMIN6, PABPC1, NCBP1, SMNDC1, CPSF3, SLU7, SNW1, TXNL4A, NONO
MRNA_METABOLIC_PROCESS	62	0.15869495	GEMIN5, USP39, PAIP1, PRPF18, KHDRBS1, LSM3, CRNKL1, SF3A3, SNRPD1, AUH, SNRPD2, UPF2, SLBP, GEMIN7, LSM1, DHX15, TRA2A, GEMIN6, PABPC1, NCBP1, SMNDC1, CPSF3, SLU7, SNW1, TXNL4A, NONO
DNA_CATABOLIC_PROCESS	22	0.16022184	UBE2V2, ISG20, NTHL1, ERCC1, XRN2, UBE2N, PPT1, ERCC5, BNIP3, TPD52L1, SOD1
RESPONSE_TO TEMPERATURE_STIMULUS	16	0.19627394	EIF2B2, FGF16, EIF2B1, EIF2B3, GCLC, EIF2B4, GMPR, SOD1, DNAJB4
NUCLEOBASENUCLEOSIDENUCLEOTIDE AND_NUCLEIC_ACID_TRANSPORT	28	0.20150258	CKAP5, SLC29A1, KHDRBS1, EIF5A, UPF2, DDX25, NUP107, SLC23A1, NUP133, SLC23A2, NUDT4, SMG7, DDX19B, VDAC3, RAE1, SMG1
CELLULAR_PROTEIN CATABOLIC_PROCESS	53	0.20583959	PSMD14, UBE2A, PCNP, ARIH1, RNF11, UBE2G1, TSG101, ANAPC10, STUB1, UFD1L, UBE2D1, BTRC, UBB, SMURF2, UBE3A, UBE2E1, UHRF2, ANAPC5, EDEM1, ANAPC4, FBXO7, CDC23, UBE2I, SELS, PARK2, DERL1, UBE4A
NITROGEN_COMPOUND BIOSYNTHETIC_PROCESS	25	0.2059484	HSP90AB1, OAZ1, MAT2B, HSP90AA1, PAH, GLA, GCHFR, PYCR1, BBOX1, GCH1
CASPASE_ACTIVATION	24	0.20746294	APAF1, MTCH1, IFNB1, MOAP1, LCK, STAT1, DIABLO
MICROTUBULE_BASED_MOVEMENT	16	0.21176434	KIF3B, KIF5B, RHOT1, KIF1A, LRPPRC, UXT, KIF4A, KIF13B, OPA1
RNA_EXPORT_FROM_NUCLEUS	18	0.21735795	KHDRBS1, EIF5A, UPF2,



			DDX25, NUP107, NUP133, NUDT4, SMG7, DDX19B, RAE1, SMG1
--	--	--	--

**Tabella 10:Analisi per Biological Process: gene set associati al fenotipo Normal.**

GENE SET NAME	SIZE	FDR	GENE IN LEADING EDGE SUBSET
TRANSLATION_INITIATION FACTOR_ACTIVITY	17	0.0069216536	EIF2A, EIF2B2, EIF2S2, EIF4E, EIF5A, EIF2B1, EIF2B3, COPS5, EIF2S3, EIF2B4
SINGLE_STRANDED_DNA_BINDING	32	0.01631071	HNRPDL, MLH1, MYT2, TERF2IP, WBP11, RAD51, PCBP1, MSH3, RPA4, IGHMBP2, ERCC1, HMGB2, YBX1, RPA2, ERCC5, XPC, MSH2
DOUBLE_STRANDED_DNA_BINDING	31	0.06433922	XRCC6, HNRPDL, MLH1, RAD51, NTHL1, MSH3, PNKP, APTX, HMGB2, YBX1, ERCC5, MSH2
STRUCTURE_SPECIFIC_DNA_BINDING	53	0.07594784	XRCC6, HNRPDL, MLH1, MYT2, TERF2IP, WBP11, RAD51, PCBP1, NTHL1, MSH3, RPA4, PNKP, APTX, IGHMBP2, ERCC1, HMGB2, YBX1, RPA2, ERCC5, XPC, MSH2
TRANSLATION_FACTOR_ACTIVITY NUCLEIC_ACID_BINDING	29	0.09818851	EIF2A, PAIP1, EIF2B2, EIF2S2, EIF4E, EIF5A, EIF2B1, EIF2B3, COPS5, PABPC1, EIF2S3, EIF2B4, DAZL
CARBON_OXYGEN_LYASE_ACTIVITY	31	0.10472582	APEX1, PTS, FH, NTHL1, ALAD, AUH CA6, ENO3, EHHADH, CA2, CA7, HADHA, ECHS1, CBS, CA5B, ENO1, OGG1
UNFOLDED_PROTEIN_BINDING	40	0.11131139	TOMM20, CCT4, HSPD1, CCT6A, PTGES3, RP2, PPIC, MKKS, TOR1B, TCP1, CCT6B, CCT7, AFG3L2, TTC1, DNAJA3, DNAJB4, CLGN, AIP, NPM1, CCT3, GRPEL1, HSPA2, CHAF1A, LRPAP1

TRANSLATION_REGULATOR_ACTIVITY	31	0.1301958	EIF2A, PAIP2, PAIP1, EIF2B2, EIF2S2, EIF4E, EIF5A, EIF2B1, EIF2B3, COPS5, PABPC1, EIF2S3, EIF2B4, DAZL
EXONUCLEASE_ACTIVITY	18	0.21967793	APEX1, ISG20, WRN, REXO2, APTX, XRN2, ENPP3, EXOSC7
HYDRO_LYASE_ACTIVITY	27	0.22173496	FH, ALAD, AUH, CA6, ENO3, EHHADH, CA2, CA7, HADHA, ECHS1, CBS, CA5B, ENO1
LIGASE_ACTIVITY	89	0.22195892	FBXO3, RTCD1, ARIH1, UBE2M, SUCLA2, ANAPC10, FARS2, STUB1, MARS, ACACA, TRAF6, KARS, HLCS, GSS, PCCA, DZIP3, GCLC, CAD, GMPS, FPGS, ANAPC11, SMURF2, TRAF7, BRAP, UBE2E1, NARS, PAICS, RNF14, UHRF2, ACSL1, ANAPC5, ANAPC4, FBXO7, ACSL4, QARS, CDC23, PCCB, DDB2, CPS1, ASNS, PARK2

**Tabella 11: Analisi per Molecular Function: gene set associati al fenotipo Normal.**

Le analisi per CC, BP e MF non presentano gene set significativi associati al fenotipo *Diabetic*.

### 8.2.2.3 Glucose intollerant VS Diabetic

GENE SET NAME	SIZE	FDR	GENE IN LEADING EDGE SUBSET
RIBONUCLEOPROTEIN COMPLEX	119	0.0064141904	APEX1, PTGES3, HNRPD, POP7, SRP68, NR0B1, SNRPA, SNRPD1, RBM14, SNRPC, RPL7A, PRPF18, MRPL51, SNRPD2, SRP19, RPL15, MRPL10, PSMA1, RPS7, DKC1, FAU, EFTUD2, RPS4X, RALY, PRPF8, RPL9, GEMIN4, EMG1, SNRPD3, TEP1, SF3B2, MRPL40, DHX8,

			MRPS18A, SRA1, MRPS18C, SF3A3, LSM4, DDX23, DHX15, MKRN3, PTBP1, SRP14, SF3B3, SNRPB2, SNW1, SNRPG, SLBP, RPL13A, NOC4L, RRBP1, EPM2A, POP4, SMNDC1, LSM6, U2AF1, UTP18, SNRPA1
NUCLEOLUS	104	0.050093412	ZNF330, ILF2, DDX21, POP7, SRP68, MKI67IP, RPL35, NSUN2, TSPYL2, RPS7, NPM1, DKC1, ZNF274, EXOSC10, RPL3, MORF4L2, SMUG1, GEMIN4, DDX47, EMG1, ELP3, DDX56, NCL, RPL11, NOLC1, TCOF1, CCNT1, IMP4, POLA1, WRN, NUFIP1, EXOSC9, EXOSC1, REXO4, PTBP1, IMP3, DEDD, ZNF506, COIL, OASL, CIRH1A, XRN2
RIBOSOME	39	0.058163535	APEX1, SRP68, NR0B1, RPL7A, MRPL51, RPL15, MRPL10, RPS7, FAU, RPS4X, RPL9, MRPL40, MRPS18A, MRPS18C
SMALL NUCLEAR RIBONUCLEOPROTEIN COMPLEX	19	0.0604272	SNRPD1, SNRPC, SNRPD2, PRPF8, GEMIN4, SNRPD3, LSM4, DDX23, SF3B3, SNRPB2, LSM6, SNRPA1
PROTEASOME COMPLEX	22	0.063842796	PSMD11, PSMD8, PSMD14, PSMC6, PSMD13, PSMD12, PSME2, PSMC3, PSMD2, PSME1
SPLICEOSOME	40	0.06777097	SNRPA, SNRPC, PRPF18, SNRPD2, EFTUD2, PRPF8, SNRPD3, SF3B2, DHX8, SF3A3, LSM4, DDX23, DHX15, SF3B3, SNRPB2, SNW1, SNRPG, SMNDC1, U2AF1, SNRPA1
ENDOPLASMIC_RETICULUM	258	0.07443328	PTN, APEX1, SSR2, PCMT1, SRP68, SEC61B, SPCS1, ICMT, PDIA6, PLP2, RCN2, PIGS, TXNDC5, PIGA, PSEN1, FKBP2, MBTPS1, BNIP1, PIGH, HMOX1, FATE1, PRNP, SLMAP, STX8, PIGB, P4HA1, SEP15, LRMP, TMTC2, STT3B, PIGF, IFNGR2, PIGY, ANP32A, CYP46A1, DDOST, KRTCAP2, FKBP9, STS, DAD1, PIGZ, TPST2, ERP29, CRIPAK, OGN, POMT1, TMEM98, NSDHL, TMEM50B, SOAT1, RTN4R, STAU1,

			<p>HSD3B1, NRP1, TM7SF2, NT5C3, RINT1, PIGC, SYVN1, DERL3, PLA2G2A, YIPF6, UBQLN4, BNIP3L, SLC36A1, DERL1, APOA5, HERPUD1, CAV1, HSP90B1, SEC22A, CTSZ, ATF6, KIFAP3, ANK3, MAN1B1, EIF2AK3, HAX1, MINPP1, SLC27A5, ZW10, PGAP1, TRAM1, DNAJC1, EBP, YIPF4, CAMLG, SELS, HTRA2, ORMDL1, AGPAT2, ITPR2, APOC1, RPN2, CASQ1, UGT2B28, PDIA2, PIGT, POR, AADAC, DHCR24, HMGCR, SLC9A6, ITGB1, CALU, MRV1, EDEM1, DPM1, CANX, KIF1C, NPC1, ALG3, RSAD2, TMEM50A, TOR1A, CLN6, PSEN2, DNAJB11, TRPC4AP, MAL, VWF, STT3A</p>
NUCLEAR_PART	487	0.079066135	<p>PTGES3, ZNF330, XRCC6, ILF2, INTS12, HNRPDL, DDX21, POP7, SRP68, LEMD3, RTCD1, MKI67IP, SP140, SNRPA, SNRPD1, NARF, ERCC3, RBM14, SNRPC, SAP18, TERF2IP, RPL35, H2AFY, BARD1, NSUN2, NUP107, PRPF18, EXOSC3, PSEN1, CBX1, SMARCC1, SNRPD2, NFYB, SMARCA5, BNIP1, APOBEC3F, TAF9, TSPYL2, RPA4, GTF3C2, MED8, RPS7, NPM1, DKC1, TAF7, GEMIN5, ZNF274, ERCC1, EXOSC10, HIPK2, CHAF1A, NUP62, UBE2I, POLH, EFTUD2, RPL3, RALY, PRPF8, MORF4L2, SMUG1, SMARCE1, MSH6, LRPPRC, GEMIN4, DDX47, EMG1, SNRPD3, KLHDC3, CHEK2, RELA, TEP1, SF3B2, HMGB2, DYRK1A, PPARGC1A, EPAS1, ELP3, WTAP, GTF3C3, DHX8, SPTBN4, PITX2, HDAC8, DDX56, NCL, RPL11, INTS2, NOLC1, POLR2B, TCOF1, CCNT1, IMP4, ANAPC4, HDAC11, MSH2, SF3A3, TIPIN, POLA1, LSM4, IPO7, H1FNT, WRN, FIGLA, GEMIN6, HDAC2, RUVBL2, DDX23, NUFIP1, LMO4, EXOSC9, NFYA,</p>

			NXT1, ZNF148, BNIP3L, RGS12, DHX15, EXOSC1, XPO7, ACTB, MRPL19, REXO4, RPA2, CDC23, POLR3H, BNIP2, ATF6, RANBP2, PTBP1, IMP3, BRCA1, CCNH, TAF4, NAP1L3, HSPD1, SF3B3, DEDD, ZNF506, TRIM27, SNRPB2, COIL, ERCC2, EIF5A, NAP1L1, WDHD1, SNW1, OASL, FZR1, SNRPG, CIRH1A, XRN2, TBP, NUP133, EDF1, MPG, EMD
CHROMOSOME	114	0.09627138	ZNF330, TINF2, MKI67IP, MIS12, TERF2IP, H2AFY, PSEN1, CBX1, SMARCA5, RAN, RPA4, DCTN2, BIRC5, BUB3, ERCC1, UBE2I, SMARCE1, POLG2, CENPC1, LRPPRC, PCNA, KLHDC3, TERF2, HMGB2, MCM2, HDAC8, RFC4, RFC2, JUND, TIPIN, POLA1, H1FNT, NUFIP1, RGS12, UPF1, RPA2
ENDOSOME	57	0.09741414	DNER, ARF6, ALS2, TSG101, ANKFY1, LAMP1, BET1L, CD79A, CTNS, AVPR1A, WDFY1SNX17, STS, BACE1, CTSE, AVPR1B
MICROTUBULE_ASSOCIATED COMPLEX	44	0.10340312	DCTN3, MAP2, KIF25, DCTN2, DYNLL1, DYNC1LI2, ACTR1B, KIF3B, STAU1, KIF1B, EML1, ACTR1A, KIF5B, RABGAP1, KIFAP3, PXN, ATG4A, DCX, MAP1B, PAFAH1B1, MAP1A
DNA_DIRECTED RNA POLYMERASEII HOLOENZYME	54	0.10401403	INTS12, ERCC3, TAF9, MED8, TAF7, PPARGC1A, ELP3, INTS2, POLR2B, ZNF148, TAF4, ERCC2, TBP, EDF1, TAF2, TAF10, TAF1, POLR2J, POLR2A, SHFM1, GTF2H1, ERCC5, TAF12, TAF5, INTS8, MED4, INTS5, POLR2L, INTS1
NUCLEAR_MEMBRANE_PART	39	0.10873131	LEMD3, NARF, NUP107, PSEN1, NUP62, IPO7, NXT1, XPO7, RANBP2, EIF5A, NUP133, SNUPN, LBR, NUP54, SENP2, NUP50, PSEN2, KPNB1, NUP98
NUCLEAR_CHROMOSOME	50	0.1089988	MKI67IP, TERF2IP, H2AFY, CBX1, RPA4, ERCC1, UBE2I, SMARCE1, LRPPRC, KLHDC3, HMGB2, HDAC8, TIPIN, POLA1, H1FNT,

			NUFIP1, RGS12, RPA2
CHROMOSOMAL_PART	88	0.11040027	ZNF330, TINF2, MIS12, H2AFY, PSEN1, CBX1, RAN, RPA4, DCTN2, BIRC5, BUB3, ERCC1, UBE2I, CENPC1, PCNA, KLHDC3, TERF2, MCM2, RFC4, RFC2, JUND, TIPIN, POLA1, H1FNT, NUFIP1, UPF1, RPA2
NUCLEAR_MEMBRANE	47	0.1117122	LEMD3, NARF, NUP107, PSEN1, NUP62, WTAP, IPO7, NXT1, XPO7, MRPL19, RANBP2, TRIM27, EIF5A, NUP133, SNUPN, LBR, NUP54, SENP2, NUP50, PSEN2, KPNB1, NUP98, PRICKLE1, NUTF2
UBIQUITIN_LIGASE_COMPLEX	24	0.11263588	BARD1, UBE4A, FBXO7, ARIH1, TRAF7, ANAPC4, CAND1, CDC23, BRCA1, FBXL4, FZR1
LIPID_RAFT	27	0.11386945	SPCS1, LEMD3, SP140, PLP2, PIGS, NARF, RER1, PIGA, NUP107, PSEN1, BNIP1, FIS1, PIGB, LRMP, STT3B, PIGF, NUP62, AP1S1, PIGY, DDOST, KRTCAP2, DAD1, COG2, WTAP, COPB2, NRBP1, POMT1, KIF1B, HSD3B1, POLA1, IPO7, BET1, PIGC, DERL3, NXT1, UBQLN4, BNIP3L, DERL1, HERPUD1, ICA1, CAV1, STX10, NBEA, XPO7, HSP90B1, SEC22A, MRPL19, BNIP2, ATF6, RANBP2, COPZ1, TRIM27, EIF5A, STEAP2, NUP133, EMD, PCSK7, SELS, SLC2A4, HTRA2, ORMDL1, COPB1, SNUPN, RPN2, GOLGA3, LBR, PIGT, NUP54, AADAC, TMED2, HMGCR, SLC9A6, MALL, AP1G2, MRV11, SENP2, EDEM1, DPM1, GOLGB1, BNIP3, NPC1, ALG3
ENDOMEMBRANE_SYSTEM	194	0.11859294	SPCS1, LEMD3, SP140, PLP2, PIGS, NARF, RER1, PIGA, NUP107, PSEN1, BNIP1, FIS1, PIGB, LRMP, STT3B, PIGF, NUP62, AP1S1, PIGY, DDOST, KRTCAP2, DAD1, COG2, WTAP, COPB2, NRBP1, POMT1, KIF1B, HSD3B1, POLA1, IPO7, BET1, PIGC, DERL3, NXT1, UBQLN4, BNIP3L, DERL1, HERPUD1, ICA1, CAV1, STX10, NBEA, XPO7,

			HSP90B1, SEC22A, MRPL19, BNIP2, ATF6, RANBP2, COPZ1, TRIM27, EIF5A, STEAP2, NUP133, EMD, PCSK7, SELS, SLC2A4, HTRA2, ORMDL1, COPB1, SNUPN, RPN2, GOLGA3, LBR, PIGT, NUP54, AADAC, TMED2, HMGCR, SLC9A6, MALL, AP1G2, MRV1, SENP2, EDEM1, DPM1, GOLGB1, BNIP3, NPC1, ALG3
NUCLEAR_LUMEN	325	0.11988268	ZNF330, XRCC6, ILF2, INTS12, DDX21, POP7, SRP68, RTCD1, MKI67IP, SP140, NARF, ERCC3, RBM14, SAP18, RPL35, NSUN2, CBX1, SMARCC1, NFYB, SMARCA5, TAF9, TSPYL2, GTF3C2, MED8, RPS7, NPM1, DKC1, TAF7, GEMIN5, ZNF274, EXOSC10, HIPK2, CHAF1A, UBE2I, POLH, RPL3, MORF4L2, SMUG1, SMARCE1, GEMIN4, DDX47, EMG1, CHEK2, RELA, TEP1, DYRK1A, PPARGC1A, EPAS1, ELP3, GTF3C3, SPTBN4, PITX2, HDAC8, DDX56, NCL, RPL11, INTS2, NOLC1, POLR2B, TCOF1, CCNT1, IMP4, HDAC11, POLA1, WRN, FIGLA, GEMIN6, HDAC2, RUVBL2, NUFIP1, LMO4, EXOSC9, NFYA, ZNF148, EXOSC1, ACTB, REXO4, POLR3H, ATF6, PTBP1, IMP3, TAF4, NAP1L3, DEDD, ZNF506, TRIM27, COIL, ERCC2, NAP1L1, WDHD1, OASL, CIRH1A, XRN2, TBP, EDF1, MPG, TAF2, NCOA6, HDAC3, NOC4L, TAF10, TOP2B, TAF1, POLR3C, POLR2J, PHF21A, POLR2A, SHFM1, NXF3, NOL3, U2AF1, NUP54, NR6A1, UTP18, RPAIN, DEDD2, ASH2L, FXR1, ASF1A, BNIP3, NCBP1, SUDS3, GTF2H1, ERCC5, HDAC4, CPSF3, TAF12, POLQ, RPL36, YEATS4, SMARCD3, APTX, TAF5, XPO1, PML, INTS8, MED4, SUZ12, RBM39, GIT2, NUP98, INTS5, POLR2L, ZNF259, KIN
SPINDLE_POLE	17	0.1653297	LATS1, NUMA1, KNTC1, KIF11, CETN1

MITOCHONDRIAL_RESPIRATORY_CHAIN	23	0.23195103	NDUFA2, NDUFS4, NDUFAB1, UQCRB, NDUFS1, SDHA, NDUFS2, NDUFS7, NDUFA9, COX15, NDUFS3, UQCRC1
---------------------------------	----	------------	---

**Tabella 12: Analisi per Cellular Component: in bianco i gene set up-regolati (associati al fenotipo Glucose Intollerant); in grigio i gene set down-regolati (associati al fenotipo Diabetic).**

GENE SET NAME	SIZE	FDR	GENE IN LEADING EDGE SUBSET
RIBONUCLEOPROTEIN COMPLEX BIOGENESIS_AND_ASSEMBLY	56	0.016651664	EIF2B2, SNRPD1, RPS3A, EIF2B3, EXOSC3, SNRPD2, EIF2B1, FBL, DAZL, NPM1, DKC1, GEMIN5, EIF1AX, GEMIN4, CRNKL1, EIF2B4, NOLC1, SF3A3, GEMIN6, MTIF3, EXOSC2, EIF2AK3, EIF5A, EIF4B, EIF1
TRANSLATION	154	0.02448767	SIGIRR, RPL6, EIF2B2, FARS2, RPS12, RPL22, KARS, RPS3A, EIF2B3, RPL30, RPL35, RPL7A, MRPL51, CD276, RPL28, RPL24, RPS9, RPL15, MRPL10, GHRL, EIF2B1, RPS3, RPS27, DAZL, RPS5, APBB1, EEF2K, RPL29, RPL13, IL29, EIF1AX, RPL3, RPL9, PABPC4, SAMD4A, RPL14, RPL7, IRF4, IL9, EIF2B4, RPL11, DHPS, RPL31, TSC1, RPS17, AZU1, YBX2, MTIF3, NACA, UPF1, WARS, RPL23A, HRSP12, YARS, EEF1A1, BCL10, EIF2AK3, TLR7, EIF5A, EIF4B, EIF1, GHSR, EIF2AK1
RNA_SPLICING	66	0.033303328	SNRPD1, SNRPC, PRPF18, SNRPD2, DBR1, GEMIN5, EFTUD2, DHX38, PRPF8, SNRPD3, CRNKL1, SF3B2, PPARGC1A, RNPS1, PRPF3, DHX8, ZRANB2, SF3A3, DHX16, LSM4, GEMIN6, DDX23, KHSRP, NONO, TRA2A, PPAN, PTBP1, SF3B3, SNRPB2, SRRM1, SNW1, SNRPG, SFPQ, SMNDC1, LSM6,



			NOL3, U2AF1, SNRPA1
PHOSPHOINOSITIDE METABOLIC_PROCESS	24	0.036536094	PIGS, PIGA, PIGH, PIGB, PIGF, PIGY, PIGZ, PIGC
RNA_PROCESSING	133	0.042084318	INTS12, HNRPDL, FARS2, SNRPD1, SNRPC, NSUN2, ELAVL4, PRPF18, EXOSC3, SNRPD2, FBL, DKC1, DBR1, GEMIN5, EFTUD2, DHX38, PRPF8, GEMIN4, SNRPD3, CRNKL1, PABPC4, SF3B2, PPARGC1A, RNPS1, PRPF3, PABPC1, RBM3, DHX8, INTS2, NOLC1, ZRANB2, SF3A3, DHX16, LSM4, GEMIN6, DDX23, KHSRP, NONO, NUFIP1, BICD1, TRA2A, DHX15, UPF1, PPAN, PTBP1, EXOSC2, SF3B3, KHDRBS1, SNRPB2, RBM5, SRRM1, SNW1, SNRPG, SLBP, IGF2BP3, LSM3, POP4, SFPQ, SMNDC1, DUSP11, SARS, LSM6, NOL3, U2AF1, SNRPA1
ANTI_APOPTOSIS	109	0.043893814	SEMA4D, BRAF, GDNF, SERPINB2, TRIAP1, SOCS2, BECN1, NME5, IER3, TXNDC5, PSEN1, BNIP1, NFKB1, BIRC5, NPM1, TNFRSF18, PEA15, SOCS3, RELA, RTKN, DAD1, MPO, PRKCZ, TNFAIP8, NOTCH2, TNFRSF10D, SH3GLB1, BFAR, ANXA5, SNCA, BCL2L2, AATF, BNIP3L, BCL2L1, HSP90B1, GSTP1, BNIP2, GLO1, HMGB1, CRYAB, SFRP1, TAX1BP1, BAG1, IGF1R, ARHGDI, YWHAZ, HDAC3, HSPB1, IFI6, CBX4, AVEN, RNF7, NOL3, DHCR24
PHOSPHOINOSITIDE BIOSYNTHETIC PROCESS	20	0.044698954	PIGS, PIGA, PIGH, PIGB, PIGF, PIGY, PIGZ, PIGC
RNA_SPLICINGVIA TRANSESTERIFICATION REACTIONS	21	0.047923114	SNRPD1, SNRPD2, DBR1, GEMIN5, DHX38, CRNKL1, SF3A3, GEMIN6, TRA2A

PROTEIN_AMINO_ACID LIPIDATION	23	0.048859004	PIGS, PIGA, PIGH, PIGB, PIGF, PIGY, PIGZ, PGGT1B, PIGC, FNTA, PIGT, DPM1, CHM
RNA_EXPORT FROM NUCLEUS	18	0.05774205	UPF2, NUP107, NXF5, TSC1, UPF1, KHDRBS1, EIF5A, NUP133, NUDT4
LIPOPROTEIN BIOSYNTHETIC PROCESS	25	0.069362655	PIGS, PIGA, PIGH, PIGB, PIGF, PIGY, PIGZ, PGGT1B, PIGC, FNTA, PIGT, DPM1, CHM
GLYCEROPHOSPHOLIPID BIOSYNTHETIC PROCESS	26	0.06983235	PIGS, PIGA, PIGH, PIGB, PIGF, PIGY, PIGZ, PIGC, AGPAT2, PIGT, DPM1, PEMT
NUCLEOBASENUCLEOSIDENUCLEOTIDE AND NUCLEIC_ACID_TRANSPORT	28	0.07026562	UPF2, NUP107, CKAP5, NXF5, SLC23A2, TSC1, UPF1, SLC23A1, KHDRBS1, EIF5A, NUP133, SLC35B2, NUDT4
MACROMOLECULE BIOSYNTHETIC PROCESS	286	0.07050821	TSPAN8, SIGIRR, RPL6, EIF2B2, FARS2, RPS12, RPL22, KARS, PIGS, ATF4, RPS3A, IL18, EIF2B3, RPL30, RPL35, UAP1, RPL7A, PIGA, MRPL51, CD276, RPL28, ALG2, RPL24, PIGH, RPS9, RPL15, MRPL10, GHRL, EIF2B1, RPS3, RPS27, DAZL, PIGB, RPS5, APBB1, EEF2K, RPL29, STT3B, PIGF, RPL13, IL29, EIF1AX, RPL3, RPL9, PIGY, FUT8, PABPC4, SAMD4A, RPL14, RPL7, PPARGC1A, B3GNT8, COG2, PIGZ, IRF4, IL9, CHST6, EIF2B4, POMT1, RPL11, DHPS, ST8SIA2, RPL31, GCNT1, TSC1, PGGT1B, SDF2, RPS17, AZU1, PIGC, TSPAN7, YBX2, MTIF3, NACA, UPF1, MPDU1, WARS, RPL23A, HRSP12, FUT5, YARS, EEF1A1, HS3ST3B1, FNTA, BCL10, EIF2AK3, TLR7, GALNT2, EIF5A, B3GALT2, EIF4B, EIF1

NEGATIVE_REGULATION OF APOPTOSIS	137	0.07409686	SEMA4D, BRAF, GDNF, SERPINB2, TRIAP1, SOCS2, ALB, BECN1, NME5, IER3, BARD1, TXNDC5, PSEN1, BNIP1, NFKB1, GHRL, BIRC5, NPM1, TNFRSF18, NUP62, PEA15, SOCS3, RELA, RTKN, DAD1, MPO, PRKCZ, TNFAIP8, NOTCH2, TNFRSF10D, SOD1, SH3GLB1, BFAR, ANXA5, SNCA, BCL2L2, AATF, BNIP3L, BCL2L1, HSP90B1, GSTP1, BNIP2, GLO1, HMGB1, BCL10, CRYAB, SFRP1, TAX1BP1, BAG1, IGF1R, ARHGDI, YWHAZ, HDAC3, HSPB1, IFI6, CBX4, CASP3, AVEN, RNF7, NOL3, DHCR24
PROTEIN_FOLDING	55	0.07523714	CCT4, PDIA6, HSP90AA1, ARL2, DNAJA1, CCT7, TBCA, SEP15, CCT6B, CCT3, LRPAP1, AHSA1, FKBP9, PPIA, ERP29, DNAJB6, RP2, RUVBL2, DNAJC7, ATF6
PROTEIN_RNA_COMPLEX ASSEMBLY	44	0.075487114	EIF2B2, SNRPD1, RPS3A, EIF2B3, SNRPD2, EIF2B1, DAZL, NPM1, EIF1AX, CRNKL1, EIF2B4, SF3A3, MTIF3, EIF2AK3, EIF5A, EIF4B, EIF1
NEGATIVE_REGULATION OF PROGRAMMED_CELL_DEATH	138	0.07616706	SEMA4D, BRAF, GDNF, SERPINB2, TRIAP1, SOCS2, ALB, BECN1, NME5, IER3, BARD1, TXNDC5, PSEN1, BNIP1, NFKB1, GHRL, BIRC5, NPM1, TNFRSF18, NUP62, PEA15, SOCS3, RELA, RTKN, DAD1, MPO, PRKCZ, TNFAIP8, NOTCH2, TNFRSF10D, SOD1, SH3GLB1, BFAR, ANXA5, SNCA, BCL2L2, AATF, BNIP3L, BCL2L1, HSP90B1, GSTP1, BNIP2, GLO1, HMGB1, BCL10, CRYAB, SFRP1, TAX1BP1, BAG1, IGF1R, ARHGDI, YWHAZ, HDAC3, HSPB1, IFI6, CBX4, CASP3, AVEN, RNF7, NOL3, DHCR24

POSITIVE_REGULATION OF_CELLULAR_COMPONENT ORGANIZATION_AND_BIOGENESIS	34	0.086258546	ARF6, NCK1, GHRL, DAZL, SPACA3, FAF1, PPARGC1A, AZU1, AHSG, NCK2, CDC42EP4, EIF5A, SORBS3, SLIT2, AMIGO1, ROBO2, CDC42
ONE_CARBON_COMPOUND METABOLIC_PROCESS	26	0.123925045	PCMT1, ICMT, FOS, PRMT2, MAT2B, NSUN2, DMAP1, DNMT3B, PRMT5, PRMT1

**Tabella 13: Analisi per Biological Process: gene set associati al fenotipo Glucose Intollerant.**

GENE SET NAME	SIZE	FDR	GENE IN LEADING EDGE SUBSET
RNA_BINDING	213	0.0010982659	RPS8, RPS29, ILF2, STAU2, HNRPDL, RPL6, FARS2, RPS12, RTCD1, MKI67IP, RPL22, NR0B1, SNRPD1, RPS3A, RPS20, RPL30, RPL35, BARD1, ELAVL4, RPL7A, APOBEC1, RPL28, ZFP36L1, RPL24, APOBEC3F, RPL15, RPS3, DAZAP1, PSMA1, FBL, RPS7, RPS27, DAZL, NPM1, DKC1, FAU, RPS5, RPS6, GEMIN5, RPL29, RPL13, RPL3, RALY, PRPF8, RPL9, PABPC4, TEP1, RPL14, RPLP2, RPL7, PPARGC1A, RPS13, RNPS1, PRPF3, PABPC1, RBM3, NXF5, HEXIM1, NCL, RNASEH1, STAU1, RBM4, RPL31, ZRANB2, SF3A3, CIRBP, DDX23, KHSRP, NUFIP1, RPS17, PPP1R8, PABPN1, HSP90B1, PTBP1, EXOSC2, PABPC3, SRP14, SF3B3, KHDRBS1, RBM5, FXR2, SRRM1, TLR7, EIF5A, OASL, SLBP, EIF4B
STRUCTURE_SPECIFIC_DNA_BINDING	53	0.0848704	XRCC6, HNRPDL, MLH1, MYT2, NR0B1, XPC, TERF2IP, RPA4, ERCC1, MSH6, TERF2, HMGB2, WBP11, MSH2, PNKP, RPA2, FOXC2, FUBP1, TERF1, FEN1

GTPASE_BINDING	32	0.08814762	ALS2, RANBP1, RTKN, IPO8, RAPGEF6, IPO7, GGA2, GOLGA5, PLCE1, FGD2, IPO9, RANBP2
TRANSCRIPTION_ACTIVATOR_ACTIVITY	151	0.088647865	APEX1, RXRA, CREBBP, ATF4, RBM14, PQBP1, SMARCC1, TAF9, NPM1, TAF7, SP4, RFXAP, ELF3, RREB1, SMARCE1, ZFX, MAML3, BRDT, PSMC3, MITF, NFATC1, TFDP1, CDH1, PPARGC1A, EPAS1, IRF4, BRCA2, POU2AF1, ESR2, ZNF148, E4F1, FOXC2, JMY, ATF6, MAX, GMEB1, PRRX1, TAF4, BCL10, BRD8, ERCC2, FOXH1, YY1, SNW1, CALCOCO1, EDF1, NR2C2, RNF14, NFKBIB, SP1, NCOA6, GTF2A1, TRIM24, CXXC1, NFATC2, RARA, FUBP3, SOX10, KLF6, TFAP4, ZNF423, UBE2V1, PCBD1, TBPL1
SINGLE_STRANDED_DNA_BINDING	32	0.09384113	HNRPDL, MLH1, MYT2, XPC, TERF2IP, RPA4, ERCC1, TERF2, HMGB2, WBP11, MSH2, RPA2, FUBP1
RAS_GTPASE_BINDING	23	0.09494338	ALS2, RANBP1, RTKN, IPO8, RAPGEF6, IPO7, GOLGA5, PLCE1, IPO9, RANBP2
TUBULIN_BINDING	42	0.0961993	GABARAPL1, PRNP, BIRC5, UXT, LRPPRC, MAPRE1, MARK4, KATNA1, BRCA2, GABARAPL2, KIF5B, RABGAP1, ATG4A, DCX, MAP1S, ARHGEF2, ARL8A
RNA_HELICASE_ACTIVITY	21	0.09789644	DDX21, DHX38, DHX8, DDX56, DDX1, DHX16, DDX23, DHX15, UPF1
SMALL_GTPASE_BINDING	31	0.098368794	ALS2, RANBP1, RTKN, IPO8, RAPGEF6, IPO7, GGA2, GOLGA5, PLCE1, FGD2, IPO9, RANBP2
UNFOLDED_PROTEIN_BINDING	40	0.09853781	

			PTGES3, CCT4, HSPA2, TOMM20, CCT7, NPM1, CCT6B, CCT3, CHAF1A, LRPAP1, PPIA, RP2, RUVBL2, HSPD1, PPIC, AFG3L2
STRUCTURAL_MOLECULE_ACTIVITY	226	0.10262195	MYLPF, COMP, RPS8, RPS29, DSP, MYBPC2, RPL6, RPS12, ARPC3, PLP1, PRELP, RPL22, DCTN3, RPS3A, RPS20, RPL30, RPL35, LOR, RPL7A, MRPL51, MYH6, RPL28, PGM5, IMPG2, RPL24, RPS9, MATN3, RPL15, MRPL10, RPS3, RPS7, MATN1, RPS27, ARPC5, FAU, RPS5, RPS6, RPL29, RPL13, CYLC2, RPS4X, RPL3, RPL9, RPS18, RPL14, KAL1, RPLP2, MYBPC3, RPL7, RPS13, DSPP, MYOM2, SPTBN4, EPB41, RPL11, KRT15, RPL31, ARPC4, KRT5, BFAR, ACTC1, LLGL1, RPS17, BICD1, CAV1, LAMA2, ACTB, SPTAN1, RPL23A, RPS28, OPTC, KRT38
TRANSCRIPTION_COACTIVATOR_ACTIVITY	111	0.10296446	APEX1, RXRA, CREBBP, RBM14, PQBP1, SMARCC1, TAF9, NPM1, TAF7, SP4, RFXAP, ELF3, SMARCE1, ZFX, MAML3, BRDT, PSMC3, TFDP1, PPARGC1A, EPAS1, POU2AF1, ESR2, E4F1, JMY, ATF6, MAX, GMEB1, PRRX1, TAF4, BCL10, BRD8, YY1, SNW1, CALCOCO1, EDF1, NR2C2, RNF14, NFKBIB, NCOA6, GTF2A1, TRIM24, RARA, SOX10, TFAP4, PCBD1, TBPL1
MICROTUBULE_BINDING	28	0.10335982	PRNP, BIRC5, UXT, LRPPRC, MAPRE1, MARK4, KATNA1, KIF5B, ATG4A, DCX, MAP1S, ARHGEF2
PHOSPHOTRANSFERASE_ACTIVITY_PHOSPHATE_GROUP_AS_ACCEPTOR	17	0.103953294	TJP2, NME5, DTYMK, PMVK, AK1, CASK
GLYCOSAMINOGLYCAN_BINDING	29	0.10590699	HBEGF, CHODL, TNFAIP6, IMPG2, MDK, RPL29, LRPAP1, POSTN,

			AAMP, APOA5, TNXB
HELICASE_ACTIVITY	45	0.10703216	XRCC6, DDX21, ERCC3, SMARCA5, DHX38, DHX8, DDX56, DDX1, DHX16, WRN, RUVBL2, DDX23, DHX15, UPF1
GENERAL_RNA_POLYMERASE_II_TRANSCRIPTION_FACTOR_ACTIVITY	24	0.11096814	GTF2B, TAF9, TAF7, PPARGC1A, TAF1A, GTF2F2, TAF4, TBP, GTF2A1, GTF2H1, TAF12, MED4, TCEA1
HEPARIN_BINDING	21	0.11149307	HBEGF, MDK, RPL29, LRPAP1, POSTN, AAMP, APOA5, TNXB, THBS4
TRANSLATION_REGULATOR_ACTIVITY	31	0.11455012	EIF2B2, EIF2B3, EIF2B1, DAZL, EEF2K, SAMD4A, PABPC1, EIF2B4

**Tabella 14: Analisi per Molecular Function: gene set associati al fenotipo Glucose Intollerant.**

Le analisi per BP e MF non presentano gene set significativi associati al fenotipo *Diabetic*.

### 8.3 Risultati di Analisi della Varianza

Nel confronto tra soggetti normali e intolleranti al glucosio risulta che su 54675 geni analizzati la varianza nei “normali” è maggiore che negli “intolleranti al glucosio” tranne che per 54 geni. Per tali geni risulta infatti che FDR q-value < 0.05. I geni selezionati sono annotati in *Tabella 15*:

Probe Set ID	Gene Symbol	Gene Title	go biological process term	go molecular function term	go cellular component term
1552325_at	CCDC11	coiled-coil domain containing 11	---	---	---
1557094_at	LOC100652762	uncharacterized LOC100652762	---	---	---
1562049_at	LOC100507959	uncharacterized LOC100507959	---	---	---
1562292_at	ANKRD30B	ankyrin repeat domain 30B	---	---	---
1562321_at	PDK4	pyruvate dehydrogenase kinase, isozyme 4	two-component signal transduction system (phosphorelay) ///	two-component sensor activity /// nucleotide binding /// protein kinase activity ///	mitochondrion /// mitochondrion /// mitochondrial inner membrane ///

			carbohydrate metabolic process /// glucose metabolic process /// glucose metabolic process /// pyruvate metabolic process /// protein phosphorylation /// protein phosphorylation /// protein phosphorylation /// signal transduction /// regulation of acetyl-CoA biosynthetic process from pyruvate /// phosphorylation /// peptidyl-histidine phosphorylation /// signal transduction by phosphorylation /// small molecule metabolic process	protein kinase activity /// protein histidine kinase activity /// pyruvate dehydrogenase (acetyl-transferring) kinase activity /// pyruvate dehydrogenase (acetyl-transferring) kinase activity /// ATP binding /// ATP binding /// kinase activity /// transferase activity	mitochondrial matrix /// mitochondrial matrix
1570452_at	---	---	---	---	---
201909_at	RPS4Y1	ribosomal protein S4, Y-linked 1	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay /// translation /// translation /// translational initiation /// translational elongation /// translational termination /// SRP-dependent cotranslational protein targeting to membrane /// multicellular organismal development /// gene expression /// viral reproduction /// RNA metabolic process /// mRNA metabolic process /// viral infectious cycle /// viral transcription /// cellular protein metabolic process	RNA binding /// RNA binding /// structural constituent of ribosome /// structural constituent of ribosome /// structural constituent of ribosome /// rRNA binding	intracellular /// cytosol /// ribosome /// polysome /// cytosolic small ribosomal subunit /// cytosolic small ribosomal subunit /// ribonucleoprotein complex
204409_s_at	EIF1AY	eukaryotic translation initiation factor 1A, Y-linked	translation /// translational initiation	RNA binding /// translation initiation factor activity /// protein binding	cytoplasm
204622_x_at	NR4A2	nuclear receptor subfamily 4, group A, member 2	response to hypoxia /// neuron migration /// response to amphetamine /// transcription, DNA-dependent /// transcription, DNA-dependent /// regulation of transcription, DNA-dependent /// transcription initiation from RNA polymerase II promoter ///	DNA binding /// sequence-specific DNA binding transcription factor activity /// steroid hormone receptor activity /// receptor activity /// ligand-activated sequence-specific DNA binding RNA polymerase II transcription factor activity /// ligand-activated sequence-specific DNA binding	nucleus /// nucleus /// nucleus /// nucleoplasm /// cytoplasm



			<p>response to stress  /// signal  transduction ///  nervous system  development ///  adult locomotory  behavior /// post-  embryonic  development ///  response to  inorganic substance  /// gene expression  /// gene expression  /// death /// response  to insecticide ///  central nervous  system projection  neuron  axonogenesis ///  central nervous  system neuron  differentiation ///  habenula  development ///  neuron  differentiation ///  intracellular receptor  mediated signaling  pathway ///  intracellular receptor  mediated signaling  pathway /// cellular  response to  extracellular  stimulus /// cellular  response to  extracellular  stimulus /// cellular  response to stress  /// regulation of  dopamine metabolic  process ///  dopamine  biosynthetic process  /// dopamine  metabolic process ///  neuron maturation ///  positive regulation of  catalytic activity ///  steroid hormone  mediated signaling  pathway /// negative  regulation of neuron  apoptotic process ///  regulation of  respiratory gaseous  exchange /// positive  regulation of  transcription, DNA-  dependent ///  positive regulation of  transcription from  RNA polymerase II  promoter ///  dopaminergic  neuron  differentiation ///  dopaminergic  neuron  differentiation</p>	<p>RNA polymerase II  transcription factor  activity /// protein  binding /// zinc ion  binding /// sequence-  specific DNA binding ///  metal ion binding ///  protein  heterodimerization  activity</p>	
204840_s_at	EEA1	early endosome antigen 1	<p>endocytosis ///  vesicle fusion ///  synaptic vesicle to  endosome fusion ///</p>	<p>protein binding ///  calmodulin binding ///  1-phosphatidylinositol  binding /// 1-</p>	<p>intracellular ///  membrane fraction ///  cytoplasm ///  cytoplasm ///</p>

			early endosome to late endosome transport	phosphatidylinositol binding /// zinc ion binding /// zinc ion binding /// GTP-dependent protein binding /// protein homodimerization activity /// metal ion binding	endosome /// early endosome /// early endosome /// early endosome /// cytosol /// serine-pyruvate aminotransferase complex /// membrane /// extrinsic to plasma membrane /// early endosome membrane
205029_s_at	FABP7	fatty acid binding protein 7, brain	startle response /// transport /// nervous system development /// negative regulation of cell proliferation /// cell proliferation in forebrain /// neurogenesis /// epithelial cell proliferation /// prepulse inhibition	transporter activity /// lipid binding /// lipid binding	nucleus /// cytoplasm /// cell projection /// neuronal cell body
205048_s_at	PSPH	phosphoserine phosphatase	L-serine metabolic process /// L-serine metabolic process /// L-serine biosynthetic process /// L-serine biosynthetic process /// L-serine biosynthetic process /// metabolic process /// cellular amino acid biosynthetic process /// cellular amino acid biosynthetic process /// response to mechanical stimulus /// dephosphorylation /// response to nutrient levels /// response to testosterone stimulus /// cellular nitrogen compound metabolic process /// small molecule metabolic process	magnesium ion binding /// catalytic activity /// phosphoserine phosphatase activity /// phosphoserine phosphatase activity /// phosphoserine phosphatase activity /// calcium ion binding /// hydrolase activity /// phosphatase activity /// protein homodimerization activity /// metal ion binding	cytoplasm /// cytosol /// synaptosome
205421_at	SLC22A3	solute carrier family 22 (extraneuronal monoamine transporter), member 3	transport /// ion transport /// organic cation transport /// organic cation transport /// organic cation transport /// quaternary ammonium group transport /// quaternary ammonium group transport /// monoamine transport /// dopamine transport /// regulation of appetite /// ion transmembrane transport /// histamine transport /// histamine uptake /// transmembrane transport /// transmembrane transport	transporter activity /// dopamine transmembrane transporter activity /// protein binding /// ion transmembrane transporter activity /// organic cation transmembrane transporter activity /// organic cation transmembrane transporter activity /// organic cation transmembrane transporter activity /// quaternary ammonium group transmembrane transporter activity /// quaternary ammonium group transmembrane transporter activity /// toxin transporter activity	membrane fraction /// plasma membrane /// integral to plasma membrane /// membrane /// integral to membrane

206700_s_at	KDM5D	lysine (K)-specific demethylase 5D	spermatogenesis /// chromatin modification /// histone H3-K4 demethylation /// oxidation-reduction process	DNA binding /// zinc ion binding /// oxidoreductase activity /// oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen /// oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors /// histone demethylase activity (H3-K4 specific) /// metal ion binding	intracellular /// nucleus
206717_at	MYH8	myosin, heavy chain 8, skeletal muscle, perinatal	muscle contraction /// muscle filament sliding	nucleotide binding /// motor activity /// actin binding /// calmodulin binding /// ATP binding /// structural constituent of muscle /// structural constituent of muscle	cytoplasm /// cytosol /// muscle myosin complex /// muscle myosin complex /// myofibril /// myosin filament
206891_at	ACTN3	actinin, alpha 3	muscle filament sliding /// regulation of apoptotic process /// focal adhesion assembly	actin binding /// integrin binding /// calcium ion binding /// protein binding /// structural constituent of muscle /// protein homodimerization activity	cytosol /// actin filament /// focal adhesion /// pseudopodium
207978_s_at	NR4A3	nuclear receptor subfamily 4, group A, member 3	mesoderm formation /// transcription, DNA-dependent /// regulation of transcription, DNA-dependent /// transcription initiation from RNA polymerase II promoter /// axon guidance /// gene expression /// hippocampus development /// intracellular receptor mediated signaling pathway /// adult behavior /// organ regeneration /// cellular response to stress /// inner ear morphogenesis /// negative regulation of apoptotic process /// steroid hormone mediated signaling pathway /// steroid hormone mediated signaling pathway /// response to peptide hormone stimulus /// negative regulation of neuron apoptotic process /// positive regulation of cell	RNA polymerase II core promoter proximal region sequence-specific DNA binding transcription factor activity involved in positive regulation of transcription /// DNA binding /// DNA binding /// sequence-specific DNA binding transcription factor activity /// steroid hormone receptor activity /// steroid hormone receptor activity /// receptor activity /// ligand-activated sequence-specific DNA binding RNA polymerase II transcription factor activity /// thyroid hormone receptor activity /// binding /// zinc ion binding /// sequence-specific DNA binding /// metal ion binding	nucleus /// nucleus /// nucleoplasm /// transcription factor complex

			<p>cycle /// positive regulation of transcription from RNA polymerase II promoter /// semicircular canal morphogenesis /// neuromuscular process controlling balance /// vestibular reflex /// positive regulation of leukocyte apoptotic process</p>		
209189_at	FOS	FBJ murine osteosarcoma viral oncogene homolog	<p>conditioned taste aversion /// toll-like receptor signaling pathway /// MyD88-dependent toll-like receptor signaling pathway /// MyD88-independent toll-like receptor signaling pathway /// DNA methylation /// regulation of transcription, DNA-dependent /// regulation of transcription from RNA polymerase II promoter /// transcription from RNA polymerase II promoter /// transcription from RNA polymerase II promoter /// response to stress /// inflammatory response /// transforming growth factor beta receptor signaling pathway /// nervous system development /// female pregnancy /// aging /// Toll signaling pathway /// response to cold /// response to light stimulus /// response to mechanical stimulus /// response to gravity /// response to toxin /// response to organic cyclic compound /// sleep /// cellular response to extracellular stimulus /// response to lipopolysaccharide /// response to progesterone stimulus /// cellular response to hormone stimulus /// response to cytokine stimulus /// toll-like receptor 1 signaling pathway /// toll-like receptor 2 signaling</p>	<p>DNA binding /// double-stranded DNA binding /// sequence-specific DNA binding transcription factor activity /// sequence-specific DNA binding transcription factor activity /// protein binding /// sequence-specific DNA binding /// transcription regulatory region DNA binding /// protein dimerization activity /// R-SMAD binding</p>	<p>membrane fraction /// nucleus /// nucleus /// nucleus /// nucleoplasm /// transcription factor complex /// nucleolus /// synaptosome</p>

			<p>pathway /// toll-like receptor 3 signaling pathway /// toll-like receptor 4 signaling pathway /// cellular response to reactive oxygen species /// TRIF-dependent toll-like receptor signaling pathway /// response to drug /// innate immune response /// positive regulation of transcription, DNA-dependent /// positive regulation of transcription from RNA polymerase II promoter /// regulation of sequence-specific DNA binding transcription factor activity /// stress-activated MAPK cascade /// response to corticosterone stimulus /// response to cAMP /// SMAD protein signal transduction /// cellular response to calcium ion</p>		
209258_s_at	SMC3	structural maintenance of chromosomes 3	<p>M phase of mitotic cell cycle /// mitotic prometaphase /// mitotic cell cycle /// regulation of DNA replication /// DNA repair /// response to DNA damage stimulus /// cell cycle /// mitotic spindle organization /// sister chromatid cohesion /// sister chromatid cohesion /// mitosis /// mitosis /// mitotic metaphase/anaphase transition /// meiosis /// meiosis /// signal transduction /// DNA mediated transformation /// negative regulation of DNA endoreduplication /// chromosome organization /// cell division</p>	<p>nucleotide binding /// microtubule motor activity /// protein binding /// ATP binding /// dynein binding /// protein heterodimerization activity</p>	<p>chromosome, centromeric region /// chromatin /// lateral element /// spindle pole /// basement membrane /// basement membrane /// nucleus /// nucleus /// nucleoplasm /// chromosome /// cytoplasm /// cohesin complex /// nuclear matrix /// meiotic cohesin complex /// meiotic cohesin complex</p>
214218_s_at	XIST	X (inactive)-specific transcript (non-protein coding)	---	---	---
214464_at	CDC42BP A	CDC42 binding protein kinase alpha (DMPK-like)	<p>microtubule cytoskeleton organization /// protein phosphorylation /// protein phosphorylation /// protein</p>	<p>nucleotide binding /// magnesium ion binding /// magnesium ion binding /// protein kinase activity /// protein serine/threonine kinase activity /// protein</p>	<p>intracellular /// cytoplasm /// cell-cell junction /// cell-cell junction /// cell leading edge /// cell leading edge /// actomyosin</p>

			phosphorylation /// cytoskeleton organization /// nuclear migration /// phosphorylation /// cell migration /// actomyosin structure organization /// actin cytoskeleton reorganization /// actin cytoskeleton reorganization /// intracellular signal transduction /// regulation of small GTPase mediated signal transduction	serine/threonine kinase activity /// protein serine/threonine kinase activity /// small GTPase regulator activity /// protein binding /// ATP binding /// ATP binding /// phospholipid binding /// kinase activity /// transferase activity /// transferase activity, transferring phosphorus-containing groups /// identical protein binding /// identical protein binding /// metal ion binding	
214723_x_at	ANKRD36	ankyrin repeat domain 36	---	---	---
217402_at	---	---	---	---	---
219995_s_at	ZNF750	zinc finger protein 750	transcription, DNA- dependent /// regulation of transcription, DNA- dependent /// epidermis development /// amine metabolic process /// cell differentiation /// positive regulation of transcription from RNA polymerase II promoter /// oxidation-reduction process	core promoter sequence-specific DNA binding /// RNA polymerase II core promoter proximal region sequence- specific DNA binding transcription factor activity involved in positive regulation of transcription /// copper ion binding /// primary amine oxidase activity /// zinc ion binding /// metal ion binding /// quinone binding	intracellular /// nucleus /// nucleus
220504_at	KERA	keratocan	visual perception /// response to stimulus	---	extracellular region /// proteinaceous extracellular matrix /// proteinaceous extracellular matrix
221728_x_at	XIST	X (inactive)- specific transcript (non- protein coding)	---	---	---
224588_at	XIST	X (inactive)- specific transcript (non- protein coding)	---	---	---
224589_at	XIST	X (inactive)- specific transcript (non- protein coding)	---	---	---
224590_at	XIST	X (inactive)- specific transcript (non- protein coding)	---	---	---
226189_at	ITGB8	integrin, beta 8	ganglioside metabolic process /// cell adhesion /// cell adhesion /// cell- matrix adhesion /// integrin-mediated signaling pathway /// multicellular organismal development /// placenta blood vessel development	receptor activity /// receptor binding	plasma membrane /// integrin complex /// integrin complex /// membrane /// integral to membrane
227671_at	XIST	X (inactive)- specific	---	---	---

		transcript (non-protein coding)			
227827_at	---	---	---	---	---
230030_at	HS6ST2	heparan sulfate 6-O-sulfotransferase 2	---	sulfotransferase activity /// transferase activity	membrane /// integral to membrane
231486_x_at	---	---	---	---	---
235763_at	SLC44A5	solute carrier family 44, member 5	transmembrane transport	---	plasma membrane /// membrane /// integral to membrane
235888_at	GUSBP1	glucuronidase, beta pseudogene 1	carbohydrate metabolic process	catalytic activity /// hydrolase activity, hydrolyzing O-glycosyl compounds /// cation binding	---
239519_at	---	---	---	---	---
239591_at	---	---	---	---	---
240509_s_at	GREM2	gremlin 2	BMP signaling pathway	cytokine activity	extracellular region /// extracellular region /// extracellular space
241255_at	---	---	---	---	---
241955_at	HECTD1	HECT domain containing E3 ubiquitin protein ligase 1	neural tube closure /// cellular protein modification process /// cell adhesion /// protein ubiquitination /// protein ubiquitination involved in ubiquitin-dependent protein catabolic process	ubiquitin-protein ligase activity /// ubiquitin-protein ligase activity /// protein binding /// ligase activity /// acid-amino acid ligase activity /// metal ion binding	intracellular /// nucleus /// cytoplasm
242024_at	---	---	---	---	---
244106_at	---	---	---	---	---
244181_at	---	---	---	---	---
34471_at	MYH8	myosin, heavy chain 8, skeletal muscle, perinatal	muscle contraction /// muscle filament sliding	nucleotide binding /// motor activity /// actin binding /// calmodulin binding /// ATP binding /// structural constituent of muscle /// structural constituent of muscle	cytoplasm /// cytosol /// muscle myosin complex /// muscle myosin complex /// myosin complex /// myofibril /// myosin filament
AFFX-DapX-3_at	---	---	tRNA 3'-terminal CCA addition /// transcription, DNA-dependent /// regulation of transcription, DNA-dependent /// RNA processing /// cellular protein modification process /// tRNA processing /// cellular amino acid biosynthetic process /// biosynthetic process /// lysine biosynthetic process /// lysine biosynthetic process via diaminopimelate /// methylglyoxal biosynthetic process /// diaminopimelate biosynthetic process /// RNA repair /// oxidation-reduction process	nucleotide binding /// DNA binding /// RNA binding /// catalytic activity /// biotin-[acetyl-CoA-carboxylase] ligase activity /// tRNA adenylyltransferase activity /// ATP binding /// dihydrodipicolinate reductase activity /// methylglyoxal synthase activity /// tRNA cytidyltransferase activity /// oxidoreductase activity /// transferase activity /// transferase activity, transferring glycosyl groups /// nucleotidyltransferase activity /// hydrolase activity /// lyase activity /// ligase activity /// metal ion binding /// CTP:tRNA cytidyltransferase activity /// CTP:3'-	cytoplasm /// plasma membrane /// membrane /// integral to membrane

				<p>cytidine-tRNA cytidyltransferase activity /// ATP:3'- cytidine-cytidine-tRNA adenyltransferase activity /// NADPH binding</p>	
AFFX-DapX-M_at	---	---	<p>tRNA 3'-terminal CCA addition /// transcription, DNA- dependent /// regulation of transcription, DNA- dependent /// RNA processing /// cellular protein modification process /// tRNA processing /// cellular amino acid biosynthetic process /// biosynthetic process /// lysine biosynthetic process /// lysine biosynthetic process via diaminopimelate /// methylglyoxal biosynthetic process /// diaminopimelate biosynthetic process /// RNA repair /// oxidation-reduction process</p>	<p>nucleotide binding /// DNA binding /// RNA binding /// catalytic activity /// biotin- [acetyl-CoA- carboxylase] ligase activity /// tRNA adenyltransferase activity /// ATP binding /// dihydrodipicolinate reductase activity /// methylglyoxal synthase activity /// tRNA cytidyltransferase activity /// oxidoreductase activity /// transferase activity /// transferase activity, transferring glycosyl groups /// nucleotidyltransferase activity /// hydrolase activity /// lyase activity /// ligase activity /// metal ion binding /// CTP:tRNA cytidyltransferase activity /// CTP:3'- cytidine-tRNA cytidyltransferase activity /// ATP:3'- cytidine-cytidine-tRNA adenyltransferase activity /// NADPH binding</p>	<p>cytoplasm /// plasma membrane /// membrane /// integral to membrane</p>
AFFX-r2-Bs-dap-3_at	---	---	<p>tRNA 3'-terminal CCA addition /// transcription, DNA- dependent /// regulation of transcription, DNA- dependent /// RNA processing /// cellular protein modification process /// tRNA processing /// cellular amino acid biosynthetic process /// biosynthetic process /// lysine biosynthetic process /// lysine biosynthetic process via diaminopimelate /// methylglyoxal biosynthetic process /// diaminopimelate biosynthetic process /// RNA repair /// oxidation-reduction process</p>	<p>nucleotide binding /// DNA binding /// RNA binding /// catalytic activity /// biotin- [acetyl-CoA- carboxylase] ligase activity /// tRNA adenyltransferase activity /// ATP binding /// dihydrodipicolinate reductase activity /// methylglyoxal synthase activity /// tRNA cytidyltransferase activity /// oxidoreductase activity /// transferase activity /// transferase activity, transferring glycosyl groups /// nucleotidyltransferase activity /// hydrolase activity /// lyase activity /// ligase activity /// metal ion binding /// CTP:tRNA cytidyltransferase activity /// CTP:3'- cytidine-tRNA cytidyltransferase activity /// ATP:3'- cytidine-cytidine-tRNA</p>	<p>cytoplasm /// plasma membrane /// membrane /// integral to membrane</p>



				adenylyltransferase activity /// NADPH binding	
AFFX-r2-Bs-dap-5_at	---	---	tRNA 3'-terminal CCA addition /// transcription, DNA-dependent /// regulation of transcription, DNA-dependent /// RNA processing /// cellular protein modification process /// tRNA processing /// cellular amino acid biosynthetic process /// biosynthetic process /// lysine biosynthetic process /// lysine biosynthetic process via diaminopimelate /// methylglyoxal biosynthetic process /// diaminopimelate biosynthetic process /// RNA repair /// oxidation-reduction process	nucleotide binding /// DNA binding /// RNA binding /// catalytic activity /// biotin-[acetyl-CoA-carboxylase] ligase activity /// tRNA adenylyltransferase activity /// ATP binding /// dihydrodipicolinate reductase activity /// methylglyoxal synthase activity /// tRNA cytidylyltransferase activity /// oxidoreductase activity /// transferase activity /// transferase activity, transferring glycosyl groups /// nucleotidyltransferase activity /// hydrolase activity /// lyase activity /// ligase activity /// metal ion binding /// CTP:tRNA cytidylyltransferase activity /// CTP:3'-cytidine-tRNA cytidylyltransferase activity /// ATP:3'-cytidine-cytidine-tRNA adenylyltransferase activity /// NADPH binding	cytoplasm /// plasma membrane /// membrane /// integral to membrane
AFFX-r2-Bs-dap-M_at	---	---	tRNA 3'-terminal CCA addition /// transcription, DNA-dependent /// regulation of transcription, DNA-dependent /// RNA processing /// cellular protein modification process /// tRNA processing /// cellular amino acid biosynthetic process /// biosynthetic process /// lysine biosynthetic process /// lysine biosynthetic process via diaminopimelate /// methylglyoxal biosynthetic process /// diaminopimelate biosynthetic process /// RNA repair /// oxidation-reduction process	nucleotide binding /// DNA binding /// RNA binding /// catalytic activity /// biotin-[acetyl-CoA-carboxylase] ligase activity /// tRNA adenylyltransferase activity /// ATP binding /// dihydrodipicolinate reductase activity /// methylglyoxal synthase activity /// tRNA cytidylyltransferase activity /// oxidoreductase activity /// transferase activity /// transferase activity, transferring glycosyl groups /// nucleotidyltransferase activity /// hydrolase activity /// lyase activity /// ligase activity /// metal ion binding /// CTP:tRNA cytidylyltransferase activity /// CTP:3'-cytidine-tRNA cytidylyltransferase activity /// ATP:3'-cytidine-cytidine-tRNA adenylyltransferase activity /// NADPH binding	cytoplasm /// plasma membrane /// membrane /// integral to membrane

AFFX-r2-Bs-phe-5_at	---	---	two-component signal transduction system (phosphorelay) /// GTP catabolic process /// metabolic process /// cellular amino acid biosynthetic process /// aromatic amino acid family biosynthetic process /// L-phenylalanine biosynthetic process /// phosphorylation /// signal transduction by phosphorylation /// sporulation resulting in formation of a cellular spore /// response to stimulus	two-component sensor activity /// nucleotide binding /// magnesium ion binding /// GTPase activity /// prephenate dehydratase activity /// GTP binding /// kinase activity /// amino acid binding /// transferase activity /// transferase activity, transferring phosphorus-containing groups /// lyase activity	cytoplasm
AFFX-r2-Bs-thr-3_s_at	---	---	proteolysis /// cellular amino acid metabolic process /// threonine metabolic process /// metabolic process /// cellular amino acid biosynthetic process /// branched chain family amino acid biosynthetic process /// methionine biosynthetic process /// threonine biosynthetic process /// isoleucine biosynthetic process /// phosphorylation /// oxidation-reduction process	nucleotide binding /// catalytic activity /// homoserine dehydrogenase activity /// homoserine kinase activity /// threonine synthase activity /// ATP binding /// peptidase activity /// serine-type peptidase activity /// kinase activity /// oxidoreductase activity /// amino acid binding /// transferase activity /// hydrolase activity /// lyase activity /// pyridoxal phosphate binding /// NADP binding	cytoplasm
AFFX-ThrX-3_at	---	---	proteolysis /// cellular amino acid metabolic process /// threonine metabolic process /// metabolic process /// cellular amino acid biosynthetic process /// branched chain family amino acid biosynthetic process /// methionine biosynthetic process /// threonine biosynthetic process /// isoleucine biosynthetic process /// phosphorylation /// oxidation-reduction process	nucleotide binding /// catalytic activity /// homoserine dehydrogenase activity /// homoserine kinase activity /// threonine synthase activity /// ATP binding /// peptidase activity /// serine-type peptidase activity /// kinase activity /// oxidoreductase activity /// amino acid binding /// transferase activity /// hydrolase activity /// lyase activity /// pyridoxal phosphate binding /// NADP binding	cytoplasm
AFFX-ThrX-M_at	---	---	proteolysis /// cellular amino acid metabolic process /// threonine metabolic process /// metabolic process /// cellular amino acid biosynthetic process /// branched chain family amino acid biosynthetic process /// methionine	nucleotide binding /// catalytic activity /// homoserine dehydrogenase activity /// homoserine kinase activity /// threonine synthase activity /// ATP binding /// peptidase activity /// serine-type peptidase activity /// kinase activity ///	cytoplasm

			biosynthetic process /// threonine biosynthetic process /// isoleucine biosynthetic process /// phosphorylation /// oxidation- reduction process	oxidoreductase activity /// amino acid binding /// transferase activity /// hydrolase activity lyase activity pyridoxal phosphate binding /// NADP binding	
--	--	--	---	--	--

**Tabella 15: Geni per i quali la varianza dell'espressione nei soggetti normali è minore che nei soggetti intolleranti al glucosio.**

Nel confronto tra soggetti “normali” e “diabetici” risulta che su 54675 geni analizzati la varianza nei normali è maggiore che nei diabetici tranne che per 52 geni. Per tali geni risulta infatti che FDR q-value < 0.05.

Probe Set ID	Gene Symbol	Gene Title	go biological process term	go molecular function term	go cellular component term
1553191_at	DST /// LOC100 652766	dystonin /// dystonin-like	microtubule cytoskeleton organization /// transport /// cytoskeleton organization /// cell cycle arrest /// cell adhesion /// integrin- mediated signaling pathway /// axonogenesis /// retrograde axon cargo transport /// response to wounding /// maintenance of cell polarity /// regulation of microtubule polymerization or depolymerization /// cytoplasmic microtubule organization /// hemidesmosome assembly /// intermediate filament cytoskeleton organization /// intermediate filament cytoskeleton organization /// intermediate filament cytoskeleton organization /// cell motility /// transmembrane transport	actin binding /// integrin binding /// calcium ion binding /// protein binding /// ATP binding /// microtubule binding /// protein C-terminus binding /// ATPase activity, coupled to transmembrane movement of substances /// protein homodimerization activity /// microtubule plus-end binding	basement membrane /// nucleus /// nucleus /// nuclear envelope /// cytoplasm /// cytoplasm /// cytoplasm /// endoplasmic reticulum /// endoplasmic reticulum membrane /// cytoskeleton /// microtubule /// intermediate filament /// plasma membrane /// cell cortex /// basal plasma membrane /// actin cytoskeleton /// actin cytoskeleton /// microtubule cytoskeleton /// microtubule cytoskeleton /// membrane /// integral to membrane /// cytoplasmic membrane-bounded vesicle /// Z disc /// Z disc /// cell junction /// hemidesmosome /// hemidesmosome /// hemidesmosome /// axon /// cell leading edge /// H zone /// axon part /// microtubule plus end /// cell projection /// intermediate filament cytoskeleton /// neurofilament cytoskeleton
1562049_at	LOC100 507959	uncharacterized LOC100507959	---	---	---
1570452_at	---	---	---	---	---
201909_at	RPS4Y1	ribosomal protein S4, Y- linked 1	nuclear-transcribed mRNA catabolic process, nonsense- mediated decay ///	RNA binding /// RNA binding /// structural constituent of ribosome /// structural constituent	intracellular /// cytosol /// ribosome /// polysome /// cytosolic small ribosomal

			<p>translation /// translation /// translation /// translational initiation /// translational elongation /// translational termination /// SRP-dependent cotranslational protein targeting to membrane /// multicellular organismal development /// gene expression /// viral reproduction /// RNA metabolic process /// mRNA metabolic process /// viral infectious cycle /// viral transcription /// cellular protein metabolic process</p>	<p>of ribosome /// structural constituent of ribosome /// rRNA binding</p>	<p>subunit /// cytosolic small ribosomal subunit /// ribonucleoprotein complex</p>
204388_s_at	MAOA	monoamine oxidase A	<p>cellular biogenic amine metabolic process /// catecholamine metabolic process /// xenobiotic metabolic process /// synaptic transmission /// neurotransmitter secretion /// behavior /// neurotransmitter catabolic process /// neurotransmitter biosynthetic process /// dopamine catabolic process /// serotonin metabolic process /// phenylethylamine metabolic process /// small molecule metabolic process /// oxidation-reduction process</p>	<p>primary amine oxidase activity /// primary amine oxidase activity /// oxidoreductase activity /// flavin adenine dinucleotide binding /// serotonin binding</p>	<p>soluble fraction /// mitochondrion /// mitochondrial outer membrane /// mitochondrial outer membrane /// membrane /// integral to membrane</p>
204409_s_at	EIF1AY	eukaryotic translation initiation factor 1A, Y-linked	<p>translation /// translational initiation</p>	<p>RNA binding /// translation initiation factor activity /// protein binding</p>	<p>cytoplasm</p>
204607_at	HMGCS2	3-hydroxy-3-methylglutaryl-CoA synthase 2 (mitochondrial)	<p>kidney development /// liver development /// steroid biosynthetic process /// cholesterol biosynthetic process /// brain development /// midgut development /// response to nutrient /// metabolic process /// isoprenoid biosynthetic process /// lipid biosynthetic process /// response to temperature stimulus /// response to bacterium /// response to metal</p>	<p>catalytic activity /// hydroxymethylglutaryl-CoA synthase activity /// transferase activity</p>	<p>mitochondrion /// mitochondrion /// mitochondrial inner membrane /// mitochondrial matrix /// mitochondrial matrix</p>

			<p>ion /// response to organic cyclic compound /// sterol biosynthetic process /// lung development /// response to insulin stimulus /// cellular response to insulin stimulus /// cellular response to hormone stimulus /// multicellular organismal response to stress /// response to testosterone stimulus /// response to glucagon stimulus /// response to triglyceride /// response to monosaccharide stimulus /// response to prostaglandin F stimulus /// response to drug /// response to starvation /// response to peptide hormone stimulus /// cellular lipid metabolic process /// small molecule metabolic process /// response to ethanol /// cellular ketone body metabolic process /// ketone body biosynthetic process /// ketone body biosynthetic process /// response to glucocorticoid stimulus /// response to cAMP /// response to growth hormone stimulus /// adipose tissue development /// response to fatty acid /// response to linoleic acid /// cellular response to lipopolysaccharide /// cellular response to amino acid stimulus /// cellular response to glucocorticoid stimulus /// cellular response to fatty acid /// cellular response to organic cyclic compound</p>		
204621_s_at	NR4A2	nuclear receptor subfamily 4, group A, member 2	<p>response to hypoxia /// neuron migration /// response to amphetamine /// transcription, DNA-dependent /// transcription, DNA-dependent /// regulation of transcription, DNA-dependent ///</p>	<p>DNA binding /// sequence-specific DNA binding transcription factor activity /// steroid hormone receptor activity /// receptor activity /// ligand-activated sequence-specific DNA binding RNA polymerase II</p>	<p>nucleus /// nucleus /// nucleus /// nucleoplasm /// cytoplasm</p>

			<p>transcription initiation from RNA polymerase II promoter /// response to stress /// signal transduction /// nervous system development /// adult locomotory behavior /// post-embryonic development /// response to inorganic substance /// gene expression /// gene expression /// death /// response to insecticide /// central nervous system projection neuron axonogenesis /// central nervous system neuron differentiation /// habenula development /// neuron differentiation /// intracellular receptor mediated signaling pathway /// intracellular receptor mediated signaling pathway /// cellular response to extracellular stimulus /// cellular response to extracellular stimulus /// cellular response to stress /// regulation of dopamine metabolic process /// dopamine biosynthetic process /// dopamine metabolic process /// neuron maturation /// positive regulation of catalytic activity /// steroid hormone mediated signaling pathway /// negative regulation of neuron apoptotic process /// regulation of respiratory gaseous exchange /// positive regulation of transcription, DNA-dependent /// positive regulation of transcription from RNA polymerase II promoter /// dopaminergic neuron differentiation /// dopaminergic neuron differentiation</p>	<p>transcription factor activity /// ligand-activated sequence-specific DNA binding RNA polymerase II transcription factor activity /// protein binding /// zinc ion binding /// sequence-specific DNA binding /// metal ion binding /// protein heterodimerization activity</p>	
--	--	--	---	--	--

205029_s_at	FABP7	fatty acid binding protein 7, brain	startle response /// transport /// nervous system development /// negative regulation of cell proliferation /// cell proliferation in forebrain /// neurogenesis /// epithelial cell proliferation /// prepulse inhibition	transporter activity /// lipid binding /// lipid binding	nucleus /// cytoplasm /// cell projection /// neuronal cell body
205048_s_at	PSPH	phosphoserine phosphatase	L-serine metabolic process /// L-serine metabolic process /// L-serine biosynthetic process /// L-serine biosynthetic process /// L-serine biosynthetic process /// metabolic process /// cellular amino acid biosynthetic process /// cellular amino acid biosynthetic process /// response to mechanical stimulus /// dephosphorylation /// response to nutrient levels /// response to testosterone stimulus /// cellular nitrogen compound metabolic process /// small molecule metabolic process	magnesium ion binding /// catalytic activity /// phosphoserine phosphatase activity /// phosphoserine phosphatase activity /// phosphoserine phosphatase activity /// calcium ion binding /// hydrolase activity /// phosphatase activity /// protein homodimerization activity /// metal ion binding	cytoplasm /// cytosol /// synaptosome
206700_s_at	KDM5D	lysine (K)-specific demethylase 5D	spermatogenesis /// chromatin modification /// histone H3-K4 demethylation /// oxidation-reduction process	DNA binding /// zinc ion binding /// oxidoreductase activity /// oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen /// oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors /// histone demethylase activity (H3-K4 specific) /// metal ion binding	intracellular /// nucleus
207175_at	ADIPOQ	adiponectin, C1Q and collagen domain containing	response to hypoxia /// positive regulation of protein phosphorylation /// glucose metabolic process /// glucose metabolic process /// generation of precursor metabolites and energy /// fatty acid	receptor binding /// receptor binding /// cytokine activity /// hormone activity /// hormone activity /// protein binding /// sialic acid binding /// identical protein binding /// protein homodimerization activity /// eukaryotic	extracellular region /// extracellular region /// collagen /// extracellular space /// extracellular space /// endoplasmic reticulum /// endoplasmic reticulum

			<p> beta-oxidation ///  fatty acid beta-oxidation ///  response to nutrient ///  response to sucrose stimulus ///  response to glucose stimulus ///  response to glucose stimulus ///  positive regulation of signal transduction ///  positive regulation of signal transduction ///  negative regulation of platelet-derived growth factor receptor signaling pathway ///  positive regulation of protein kinase A signaling cascade ///  negative regulation of macrophage derived foam cell differentiation ///  negative regulation of tumor necrosis factor-mediated signaling pathway ///  positive regulation of cholesterol efflux ///  regulation of glucose metabolic process ///  negative regulation of smooth muscle cell migration ///  fatty acid oxidation ///  fatty acid oxidation ///  negative regulation of cell migration ///  negative regulation of granulocyte differentiation ///  negative regulation of protein autophosphorylation ///  positive regulation of cellular protein metabolic process ///  negative regulation of tumor necrosis factor production ///  positive regulation of interleukin-8 production ///  cellular response to insulin stimulus ///  cellular response to insulin stimulus ///  positive regulation of myeloid cell apoptotic process ///  adiponectin-mediated signaling pathway ///  negative regulation of heterotypic cell-cell adhesion ///  low-density lipoprotein particle clearance ///  response to tumor </p>	<p>cell surface binding</p>	
--	--	--	---	-----------------------------	--



			<p>necrosis factor ///  cellular response to  drug /// response to  drug /// glucose  homeostasis ///  glucose  homeostasis ///  positive regulation of  I-kappaB kinase/NF-  kappaB cascade ///  positive regulation of  I-kappaB kinase/NF-  kappaB cascade ///  negative regulation  of I-kappaB  kinase/NF-kappaB  cascade /// negative  regulation of MAP  kinase activity ///  response to ethanol  /// negative  regulation of fat cell  differentiation ///  negative regulation  of macrophage  differentiation ///  negative regulation  of low-density  lipoprotein particle  receptor biosynthetic  process /// negative  regulation of  gluconeogenesis ///  negative regulation  of gluconeogenesis  /// negative  regulation of blood  pressure /// positive  regulation of blood  pressure /// positive  regulation of protein  kinase activity ///  negative regulation  of transcription,  DNA-dependent ///  positive regulation of  fatty acid metabolic  process /// positive  regulation of fatty  acid metabolic  process /// positive  regulation of glucose  import /// positive  regulation of glucose  import /// negative  regulation of  hormone secretion  /// negative  regulation of smooth  muscle cell  proliferation ///  negative regulation  of inflammatory  response ///  negative regulation  of inflammatory  response /// positive  regulation of  peptidyl-tyrosine  phosphorylation ///  negative regulation  of phagocytosis ///  negative regulation  of synaptic</p>	
--	--	--	---	--

			<p>transmission /// brown fat cell differentiation /// brown fat cell differentiation /// protein homooligomerization /// protein homooligomerization /// response to glucocorticoid stimulus /// membrane depolarization /// membrane hyperpolarization /// protein heterotrimerization /// negative regulation of ERK1 and ERK2 cascade /// detection of oxidative stress /// positive regulation of monocyte chemotactic protein- 1 production /// protein localization in plasma membrane /// negative regulation of intracellular protein transport /// negative regulation of DNA biosynthetic process /// negative regulation of eukaryotic cell surface binding /// positive regulation of glycogen (starch) synthase activity /// positive regulation of glycogen (starch) synthase activity /// positive regulation of metanephric glomerular visceral epithelial cell development /// positive regulation of cAMP-dependent protein kinase activity /// positive regulation of renal albumin absorption /// negative regulation of platelet-derived growth factor receptor-alpha signaling pathway /// negative regulation of metanephric mesenchymal cell migration</p>		
207709_at	PRKAA2	protein kinase, AMP-activated, alpha 2 catalytic subunit	<p>energy reserve metabolic process /// transcription, DNA- dependent /// regulation of transcription, DNA- dependent /// protein phosphorylation ///</p>	<p>nucleotide binding /// chromatin binding /// chromatin binding /// protein kinase activity /// protein kinase activity /// protein serine/threonine kinase activity /// protein</p>	<p>nucleus /// nucleoplasm /// nucleoplasm /// cytoplasm /// cytosol /// apical plasma membrane /// AMP- activated protein kinase complex ///</p>

			<p>protein phosphorylation /// fatty acid biosynthetic process /// steroid biosynthetic process /// cholesterol biosynthetic process /// carnitine shuttle /// autophagy /// response to stress /// response to stress /// cell cycle arrest /// signal transduction /// insulin receptor signaling pathway /// lipid biosynthetic process /// lipid biosynthetic process /// positive regulation of autophagy /// positive regulation of autophagy /// response to activity /// Wnt receptor signaling pathway /// sterol biosynthetic process /// phosphorylation /// chromatin modification /// regulation of lipid metabolic process /// response to caffeine /// cellular response to nutrient levels /// cellular response to nutrient levels /// negative regulation of TOR signaling cascade /// negative regulation of TOR signaling cascade /// histone-serine phosphorylation /// histone-serine phosphorylation /// cellular response to glucose starvation /// cellular response to glucose starvation /// regulation of fatty acid biosynthetic process /// glucose homeostasis /// glucose homeostasis /// regulation of circadian rhythm /// regulation of circadian rhythm /// negative regulation of apoptotic process /// negative regulation of apoptotic process /// cellular lipid metabolic process /// small molecule metabolic process /// positive regulation of glycolysis /// regulation of fatty</p>	<p>serine/threonine kinase activity /// AMP-activated protein kinase activity /// AMP-activated protein kinase activity /// protein binding /// ATP binding /// kinase activity /// transferase activity /// transferase activity, transferring phosphorus-containing groups /// protein binding, bridging /// histone serine kinase activity /// histone serine kinase activity /// metal ion binding</p>	<p>protein complex</p>
--	--	--	---	--	------------------------

			<p>acid oxidation /// rhythmic process /// protein heterooligomerization /// fatty acid homeostasis /// cellular response to organic nitrogen /// regulation of energy homeostasis</p>		
209291_at	ID4	inhibitor of DNA binding 4, dominant negative helix-loop-helix protein	<p>G1/S transition of mitotic cell cycle /// negative regulation of transcription from RNA polymerase II promoter /// regulation of transcription from RNA polymerase II promoter /// neuroblast proliferation /// brain development /// cell proliferation /// positive regulation of cell proliferation /// hippocampus development /// cerebral cortex neuron differentiation /// central nervous system myelination /// cellular protein localization /// fat cell differentiation /// negative regulation of fat cell differentiation /// negative regulation of neuron differentiation /// positive regulation of osteoblast differentiation /// negative regulation of transcription, DNA-dependent /// negative regulation of astrocyte differentiation /// negative regulation of oligodendrocyte differentiation</p>	RNA polymerase II transcription factor binding /// transcription corepressor activity /// protein binding	nucleus /// nucleus /// cytoplasm
210479_s_at	RORA	RAR-related orphan receptor A	<p>transcription, DNA-dependent /// regulation of transcription, DNA-dependent /// transcription initiation from RNA polymerase II promoter /// nitric oxide biosynthetic process /// signal transduction /// gene expression /// cerebellar Purkinje cell differentiation /// intracellular receptor mediated signaling pathway /// intracellular receptor mediated signaling</p>	<p>DNA binding /// sequence-specific DNA binding transcription factor activity /// sequence-specific DNA binding transcription factor activity /// steroid hormone receptor activity /// receptor activity /// ligand-activated sequence-specific DNA binding RNA polymerase II transcription factor activity /// ligand-activated sequence-specific DNA binding RNA polymerase II transcription factor</p>	nucleus /// nucleoplasm

			<p>pathway /// regulation of macrophage activation /// steroid hormone mediated signaling pathway /// positive regulation of transcription from RNA polymerase II promoter /// positive regulation of transcription from RNA polymerase II promoter /// cGMP metabolic process</p>	<p>activity /// protein binding /// zinc ion binding /// sequence-specific DNA binding /// metal ion binding</p>	
212592_at	IGJ	immunoglobulin J polypeptide, linker protein for immunoglobulin alpha and mu polypeptides	immune response	antigen binding	extracellular region /// extracellular region
214218_s_at	XIST	X (inactive)-specific transcript (non-protein coding)	---	---	---
216248_s_at	NR4A2	nuclear receptor subfamily 4, group A, member 2	<p>response to hypoxia /// neuron migration /// response to amphetamine /// transcription, DNA-dependent /// transcription, DNA-dependent /// regulation of transcription, DNA-dependent /// transcription initiation from RNA polymerase II promoter /// response to stress /// signal transduction /// nervous system development /// adult locomotory behavior /// post-embryonic development /// response to inorganic substance /// gene expression /// gene expression /// death /// response to insecticide /// central nervous system projection neuron axonogenesis /// central nervous system neuron differentiation /// habenula development /// neuron differentiation /// intracellular receptor mediated signaling pathway /// intracellular receptor mediated signaling pathway /// cellular response to extracellular</p>	<p>DNA binding /// sequence-specific DNA binding transcription factor activity /// steroid hormone receptor activity /// receptor activity /// ligand-activated sequence-specific DNA binding RNA polymerase II transcription factor activity /// ligand-activated sequence-specific DNA binding RNA polymerase II transcription factor activity /// protein binding /// zinc ion binding /// sequence-specific DNA binding /// metal ion binding /// protein heterodimerization activity</p>	nucleus /// nucleus /// nucleus /// nucleoplasm /// cytoplasm

			stimulus /// cellular response to extracellular stimulus /// cellular response to stress /// regulation of dopamine metabolic process /// dopamine biosynthetic process /// dopamine metabolic process /// neuron maturation /// positive regulation of catalytic activity /// steroid hormone mediated signaling pathway /// negative regulation of neuron apoptotic process /// regulation of respiratory gaseous exchange /// positive regulation of transcription, DNA-dependent /// positive regulation of transcription from RNA polymerase II promoter /// dopaminergic neuron differentiation /// dopaminergic neuron differentiation		
216550_x_at	ANKRD1 2	ankyrin repeat domain 12	---	---	nucleus /// nucleus /// nucleolus /// cytoplasm /// ribosome
219759_at	ERAP2	endoplasmic reticulum aminopeptidase 2	antigen processing and presentation of peptide antigen via MHC class I /// proteolysis /// regulation of blood pressure /// antigen processing and presentation of endogenous peptide antigen via MHC class I	aminopeptidase activity /// aminopeptidase activity /// peptidase activity /// metallopeptidase activity /// metallopeptidase activity /// zinc ion binding /// zinc ion binding /// hydrolase activity /// metal ion binding	endoplasmic reticulum /// endoplasmic reticulum lumen /// endoplasmic reticulum membrane /// membrane /// integral to membrane
221728_x_at	XIST	X (inactive)-specific transcript (non-protein coding)	---	---	---
224588_at	XIST	X (inactive)-specific transcript (non-protein coding)	---	---	---
224589_at	XIST	X (inactive)-specific transcript (non-protein coding)	---	---	---
224590_at	XIST	X (inactive)-specific transcript (non-protein coding)	---	---	---
227671_at	XIST	X (inactive)-specific transcript (non-protein coding)	---	---	---
227952_at	---	---	---	---	---

230577_at	LOC100507008	uncharacterized LOC100507008	---	---	---
231486_x_at	---	---	---	---	---
232636_at	SLITRK4	SLIT and NTRK-like family, member 4	---	---	membrane /// integral to membrane
235763_at	SLC44A5	solute carrier family 44, member 5	transmembrane transport	---	plasma membrane /// membrane /// integral to membrane
236439_at	---	---	---	---	---
236962_at	---	---	---	---	---
238355_at	---	---	---	---	---
239591_at	---	---	---	---	---
241683_at	HECTD1	HECT domain containing E3 ubiquitin protein ligase 1	neural tube closure /// cellular protein modification process /// cell adhesion /// protein ubiquitination /// protein ubiquitination involved in ubiquitin-dependent protein catabolic process	ubiquitin-protein ligase activity /// ubiquitin-protein ligase activity /// protein binding /// ligase activity /// acid-amino acid ligase activity /// metal ion binding	intracellular /// nucleus /// cytoplasm
241685_x_at	---	---	---	---	---
244230_at	---	---	---	---	---
244503_at	---	---	---	---	---
34471_at	MYH8	myosin, heavy chain 8, skeletal muscle, perinatal	muscle contraction /// muscle filament sliding	nucleotide binding /// motor activity /// actin binding /// calmodulin binding /// ATP binding /// structural constituent of muscle /// structural constituent of muscle	cytoplasm /// cytosol /// muscle myosin complex /// muscle myosin complex /// myosin complex /// myofibril /// myosin filament
36711_at	MAFF	v-maf musculoaponeurotic fibrosarcoma oncogene homolog F (avian)	transcription, DNA-dependent /// regulation of transcription, DNA-dependent /// transcription from RNA polymerase II promoter /// response to stress /// parturition /// blood coagulation	DNA binding /// sequence-specific DNA binding transcription factor activity /// sequence-specific DNA binding	nucleus /// nucleoplasm
AFFX-DapX-3_at	---	---	tRNA 3'-terminal CCA addition /// transcription, DNA-dependent /// regulation of transcription, DNA-dependent /// RNA processing /// cellular protein modification process /// tRNA processing /// cellular amino acid biosynthetic process /// biosynthetic process /// lysine biosynthetic process /// lysine biosynthetic process via diaminopimelate /// methylglyoxal biosynthetic process /// diaminopimelate biosynthetic process /// RNA repair /// oxidation-reduction process	nucleotide binding /// DNA binding /// RNA binding /// catalytic activity /// biotin-[acetyl-CoA-carboxylase] ligase activity /// tRNA adenylyltransferase activity /// ATP binding /// dihydrodipicolinate reductase activity /// methylglyoxal synthase activity /// tRNA cytidyltransferase activity /// oxidoreductase activity /// transferase activity /// transferase activity, transferring glycosyl groups /// nucleotidyltransferase activity /// hydrolase activity /// lyase activity /// ligase activity /// metal ion binding /// CTP:tRNA	cytoplasm /// plasma membrane /// membrane /// integral to membrane

				<p>cytidyltransferase activity /// CTP:3'-cytidine-tRNA  cytidyltransferase activity /// ATP:3'-cytidine-cytidine-tRNA  adenyltransferase activity /// NADPH binding</p>	
AFFX-DapX-5_at	---	---	<p>tRNA 3'-terminal CCA addition ///  transcription, DNA-dependent ///  regulation of transcription, DNA-dependent ///  RNA processing ///  cellular protein modification process ///  tRNA processing ///  cellular amino acid biosynthetic process ///  biosynthetic process ///  lysine biosynthetic process ///  lysine biosynthetic process via diaminopimelate ///  methylglyoxal biosynthetic process ///  diaminopimelate biosynthetic process ///  RNA repair ///  oxidation-reduction process</p>	<p>nucleotide binding ///  DNA binding ///  RNA binding ///  catalytic activity ///  biotin-[acetyl-CoA-carboxylase] ligase activity ///  tRNA adenyltransferase activity ///  ATP binding ///  dihydrodipicolinate reductase activity ///  methylglyoxal synthase activity ///  tRNA cytidyltransferase activity ///  oxidoreductase activity ///  transferase activity ///  transferase activity, transferring glycosyl groups ///  nucleotidyltransferase activity ///  hydrolase activity ///  lyase activity ///  ligase activity ///  metal ion binding ///  CTP:tRNA  cytidyltransferase activity ///  CTP:3'-cytidine-tRNA  cytidyltransferase activity ///  ATP:3'-cytidine-cytidine-tRNA  adenyltransferase activity ///  NADPH binding</p>	<p>cytoplasm ///  plasma membrane ///  integral to membrane</p>
AFFX-r2-Bs-dap-3_at	---	---	<p>tRNA 3'-terminal CCA addition ///  transcription, DNA-dependent ///  regulation of transcription, DNA-dependent ///  RNA processing ///  cellular protein modification process ///  tRNA processing ///  cellular amino acid biosynthetic process ///  biosynthetic process ///  lysine biosynthetic process ///  lysine biosynthetic process via diaminopimelate ///  methylglyoxal biosynthetic process ///  diaminopimelate biosynthetic process ///  RNA repair ///  oxidation-reduction process</p>	<p>nucleotide binding ///  DNA binding ///  RNA binding ///  catalytic activity ///  biotin-[acetyl-CoA-carboxylase] ligase activity ///  tRNA adenyltransferase activity ///  ATP binding ///  dihydrodipicolinate reductase activity ///  methylglyoxal synthase activity ///  tRNA cytidyltransferase activity ///  oxidoreductase activity ///  transferase activity ///  transferase activity, transferring glycosyl groups ///  nucleotidyltransferase activity ///  hydrolase activity ///  lyase activity ///  ligase activity ///  metal ion binding ///  CTP:tRNA  cytidyltransferase activity ///  CTP:3'-cytidine-tRNA  cytidyltransferase</p>	<p>cytoplasm ///  plasma membrane ///  integral to membrane</p>



				activity /// ATP:3'-cytidine-cytidine-tRNA adenylyltransferase activity /// NADPH binding	
AFFX-r2-Bs-dap-5_at	---	---	tRNA 3'-terminal CCA addition /// transcription, DNA-dependent /// regulation of transcription, DNA-dependent /// RNA processing /// cellular protein modification process /// tRNA processing /// cellular amino acid biosynthetic process /// biosynthetic process /// lysine biosynthetic process /// lysine biosynthetic process via diaminopimelate /// methylglyoxal biosynthetic process /// diaminopimelate biosynthetic process /// RNA repair /// oxidation-reduction process	nucleotide binding /// DNA binding /// RNA binding /// catalytic activity /// biotin-[acetyl-CoA-carboxylase] ligase activity /// tRNA adenylyltransferase activity /// ATP binding /// dihydrodipicolinate reductase activity /// methylglyoxal synthase activity /// tRNA cytidyltransferase activity /// oxidoreductase activity /// transferase activity /// transferase activity, transferring glycosyl groups /// nucleotidyltransferase activity /// hydrolase activity /// lyase activity /// ligase activity /// metal ion binding /// CTP:tRNA cytidyltransferase activity /// CTP:3'-cytidine-tRNA cytidyltransferase activity /// ATP:3'-cytidine-cytidine-tRNA adenylyltransferase activity /// NADPH binding	cytoplasm /// plasma membrane /// membrane /// integral to membrane
AFFX-r2-Bs-dap-M_at	---	---	tRNA 3'-terminal CCA addition /// transcription, DNA-dependent /// regulation of transcription, DNA-dependent /// RNA processing /// cellular protein modification process /// tRNA processing /// cellular amino acid biosynthetic process /// biosynthetic process /// lysine biosynthetic process /// lysine biosynthetic process via diaminopimelate /// methylglyoxal biosynthetic process /// diaminopimelate biosynthetic process /// RNA repair /// oxidation-reduction process	nucleotide binding /// DNA binding /// RNA binding /// catalytic activity /// biotin-[acetyl-CoA-carboxylase] ligase activity /// tRNA adenylyltransferase activity /// ATP binding /// dihydrodipicolinate reductase activity /// methylglyoxal synthase activity /// tRNA cytidyltransferase activity /// oxidoreductase activity /// transferase activity /// transferase activity, transferring glycosyl groups /// nucleotidyltransferase activity /// hydrolase activity /// lyase activity /// ligase activity /// metal ion binding /// CTP:tRNA cytidyltransferase activity /// CTP:3'-cytidine-tRNA cytidyltransferase activity /// ATP:3'-cytidine-cytidine-tRNA adenylyltransferase activity /// NADPH	cytoplasm /// plasma membrane /// membrane /// integral to membrane

				binding	
AFFX-r2-Bs-lys-5_at	---	---	cellular amino acid biosynthetic process /// lysine biosynthetic process /// lysine biosynthetic process via diaminopimelate	catalytic activity /// diaminopimelate decarboxylase activity /// lyase activity /// carboxy-lyase activity	---
AFFX-r2-Bs-lys-M_at	---	---	cellular amino acid biosynthetic process /// lysine biosynthetic process /// lysine biosynthetic process via diaminopimelate	catalytic activity /// diaminopimelate decarboxylase activity /// lyase activity /// carboxy-lyase activity	---
AFFX-r2-Bs-phe-5_at	---	---	two-component signal transduction system (phosphorelay) /// GTP catabolic process /// metabolic process /// cellular amino acid biosynthetic process /// aromatic amino acid family biosynthetic process /// L-phenylalanine biosynthetic process /// phosphorylation /// signal transduction by phosphorylation /// sporulation resulting in formation of a cellular spore /// response to stimulus	two-component sensor activity /// nucleotide binding /// magnesium ion binding /// GTPase activity /// prephenate dehydratase activity /// GTP binding /// kinase activity /// amino acid binding /// transferase activity /// transferase activity, transferring phosphorus-containing groups /// lyase activity	cytoplasm
AFFX-r2-Bs-thr-5_s_at	---	---	proteolysis /// cellular amino acid metabolic process /// threonine metabolic process /// metabolic process /// cellular amino acid biosynthetic process /// branched chain family amino acid biosynthetic process /// methionine biosynthetic process /// threonine biosynthetic process /// isoleucine biosynthetic process /// phosphorylation /// oxidation-reduction process	nucleotide binding /// catalytic activity /// homoserine dehydrogenase activity /// homoserine kinase activity /// threonine synthase activity /// ATP binding /// peptidase activity /// serine-type peptidase activity /// kinase activity /// oxidoreductase activity /// amino acid binding /// transferase activity /// hydrolase activity /// lyase activity /// pyridoxal phosphate binding /// NADP binding	cytoplasm
AFFX-ThrX-3_at	---	---	proteolysis /// cellular amino acid metabolic process /// threonine metabolic process /// metabolic process /// cellular amino acid biosynthetic process /// branched chain family amino acid biosynthetic process /// methionine biosynthetic process /// threonine biosynthetic process /// isoleucine biosynthetic process /// phosphorylation	nucleotide binding /// catalytic activity /// homoserine dehydrogenase activity /// homoserine kinase activity /// threonine synthase activity /// ATP binding /// peptidase activity /// serine-type peptidase activity /// kinase activity /// oxidoreductase activity /// amino acid binding /// transferase activity /// hydrolase activity /// lyase activity /// pyridoxal phosphate	cytoplasm

			/// oxidation-reduction process	binding /// NADP binding	
AFFX-ThrX-5_at	---	---	proteolysis /// cellular amino acid metabolic process /// threonine metabolic process /// metabolic process /// cellular amino acid biosynthetic process /// branched chain family amino acid biosynthetic process /// methionine biosynthetic process /// threonine biosynthetic process /// isoleucine biosynthetic process /// phosphorylation /// oxidation-reduction process	nucleotide binding /// catalytic activity /// homoserine dehydrogenase activity /// homoserine kinase activity /// threonine synthase activity /// ATP binding /// peptidase activity /// serine-type peptidase activity /// kinase activity /// oxidoreductase activity /// amino acid binding /// transferase activity /// hydrolase activity /// lyase activity /// pyridoxal phosphate binding /// NADP binding	cytoplasm
AFFX-ThrX-M_at	---	---	proteolysis /// cellular amino acid metabolic process /// threonine metabolic process /// metabolic process /// cellular amino acid biosynthetic process /// branched chain family amino acid biosynthetic process /// methionine biosynthetic process /// threonine biosynthetic process /// isoleucine biosynthetic process /// phosphorylation /// oxidation-reduction process	nucleotide binding /// catalytic activity /// homoserine dehydrogenase activity /// homoserine kinase activity /// threonine synthase activity /// ATP binding /// peptidase activity /// serine-type peptidase activity /// kinase activity /// oxidoreductase activity /// amino acid binding /// transferase activity /// hydrolase activity /// lyase activity /// pyridoxal phosphate binding /// NADP binding	cytoplasm

**Tabella 16: Geni per i quali la varianza dell'espressione nei soggetti normali è minore che nei soggetti diabetici.**



## 9. CONCLUSIONI E SVILUPPI FUTURI

Nello studio dei sistemi biologici, l'analisi differenziale tra i gruppi fenotipici si basa, in genere, sullo studio delle differenze tra i valori medi delle distribuzioni delle variabili nei gruppi. Questo tipo di analisi non sempre riesce a mettere in luce le differenze di espressione nei gruppi fenotipici. In questo lavoro si è scelto di studiare l'espressione differenziale focalizzando l'attenzione sulla variabilità delle distribuzioni. In questo contesto la varianza è solitamente utilizzata come misura della significatività delle differenze medie (tramite l'utilizzo di misure statistiche come il t-test o ANOVA). Raramente la variabilità di una popolazione è stata presa in considerazione nell'analisi delle differenze trascrizionali nello studio delle popolazioni. Probabilmente la varianza è stata largamente ignorata perché viene considerata solitamente nel contesto della riproducibilità sperimentale come un parametro da ridurre o un "rumore" da minimizzare attraverso precise procedure sperimentali nella costruzione dei microarray e tramite metodi di normalizzazione volti a migliorare la comparabilità tra i campioni di soggetti. Questo tipo di considerazioni erano giustificabili ai tempi della nascita dei primi microarray, ma la robustezza e la riproducibilità dell'attuale generazione di Gene Chip ci porta a riconsiderare il ruolo della varianza nello studio dell'espressione genica.

Il lavoro svolto ci conduce a delle considerazioni sull'interpretazione sia statistica che biologica dei risultati.

I test condotti hanno evidenziato che i test standard come SAM o GSEA sui dati di espressione conducono a risultati che rendono difficile se non addirittura impossibile valutare l'espressione differenziale dei geni analizzati; nel caso del test SAM, infatti, nessun gene supera la soglia della significatività statistica. Dato l'elevato numero di geni, molto maggiore del numero di microarray, e la conseguente correzione per i test multipli risulta difficile che i singoli geni raggiungano la soglia della significatività statistica, in quanto le differenze biologiche risultano essere modeste in relazione al rumore introdotto dalla tecnologia di preparazione ed elaborazione dei microarray.

Dall'analisi della varianza, invece, si riesce a capire in modo chiaro che la varianza dell'espressione genica in soggetti normali è maggiore che nei soggetti intolleranti al glucosio e diabetici. Ciò non è verificato soltanto per circa 50 geni su oltre 54000 geni analizzati; questi geni a bassa varianza presentano, inoltre, un valore medio di espressione basso, dimostrandosi misure poco affidabili considerando che valori bassi di espressione sono spesso correlati al rischio di incorrere in difetti tecnologici, come *probe* poco specifiche per il gene analizzato o artefatti e difetti dell'immagine.

Da una prima analisi dei risultati del metodo GSEA si può notare che il test sui dati della varianza porta a selezionare un maggiore numero di geni rispetto al test sui dati di espressione; dato che il software GSEA fornisce oltre all'implementazione del metodo anche l'annotazione funzionale dei dati, risulta facile notare, inoltre, che nell'analisi dei dati della varianza i gene set selezionati sono rappresentazione di una espressione differenziale coerente, in quanto sono parte di componenti cellulari o processi biologici strettamente correlati tra loro: ad esempio, la regolazione genica nei mitocondri è ampiamente presente nei dati selezionati, e le disfunzioni legate all'attività mitocondriale è risaputo sono tra le cause principale a livello cellulare del diabete mellito.

Tuttavia sviluppi futuri del presente lavoro consentiranno di valutare efficacemente le interazioni tra i geni selezionati per determinare la coerenza dell'espressione differenziale al fine di identificare quali geni caratterizzano lo stato patologico. A tal fine è necessario

avvalersi di misure che consentano di definire il livello di interazione tra i geni, in modo da poter sancire la validità biologica dei risultati. Su questa strada sarà possibile definire se l'analisi della varianza dei dati di espressione genica potrà dimostrarsi un aiuto efficace nel valutare le differenze fenotipiche che possono manifestarsi nel caso del diabete mellito, e nel capire quali pathway sono interessati da tali differenze.

# APPENDICE A

## GUIDA AI RISULTATI GSEA

L'*analysis report* presenta i risultati per ciascuno dei due fenotipi analizzati. I gene sets che hanno un ES positivo (cioè i geneset che stanno in cima alla *ranked list*) sono correlati positivamente al fenotipo analizzato, mentre quelli che presentano un ES negativo risultano correlati all'altro fenotipo.

L'*enrichment plot* fornisce una visione grafica dell'*Enrichment Score* per ogni gene set:

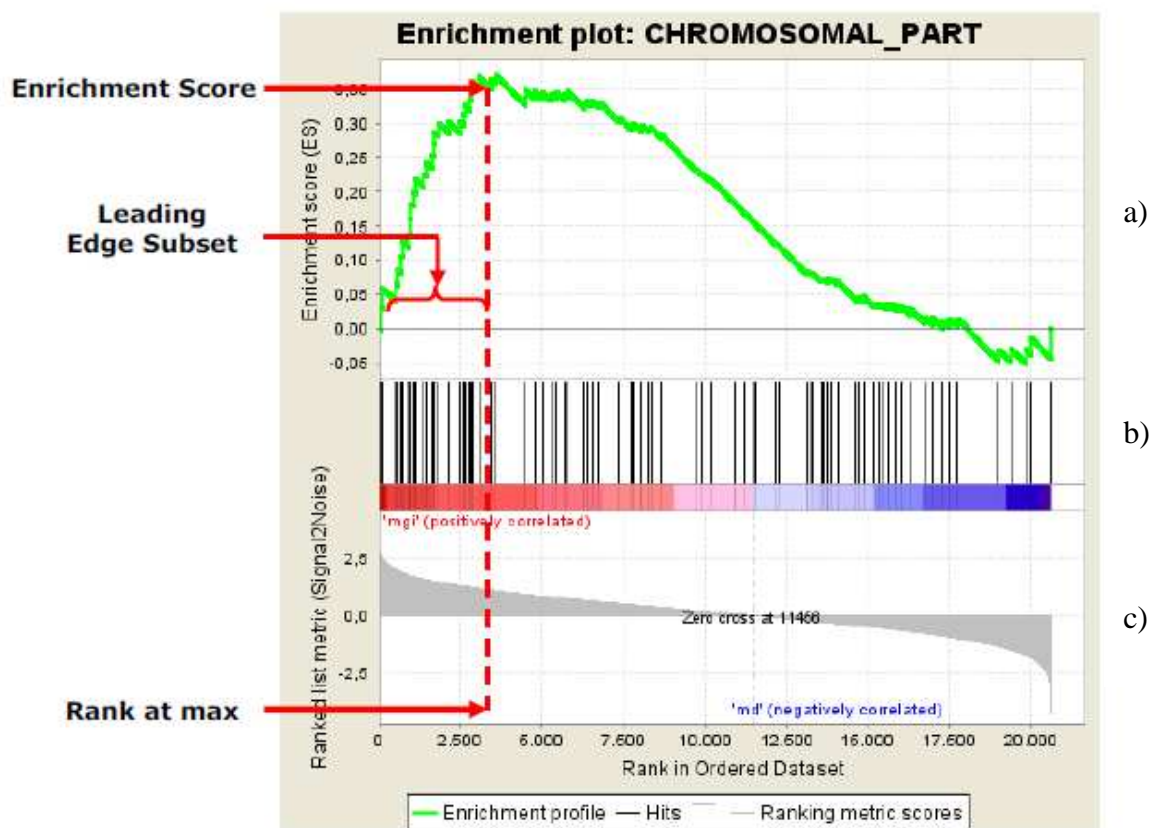


Figura A.1: Enrichment plot per il gene set “CHROMOSOMAL PART”.

- In *Figura A.1a* si mostra l'andamento dell'ES all'avanzare dell'analisi, cioè scorrendo la *ranked list*. Lo score in corrispondenza del valore di picco (il punto più lontano da 0.0) rappresenta l'ES per il gene set sottoposto all'analisi. I gene sets

che presentano picchi all'inizio o alla fine della *ranked list* sono in genere i più interessanti.

- In *Figura A.1b* si raffigura dove i membri del gene set appaiono nella *ranked list* di geni. Per ES positivi i geni appartenenti al *leading edge subset* si trovano nella *ranked list* prima che lo score assuma il valore di picco. Nel caso di ES negativi il *leading edge subset* è dato dall'insieme dei geni che si trovano dopo il valore di picco dello score.
- In *Figura A.1c* è rappresentata la correlazione di un gene con il fenotipo nello scorrere la *ranked list* (detta *ranking metric*). I valori che il *ranking metric* assume sono positivi per indicare correlazione con il primo fenotipo, mentre sono negativi per indicare correlazione con il secondo fenotipo dell'analisi.

Per ogni fenotipo l'*analysis report* mostra:

- il numero di gene sets arricchiti nell'analisi e il numero totale di gene sets analizzati
- il numero di gene sets arricchiti che sono ritenuti significativi, ovvero i gene sets che presentano un a FDR minore del 25%
- il numero di gene sets arricchiti con un p-value nominale minore dell'1% e minore del 5%
- lo *snapshot* dei risultati che riporta gli *enrichment plots* dei gene sets con NES più elevato
- vengono poi riportate note informative riguardanti la composizione del dataset, dei gene sets e della *ranked list*

Per ogni gene set oltre alle note sulla composizione vengono riportati i parametri statistici su cui si basa l'analisi dei risultati:

- i già citati ES, NES, Nominal p-value, FDR q-value
- il FWER p-value (*Familywise-error rate*) che è una stima più conservativa della probabilità che un gene set, sulla base del NES, rappresenti un falso positivo.
- RANK AT MAX: la posizione nella *ranked list* alla quale viene raggiunto l'*Enrichment Score*
- LEADING EDGE: mostra i tre valori statistici utilizzati per definire il *leading edge subset*:
  - *Tags*: la percentuale di geni raggiunta prima (per ES positivi) o dopo (per ES negativi) il picco dell'*Enrichment Score* nell'analisi corrente.
  - *List*: la percentuale di geni nella *ranked gene list* prima (per ES positivi) o dopo (per ES negativi) il picco dell'*Enrichment Score*.



- *Signal*: la combinazione delle due statistiche precedenti:

$$(Tag\%)(1 - Gene\%) \left( \frac{N}{N - N_h} \right)$$

dove N è il numero di geni nella *ranked list* e  $N_h$  è il numero di geni nel gene set. Se il gene set è completamente contenuto nelle prime  $N_h$  posizioni della *ranked list* il valore della variabile *Signal* è massimo (100%), per decrescere altrimenti via via verso lo 0%.

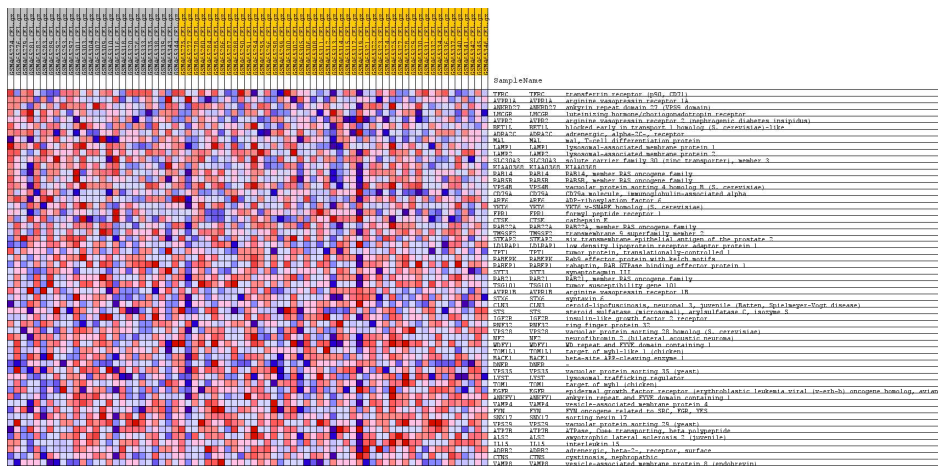


Figura A.2: Heatmap.

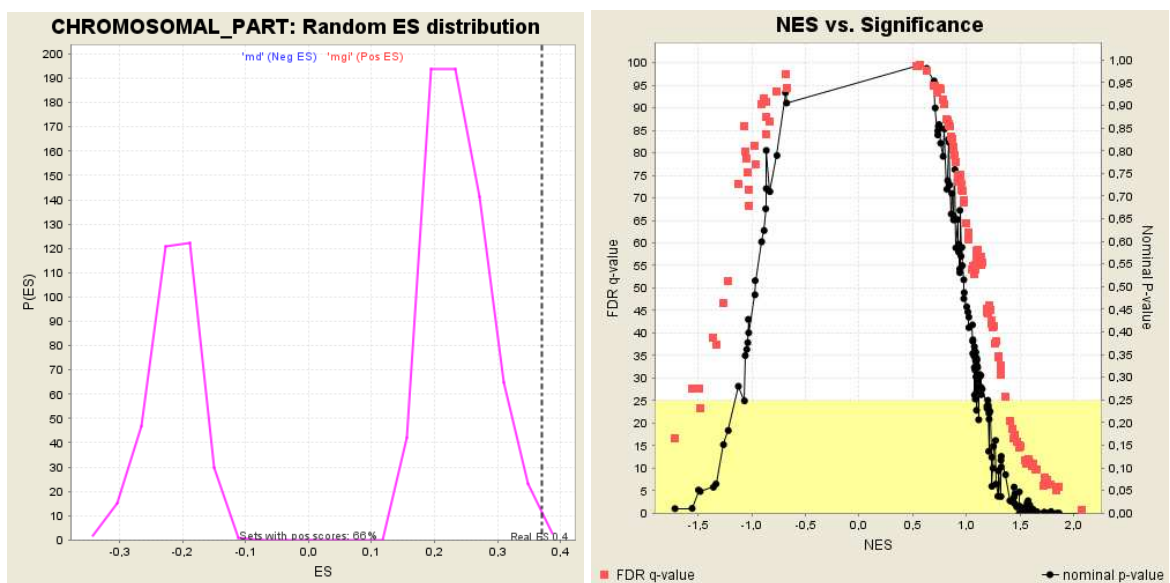


Figura A.3: p-value vs. NES e NES vs. Significance.

Butterfly plot for: DATAnorm\_collapsed\_to\_symbols.  
 muscle\_normal\_vs\_muscle\_glucoseIntollerant.  
 cls#muscle\_glucoseIntollerant\_versus\_muscle\_normal.  
 muscle\_normal\_vs\_muscle\_glucoseIntollerant.  
 cls#muscle\_glucoseIntollerant\_versus\_muscle\_normal\_r  
 epos

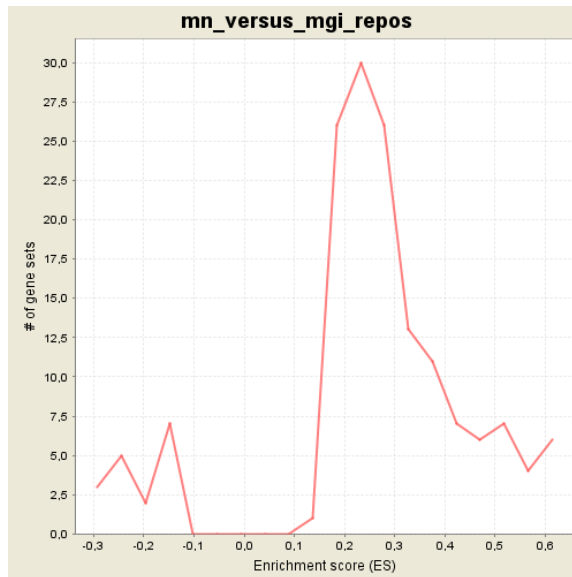
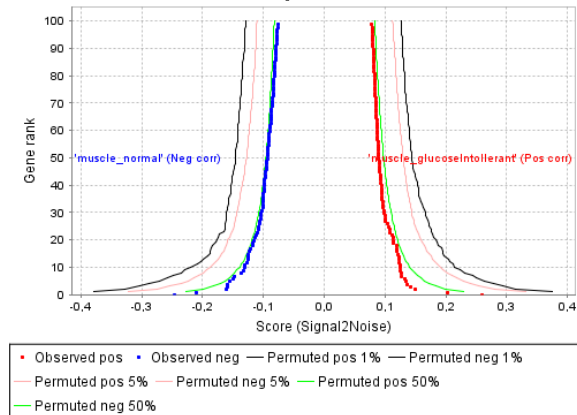


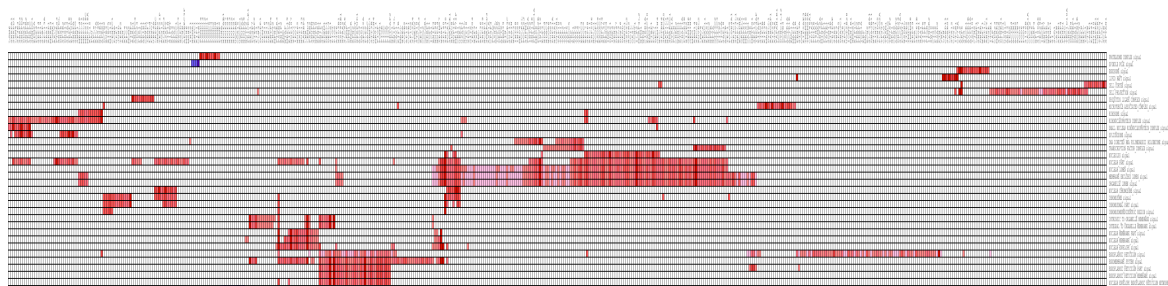
Figura A.4: butterfly plot e global ES histogram.

## LEADING EDGE ANALYSIS REPORT

I risultati della Leading Edge Analysis vengono rappresentati attraverso quattro grafici:

- Heatmap
- Set-to-set
- Gene in subsets
- Histogram

### Heatmap



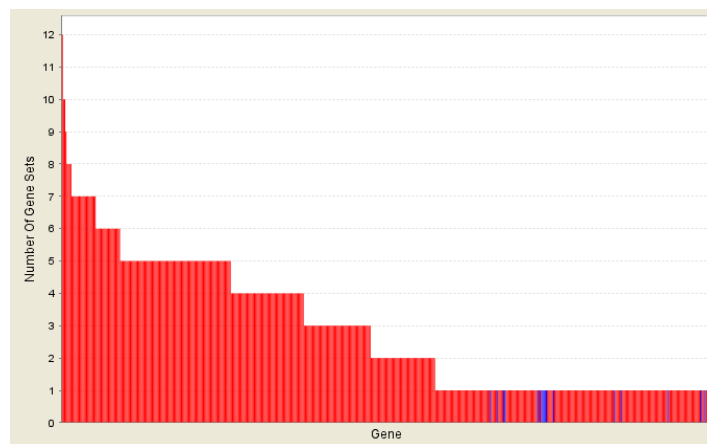
Raffigura i geni raggruppati dei vari leading edge subset per i gene sets selezionati ( FDR q-value < 25%). I valori di espressione sono rappresentati da una scala di colori che varia dal blu (per i valori di espressione più bassi) al rosso (per i valori di espressione più elevati)

## Set-to-set



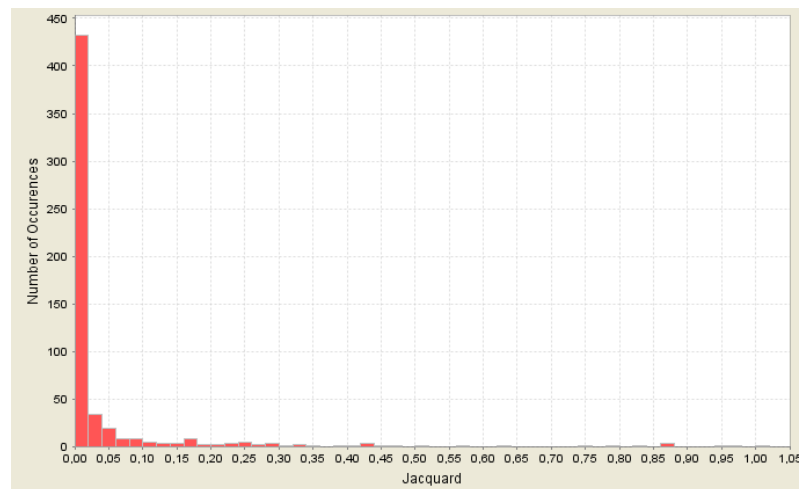
Raffigura il grado di sovrapposizione tra i leading edge subsets attraverso una scala di colori. Più il colore è intenso, maggiore è la sovrapposizione tra i subsets. In particolare l'intensità della cella relativa al confronto tra i sets A e B, corrisponde al rapporto  $X/Y$ , dove  $X$  è il numero di geni che appartengono al leading edge subset per il set A, mentre  $Y$  è l'unione dei geni dei set A e B che appartengono ai *leading edge subsets*. Una cella verde indica un elevato grado di sovrapposizione, mentre una cella bianca indica che non c'è sovrapposizione tra i due *leading edge subsets*.

## Gene in subsets



Rappresenta per ogni gene il numero di volte che compare in un *leading edge subset*.

## Histogram



L'istogramma raffigura il l'intersezione divisa per l'unione per una coppia di *leading edge subsets* (detta *Jacquard*), e il relativo numero di occorrenze, cioè il numero di coppie di *leading edge subsets* in un particolare intervallo. Se *Jacquard* = 0.0 non c'è sovrapposizione tra la coppia di subsets.

# Bibliografia

1. Jessica C. Mar, Nicholas A. Matigian, [...], and Christine A. Wells (2011), Variance of Gene Expression Identifies Altered Network Constraints in Neurological Disease
2. Subramanian, A., Kuehn, H., Gould, J., Tamayo, P., Mesirov, J.P. (2007), GSEA-P: A desktop application for Gene Set Enrichment Analysis
3. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S. & Mesirov, J. P. (2005), Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles
4. Mootha, V. K., Lindgren, C. M., Eriksson, K. F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstrale, M., Laurila, E., *et al.* (2003). PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes
5. Virginia Goss Tusher, Robert Tibshirani, and Gilbert Chu (2000), Significance analysis of microarrays applied to the ionizing radiation response
6. (2006) The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotech* 24: 1151–1161.
7. Levisky JM, Shenoy SM, Pezo RC, Singer RH (2002) Single-cell gene expression profiling. *Science* 297: 836–840.
8. Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A (2002) Regulation of noise in the expression of a single gene. *Nat Genet* 31: 69–73.
9. Cai L, Friedman N, Xie XS (2006) Stochastic protein expression in individual cells at the single molecule level. *Nature* 440: 358–362.
10. Mar JC, Rubio R, Quackenbush J (2006) Inferring steady state single-cell gene expression distributions from analysis of mesoscopic samples. *Genome Biol* 7: R119.
11. Raj A, Rifkin S, Andersen E, Oudenaarden Av (2010) Variability in gene expression underlies incomplete penetrance. *Nature* 463: 913–919.
12. Ravasi T, Wells C, Forest A, Underhill DM, Wainwright BJ, *et al.* (2002) Generation of diversity in the innate immune system: macrophage heterogeneity arises from gene-autonomous transcriptional probability of individual inducible genes. *J Immunol* 168: 44–50.

13. Colman-Lerner A, Gordon A, Serra E, Chin T, Resnekov O, et al. (2005) Regulated cell-to-cell variation in a cell-fate decision system. *Nature* 437: 699–706.
14. Eichler EE, Flint J, Gibson G, Kong A, Leal SM, et al. (2010) Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet* 11: 446–450.
15. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, et al. (2009) Finding the missing heritability of complex diseases. *Nature* 461: 747–753.
16. Feinberg AP, Irizarry RA (2010) Evolution in health and medicine Sackler colloquium: Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. *Proc Natl Acad Sci U S A* 107 Suppl 1: 1757–1764.
17. Raser J, O’Shea E (2005) Noise in gene expression: origins, consequences, and control. *Science* 309: 2010–2013.
18. Murrell W, Feron F, Wetzig A, Cameron N, Splatt K, et al. (2005) Multipotent stem cells from adult olfactory mucosa. *Dev Dyn* 233: 496–515.
19. Murrell W, Wetzig A, Donnellan M, Féron F, Burne TH, et al. (2008) Olfactory mucosa is potential source for autologous stem cell therapy for Parkinson’s disease. *Stem Cells* 26: 2183–2192.

## Web Resource

1. <http://www.broadinstitute.org/cancer/software/gsea/wiki/index.php>
2. <http://www.broadinstitute.org/gsea/doc/GSEAUUserGuideFrame.html>
3. <http://www.broadinstitute.org/gsea/msigdb/index.js>
4. <http://www.affymetrix.com>