

# Sommario

|   |           |
|---|-----------|
| <b>1. INTRODUZIONE</b> .....                              | <b>2</b>  |
| 1.1. DIABETE: CLASSIFICAZIONE ED EPIDEMIOLOGIA .....      | 3         |
| 1.2. DIABETE MELLITO.....                                 | 4         |
| 1.3. MODY: MATURITY ONSET DIABETES OF THE YOUNG.....      | 11        |
| 1.4. NEXT GENERATION SEQUENCING (NGS).....                | 22        |
| <b>2. SCOPO DELLA TESI</b> .....                          | <b>34</b> |
| <b>3. MATERIALI E METODI</b> .....                        | <b>36</b> |
| 3.1. CASISTICA ANALIZZATA .....                           | 37        |
| 3.2. ESTRAZIONE DEL DNA DA SANGUE PERIFERICO .....        | 37        |
| 3.3. DETERMINAZIONE DELLA CONCENTRAZIONE DEL DNA .....    | 38        |
| 3.4. ANALISI MOLECOLARE DEI GENI IMPLICATI NEL MODY ..... | 39        |
| 3.5. ELETTROFORESI SU GEL DI AGAROSIO .....               | 42        |
| 3.6. PURIFICAZIONE DEI PRODOTTI DI PCR.....               | 43        |
| 3.7. REAZIONE DI SEQUENZA .....                           | 43        |
| 3.8. PURIFICAZIONE DELLA REAZIONE DI SEQUENZA.....        | 44        |
| 3.9. INTERPRETAZIONE DEI RISULTATI .....                  | 45        |
| 3.10. ANALISI MOLECOLARE MEDIANTE SISTEMA ROCHE/454 ..... | 47        |
| <b>4. RISULTATI</b> .....                                 | <b>55</b> |
| <b>5. DISCUSSIONE</b> .....                               | <b>67</b> |
| <b>6. CONCLUSIONI</b> .....                               | <b>77</b> |
| <b>BIBLIOGRAFIA</b> .....                                 | <b>80</b> |

# 1. Introduzione

---

## 1.1. *Diabete: Classificazione ed epidemiologia*

### **Storia del Diabete**

Il termine diabete fu coniato da Areteo di Cappadocia (81-138d.C) medico dell'antica Roma; anche se la sua descrizione è stata ritrovata su papiri egizi nella località di Ebers (1500 a.C). Il termine mellito venne aggiunto in relazione alla caratteristica presenza di glucosio nelle urine in pazienti con diabete (glucosinuria) nel 1974 da Thomas Willis (1).

La causa responsabile del diabete venne individuata agli inizi del 1900, grazie agli studi di due medici tedeschi Josef von Mering e Oskan Minkowski, che osservarono la comparsa di sintomi attribuibili al diabete, in cani ai quali era stato asportato il pancreas. Nello stesso anno Frederick Banting riuscì a isolare l'estratto pancreatico, che noi oggi sappiamo contenere l'ormone insulina, responsabile della regolazione della concentrazione di glucosio nel sangue. In esperimenti successivi, nel 1921, Banting inietto l'estratto pancreatico in alcuni cani ai quali era stato asportato il pancreas, ripristinandone così i livelli di glucosio nel sangue. L'insulina così prodotta venne in seguito testata su un giovane paziente diabetico dal biochimico J. Collip, con conseguente miglioramento delle sue condizioni. Gli studi sull'insulina condotti da Banting e Macloed li valsero l'attribuzione del Premio Nobel nel 1923:

*“insulin is not a cure for diabetes; it is a treatment. It enables the diabetic to born sufficient carbohydrates, so that proteins and fats may be added to the diet in sufficient quantities to provide energy for the economic burdens of life”*

Nei successivi 60 anni l'insulina venne purificata e raffinata fino all'avvento della tecnologia del DNA ricombinante che portò alla sintesi dell'insulina nel 1978 (1) (2) (3).

## 1.2. *Diabete Mellito*

Il Diabete Mellito (DM) è un disordine metabolico caratterizzato da alti livelli di glucosio nel sangue. In condizioni normali a seguito di un pasto viene rilasciato dalle cellule  $\beta$  del pancreas l'ormone insulina, che stimola le cellule del muscolo e le cellule lipidiche a rimuovere lo zucchero in eccesso nel sangue. Nei soggetti affetti da diabete mellito questi livelli rimangono alti perché le cellule del pancreas non producono insulina, o non è prodotta in quantità sufficiente.

### **La biochimica del glucosio**

Il glucosio è il carburante essenziale per il nostro organismo, la sua concentrazione nel sangue è regolata principalmente dall'ormone insulina. I livelli di glucosio nel sangue non sono costanti, ma dipendono dalle esigenze dell'organismo in un dato momento, generalmente si attestano attorno ai 70-110 mg/dl. Ciò che fa innalzare tali livelli sono: la dieta, il catabolismo del glicogeno e la sintesi epatica del glucosio (4).

L'insulina è una piccola proteina prodotta dalle cellule  $\beta$  delle isole di Langerhans del pancreas. La sintesi comincia con la trascrizione di un gene presente sul cromosoma 11, che porta alla produzione di una proteina formata da 110 amminoacidi(aa), la preproinsulina, la quale in questa fase risulta è inattiva. Questa contiene nella sua sequenza un peptide segnale, di circa 24 aa, il quale viene riconosciuto e tagliato a livello della membrana cellulare, dando luogo alla formazione della proinsulina, che viene conservata nei granuli di secrezione delle cellule  $\beta$  del pancreas. Questa consta in tre domini: una catena  $\beta$  all'estremità amminoterminale, una catena  $\alpha$  all'estremità carbossi-terminale e un dominio di connessione chiamato C-peptide. Quando i livelli di glucosio nel sangue aumentano stimolano la secrezione dell'insulina, di conseguenza la proinsulina viene convertita nell'ormone attivo da parte di proteasi specifiche che rompono due legami peptidici e generano la molecola dell'insulina matura e attiva (4)(Fig.1).

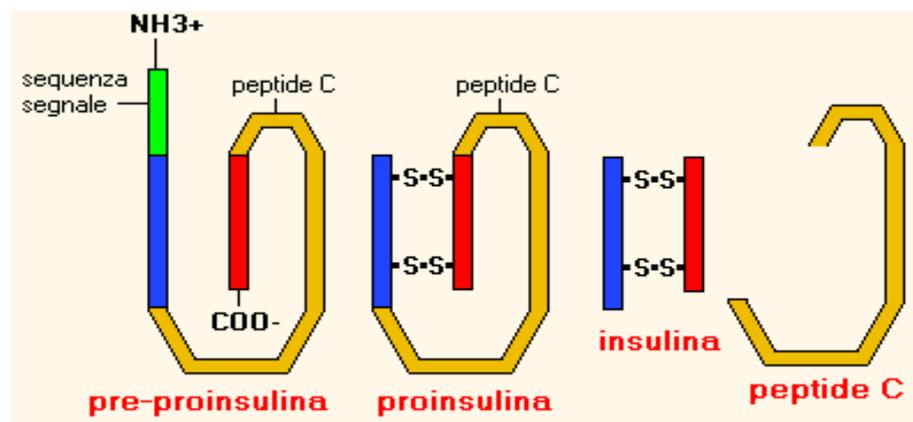
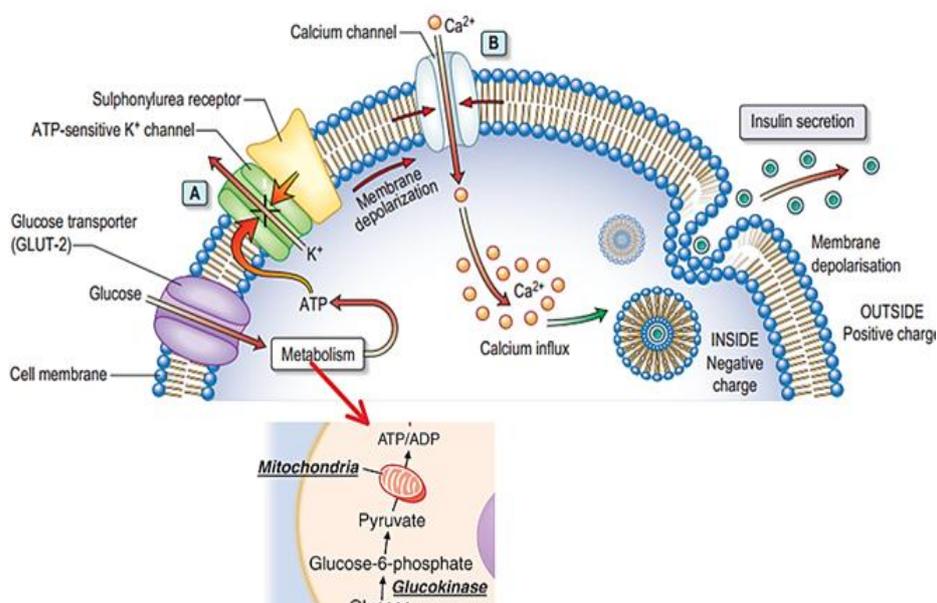


Figura 1: Maturazione dell'ormone insulina

Quando la concentrazione di glucosio nel sangue aumenta, il trasportatore di glucosio GLUT2 lo convoglia verso le cellule  $\beta$  del

pancreas dove viene convertito in glucosio 6-fosfato dall'enzima esochinasi IV (glucochinasi) ed entra nel processo glicolitico. L'aumento della velocità del catabolismo del glucosio porta all'incremento della concentrazione di ATP (adenosina trifosfato) causando la chiusura dei canali di Potassio ATP-dipendenti della membrana (bersaglio dei farmaci sulfanilureici). Questo si ripercuote sulla polarità della membrana che a causa del ridotto efflusso di  $K^+$  induce la sua depolarizzazione favorendo l'apertura dei canali del  $Ca^{2+}$ , il cui aumento della concentrazione intracellulare induce il rilascio per esocitosi di insulina (Fig.2). Oltre al glucosio altri fattori possono stimolare il rilascio di insulina dalla cellula  $\beta$ , tra cui gli ormoni, i neurotrasmettitori e l'acetilcolina; il glucosio aumenta la risposta della cellula a questi fattori (4)



**Figura 2: Regolazione della secrezione di insulina prodotta dalle cellule  $\beta$  del pancreas in base alla concentrazione di glucosio nel sangue**

L'insulina produce un'ampia varietà di effetti, che possono essere istantanei (pochi secondi), come la modulazione del trasporto di ioni  $K^+$  e del glucosio all'interno della cellula, precoci

(entro pochi minuti) come la regolazione degli enzimi metabolici o ritardati (entro ore o giorni) come gli effetti sulla crescita e sul differenziamento cellulare (5).

### Classificazione

Attualmente il DM è classificato in base al processo patogenetico che conduce all'iperglicemia. Secondo l'*American Diabetes Association* (ADA), che sancisce annualmente gli standard di questa patologia, la classificazione del diabete include quattro categorie cliniche (Tab.1):

Tabella 1: classificazione del diabete secondo L'American Diabetes Association

| <b>CLASSIFICAZIONE DEL DIABETE</b><br><i>American Diabetes Association</i> |   |
|--|---|
| <b>Tipo 1</b>  | Diabete mellito immuno-mediato  |
|  | Diabete mellito idiopatico  |
| <b>Tipo 2</b>  | Insulino-resistenza con relativa insulino-deficienza  |
| <b>Diabete Mellito Gestazionale (GDM)</b>                                  | Intolleranza ai carboidrati con iperglicemia di gravità variabile, che inizia o che viene riconosciuta per la prima volta durante la gravidanza |
| <b>Altre forme specifiche</b>  | <b>Difetti genetici della funzione <math>\beta</math>-cellulare dovuti a mutazione di:</b>  |
|  | <i>MODY</i>   |
|  | DNA mitocondriale   |
|  | Conversione proinsulina e insulina  |
|  | Difetti genetici dell'azione dell'insulina  |
|  | Insulino-resistenza di tipo A   |
|  | Leprecaunismo   |
|  | Sindrome di Rabsoon-Mendenhall  |
|  | Lipodistrofia   |
|  | Disturbi del pancreas esocrino (p.e. carcinoma pancreatico, pancreatite, pancreatectomia)   |

|  |  |
|--|--|
|  | Endocrinopatia (p.e. sindrome di Cushing, acromegalia, glucagonoma, feocromocitoma)                        |
|  | Indotto da farmaci o agenti chimici (p.e. steroidi, antiretrovirali, tiazidici, $\beta$ -bloccanti)        |
|  | Infezioni (p.e. rosolia congenita, Coxsackie, CMV)   |
|  | Forme non comuni di diabete immuno-correlato (anticorpi antirecettori dell'insulina, sindrome "stiff-man") |
|  | Altre sindromi genetiche (p.e. Sindrome di Down, Klinefelter, Turner, Wolfram, Friedreich)                 |

Il Diabete tipo 1 (DT1), interessa l'8% della popolazione affetta da diabete. Questa tipologia comprende il diabete di origine autoimmune, che deriva dalla distruzione progressiva e selettiva delle cellule  $\beta$  pancreatiche ad opera di linfociti T autoreattivi, determinando un deficit assoluto di insulina. È caratterizzato dallo sviluppo di chetoacidosi, (evento patologico acuto caratterizzato da elevati livelli ematici di glucosio e di corpi chetonici), in assenza di terapia insulinica. La chetoacidosi diabetica è inoltre aggravata dalle infezioni e altri eventi come lo stress associato ad interventi chirurgici (5).

In Italia l'incidenza è tra 6-10 su 100000 casi all'anno ed è massima entro i 29 anni di età; un caso particolare è la Sardegna che ha un'incidenza annua di 34 su 100000 casi/anno, ed è uno dei paesi del mondo, insieme alla Finlandia, ad avere la più alta diffusione della malattia (6).

Il Diabete tipo 2 (DT2) è un disordine metabolico caratterizzato da iperglicemia ed alterato metabolismo lipidico, causato dall'incapacità, nel corso del tempo, delle cellule  $\beta$  di secernere un'adeguata quantità di insulina. Si osserva una netta riduzione di tali cellule e una diminuzione della responsività dei tessuti periferici all'azione insulinica. Si presenta solitamente negli adulti obesi e si caratterizza per una glicemia modesta che

raramente conduce a chetoacidosi. Questa tipologia di diabete è spesso causa della sindrome da insulinoresistenza, ovvero una sindrome metabolica caratterizzata da ipertensione, aterosclerosi ed obesità (5).

Il DT2 rappresenta circa il 90% dei casi di diabete. Nel 2010 è stato stimato che in tutto il mondo le persone adulte affette da diabete ammontano a 285 milioni (circa 6%), di cui 18 milioni solo in America. Tale dato è destinato ad aumentare nei prossimi anni, infatti è stato stimato che entro il 2030 questo dato possa raggiungere i 439 milioni di persone (7). Alcuni studi mostrano che circa 2 milioni di persone nel mondo sono morte a causa del DT2, rappresentando la sesta causa di morte negli USA (8).

Il DT2 sta diventando ormai un'epidemia, negli USA è stato stimato che dei bambini nati dal 2000, 1 su 3 soffre di diabete. Questi dati sono estremamente allarmanti, basti pensare che fino a un decennio fa la forma di DM2 era molto rara da riscontrare tra giovani e bambini, il nome era infatti "*adult-onset diabetes*". L'aumento dei casi di diabete va di pari passo con l'aumento dei soggetti obesi. Nel 1991 gli obesi ammontavano al 12% della popolazione mondiale, dal 2001 questo dato è andando aumentando raggiungendo il 20% della popolazione, di cui 44 milioni in America. Possiamo quindi considerare il DT2 la malattia del progresso (9) (10) (11).

L'aumento del rischio è dunque associato principalmente a fattori quali: predisposizione genetica, età, etnia, familiarità, fumo, l'obesità e vita sedentaria. Le complicanze che ne conseguono includono malattie cardiovascolari, neuropatie, patologie renali, cecità e amputazione degli arti (8).

### Valutazione clinica e trattamento del paziente diabetico

Secondo L'Associazione Diabetica Americana l'iperglicemia, glucosio a digiuno, con valori di almeno 126 mg/dl o glicemia random 200 mg/dl, è il criterio diagnostico principale del diabete. La valutazione dell'emoglobina glicata è inoltre utilizzata per monitorare il controllo glicemico nei pazienti con diabete noto (5).

L'obiettivo della terapia è il controllo costante della glicemia, per contrastare le complicanze microvascolari associate al diabete.

I farmaci maggiormente impiegati nella terapia attualmente sono:

**Sulfaniluree:** agiscono aumentando il rilascio di insulina chiudendo i canali ATP del  $K^+$  nella membrana della cellula  $\beta$ . Poiché il DM1 è caratterizzato dalla distruzione delle proprie cellule  $\beta$ , questa terapia in tali soggetti, risulta essere pertanto inefficace.

**Biguanidi:** così come la metformina riducono la produzione epatica di glucosio, e aumentano l'utilizzo del glucosio mediato dall'insulina del muscolo scheletrico e negli epatociti.

**Inibitori dell' $\alpha$ -glucosidasi:** ritardano l'assorbimento intestinale dei carboidrati.

**Tiazolidinedioni:** riducono l'insulinoresistenza del muscolo scheletrico.

**Peptide simile al glucagone-1:** stimola la biosintesi insulinica e l'espressione del gene dell'insulina, ha effetti trofici e antiapoptotici sulle cellule  $\beta$ . L'uso dei GLP-1 è ancora in fase di studi clinici (5).

### 1.3. *Mody: Maturity onset diabetes of the young*

I primi a descrivere questa tipologia di diabete furono Tattersal (U.K) e Fajan (USA). Tattersal nel 1974 a seguito di osservazioni effettuate su un gruppo di giovani pazienti non trattati con insulina per due anni dopo la diagnosi (12). L'endocrinologo Fajan nel 1989, grazie a studi di *Linkage* condotti su famiglie analizzate nel corso di più generazioni, identificò il gene responsabile di tale patologia, ovvero *HNF1a*, le cui mutazioni portano al fenotipo Mody 3. Nel corso degli anni vennero progressivamente definite le caratteristiche cliniche che lo differenzia dalle altre tipologie di DM (12).

Il termine *MODY Maturity-Onset of the Young* nasce dall'esigenza di creare un nuovo sottogruppo nella vecchia classificazione del diabete (13):

- yung-onset **DT1**(insulino dipendente)
- maturity-onset **DT2**(insulino indipendente)

Il *MODY* è la forma monogenica del DM trasmessa con meccanismo autosomico dominante, ad alta penetranza caratterizzata da (13):

- insorgenza giovanile(prima dei 25 anni)
- insulino- indipendente
- moderata, ma insufficiente, concentrazione del peptide C circolante
- autoanticorpi pancreatici negativi

È dovuta alla disfunzione delle cellule β del pancreas, a causa di mutazioni a carico di fattori di trascrizione nucleari, che sono importanti per la sintesi dell'insulina, e del gene *GCK* (14).

Il primo gene ad essere stato associato al Mody è stato *GCK* che codifica per l'enzima esokinasi IV, seguito dal Fattore Nucleare degli epatociti *HNF1α* (*TCF1*), *HNF4α* (12).

### Classificazione e genetica del MODY

Ad oggi sono state individuate 13 forme di MODY (Tab.2) (12):

Tabella 2:Classificazione genetica e fenotipica dei vari sottotipi MODY

| Tipo          | Gene           | Locus       | Funzione Gene   | Difetto primario       | Freq.  |
|---------------|----------------|-------------|---|------------------------|--------|
| <b>MODY1</b>  | <i>HNF4α</i>   | 20q12-q13.1 | Fattore trascrizionale nucleare   | Pancreas               | 5-10%  |
| <b>MODY2</b>  | <i>GCK</i>     | 7p15-p13    | Esokinasi IV  | Pancreas/<br>fegato    | 35-40% |
| <b>MODY3</b>  | <i>HNF1α</i>   | 12q24.2     | Fattore trascrizionale  | Pancreas/<br>reni      | 20-30% |
| <b>MODY4</b>  | <i>IPF-1</i>   | 13q12.1     | Fattore trascrizionale  | Pancreas               | <1%    |
| <b>MODY5</b>  | <i>HNF1β</i>   | 17q12       | Fattore trascrizionale  | Reni/<br>pancreas      | 5-10%  |
| <b>MODY6</b>  | <i>NEUROD1</i> | 2q31.3      | Fattore trascrizionale(bHLH)  | Pancreas               | Rara   |
| <b>MODY7</b>  | <i>KLF11</i>   | 2p25        | Fattore di crescita e differenziamento  | Pancreas               | Rara   |
| <b>MODY8</b>  | <i>CEL</i>     | 9q34.3      | regola le funzioni endocrine ed esocrine del pancreas                               | Pancreas               | Rara   |
| <b>MODY9</b>  | <i>PAX4</i>    | 7q32        | Fattore trascrizionale  | Pancreas               | Rara   |
| <b>MODY10</b> | <i>INS</i>     | 11p15.5     | Cellule β isole di Langerhans   | NF-Kappa B             | Rara   |
| <b>MODY11</b> | <i>BLK</i>     | 8p23-p22    | Tirosinkinasi(linfoB)   | MIN6 βcell             | Rara   |
| <b>MODY12</b> | <i>ABCC8</i>   | 11p15.1     | Codificato la subunità <i>SUR 1</i>   | Celluleβ del pancreas  | Rara   |
| <b>MODY13</b> | <i>KCNJ11</i>  | 11p15.1     | responsabile della produzione della sub unità <i>KIR 6.2</i> della pompa K ATPasica | cellule β del pancreas | Rara   |

### **HNF4 $\alpha$ MODY1 e HNF1 $\alpha$ MODY3**

*HNF4 $\alpha$*  e *HNF1 $\alpha$*  fanno parte di una famiglia di fattori trascrizionali nucleari responsabili della regolazione dell'espressione genica epatica e delle cellule  $\beta$  del pancreas.

*HNF4  $\alpha$*  (20q12-q13.1) è un fattore trascrizionale appartenente alla famiglia dei recettori per gli ormoni steroidei ed è espresso principalmente a livello del fegato, reni e pancreas. Si lega al DNA in forma di omodimero e regola la trascrizione di geni coinvolti nel metabolismo glucidico e lipidico. Mutazioni in eterozigosi causano progressiva perdita della secrezione dell'insulina (*loss of function/ gene dosage effect*), oltre ad alterati livelli di trigliceridi nel sangue. Questa tipologia di MODY richiede pertanto trattamento con insulina, e può portare nel tempo a complicazioni vascolari (12) (14) (15).

*HNF1 $\alpha$*  è localizzato sul cromosoma 12q24.2. Come *HNF4 $\alpha$*  è espresso principalmente a livello di pancreas, reni e fegato; è composto da 10 esoni espressi in maniera differente nei vari tessuti; si lega al DNA in forma di omodimero o etero dimero con *HNF1 $\beta$* . Regola l'espressione di geni coinvolti nel trasporto del glucosio all'interno del mitocondrio, oltre a regolare geni coinvolti nella proliferazione cellulare ed apoptosi (12) (15).

### **CGK MODY 2**

*GCK* codifica per l'enzima esokinasi IV, responsabile della prima tappa della glicolisi. Il glucosio viene convertito in glucosio-6- fosfato (G6P) dopo il suo ingresso nella cellula  $\beta$ , grazie al trasportatore GLUT-2, altro membro del sistema "sensore" per il glucosio, con conseguente rilascio di insulina al variare della glicemia plasmatica (Fig.3) (16) (14).

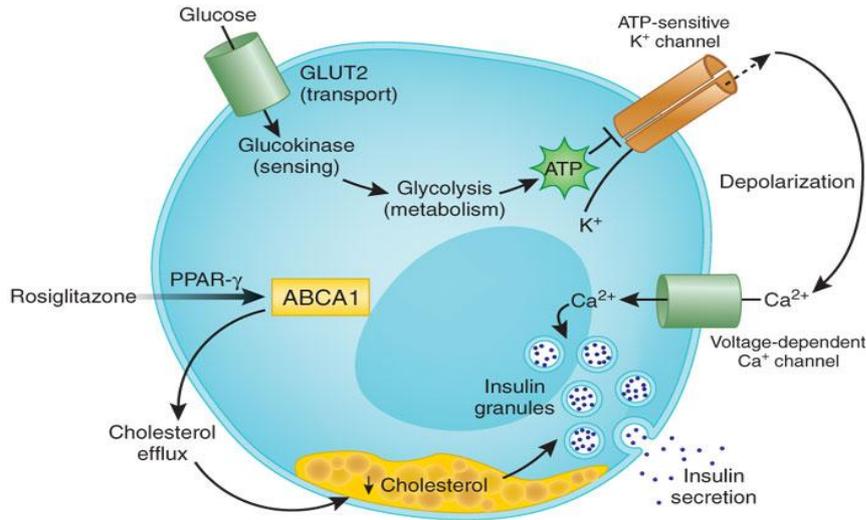


Figura 3: Meccanismo d'azione del rilascio di insulina in seguito all'aumento della concentrazione di glucosio.

Il gene *GCK* è situato sul cromosoma 7p15.3- p15.1; è espresso in pancreas, reni, fegato, cervello e cellule endocrine dell'apparato digerente; l'espressione nei vari tessuti è promotore specifica, infatti l'esone 1b e 1c sono espressi nel fegato, mentre l'esone 1a nelle cellule  $\beta$  del pancreas (16).

Mutazioni in eterozigosi del gene *GCK* portano a due diversi fenotipi: MODY2, caratterizzato da un aumento modesto della glicemia a digiuno nell'intervallo dei 110- 130 mg/dl spesso asintomatica, ed Iperinsulinemia persistente dell'infanzia (17).

### **IPF1/PDX1 MODY4**

*PDX1/IPF1* (*pancreatic and duodenal homeobox 1*), situato sul cromosoma 13q12.1, è conosciuto anche come promotore insulinico 1, è un fattore trascrizionale necessario per lo sviluppo pancreatico e della maturazione delle cellule  $\beta$ . *IPF1* (12).

Mutazioni in omozigosi sembrano essere associate ad agenesia del pancreas e in eterozigosi con MODY4 (14).

### **HNF1 $\beta$ /TCF2 MODY5**

Il fenotipo MODY5 è causato da mutazioni a livello del gene *HNF1 $\beta$*  (*hepatocyte nuclear factor 1 beta*), localizzato sul cromosoma 17q12. È anche esso un fattore trascrizionale attivo nella forma di omodimero o etero dimero assieme a *HNF1 $\alpha$* . Gioca un ruolo fondamentale nello sviluppo dei reni e differenziazione del nefrone, oltre ad essere un fattore critico nella regolazione della crescita e differenziazione del pancreas (12) (14).

Mutazioni a carico di questo gene portano ad anomalie congenite del tratto urogenitale, e disfunzioni a livello renale. Si tratta per lo più di mutazioni *non sense* distribuite su tutto il gene, è proprio l'aploinsufficienza a causare il fenotipo diabetico. In relazione al quadro clinico è stata definita una nuova sindrome: la sindrome RCAD, (*Renal Cystitis and Diabetes*). Dati i pochi casi finora identificati rimane ancora da definire una precisa correlazione genotipo-fenotipo (12) (14).

### **NEUROD1 MODY 6**

Il gene *NEUROD1* (*Neurogenic differentiation factor 1*), localizzato sul cromosoma 2q31.3, è chiamato anche  $\beta$ 2Naya, è un fattore trascrizionale della famiglia *Basic-helix-loop-helix (Bhlh)*, importante nel differenziamento di pancreas e sistema nervoso (12) (14).

### **KLF11 MODY 7**

*Krueppel-like factor 11* è una proteina che nell'uomo è codificata dal gene *KLF11*, situato sul cromosoma 2p25; regola la crescita delle cellule esocrine e si comporta come un soppressore tumorale. Tale gene gioca un ruolo importante nel *signaling* della

glicemia nelle cellule  $\beta$  del pancreas, per questo associato al MODY7 (12) (14).

### **CEL MODY 8**

Il gene *CEL* (*Carboxyl-ester lipase*) è situato sul cromosoma 9q34.3, regola le funzioni endocrine ed esocrine del pancreas. Le mutazioni che portano al fenotipo del MODY8 sono delezioni *frameshift* nelle regioni ricche di ripetizioni in tandem (VNTR) del gene (12).

### **PAX4 MODY9**

Mutazioni a carico del gene *PAX4* (*Paired box gene 4*), situato sul cromosoma 7q32, portano al fenotipo MODY9. Tale gene è membro di una famiglia di fattori trascrizionali *PAX* (*Paired Box*) chiamati così in quanto sono provvisti di particolari domini *paired box*. Svolge un ruolo importante nello sviluppo fetale, nel processo carcinogenico e nello sviluppo delle cellule pancreatiche (12) (14).

### **INS MODY 10**

*INS* codifica per il precursore della proinsulina, è localizzato sul cromosoma 11p15.5. Mutazioni in eterozigosi sono già state precedentemente associate a diabete neonatale permanente. Tuttavia studi recenti hanno associato mutazioni in *INS* a un nuovo fenotipo clinico, MODY10 (12).

### **BLK MODY 11**

*BLK*, situato sul cromosoma 8p23-p22, è un gene che codifica per una tirosin-chinasi dei linfociti B della famiglia dei proto-oncogeni *SRC*. E' espresso nelle cellule  $\beta$  pancreatiche, dove stimola la sintesi e la secrezione di insulina indotta dal glucosio,

tramite un'azione attivatoria nei confronti di alcuni fattori di trascrizione come *IPF1* (12).

### **ABCC8 MODY12 e KCNJ11 MODY 13**

Di recentissima scoperta sono le forme di MODY12 e 13 attribuibili rispettivamente a mutazioni a carico dei geni *ABCC8* e *KCNJ11* situati entrambi sul cromosoma 11p15.1. Quest'ultimo gene è responsabile della produzione della sub unità *KIR 6.2* della pompa K ATPasica, nelle cellule beta. La pompa potassio ATP dipendente (K-ATP) è un etero ottamero formato da 4 subunità *KIR 6.2* e 4 subunità *SUR 1* (*sulfonylurea receptor 1*) codificato dal gene *ABCC8*. Il canale è implicato nell'attività elettrica della membrana delle cellule beta, per questo gioca un ruolo fondamentale nella secrezione insulinica (18) (19).

### **Caratteristiche cliniche MODY**

Per poter parlare di MODY sono ormai universalmente accettati cinque criteri diagnostici:

Iperglicemia, solitamente accertata entro i 25 anni, in almeno uno o due membri della famiglia. Sono stati riportati casi di diagnosi anticipate nelle successive generazioni, probabilmente dovute ad una maggiore cognizione del diabete che porta ad effettuare test in giovane età;

Ereditarietà autosomica dominante, con una trasmissione verticale del diabete per almeno tre generazioni e un fenotipo simile condiviso tra i membri della famiglia affetti (escludendo i casi di mutazione *de novo*);

Livelli di C-peptide (marker della produzione endogena di insulina) inferiori a 1 ng/ml ed assenza di auto-anticorpi (parametri utilizzati per escludere DM1);

Livelli di insulina spesso nella norma, benché il grado di iperglicemia risulti inappropriatamente basso: ciò suggerisce un precoce deficit della funzione delle cellule  $\beta$ ;

Condizioni di obesità o sovrappeso rare, non responsabili del suo sviluppo.

La tabella 3 mostra le caratteristiche cliniche tipiche delle tipologie di MODY più frequenti.

Tabella 3: Tabella riassuntiva dei quadri clinici in relazione ai vari sottotipi MODY (14)

| Sottotipo MODY e relativo gene        | Clinica  |
|---------------------------------------|--|
| <i>HNF4a</i> (MODY1)                  | Macrosomia, iperinsulinemia ipoglicemia transiente, iperlipidemia familiare, aumentata sensibilità alle sulfaniluree |
| <i>GCK</i> (MODY 2)                   | Lieve insufficienza insulinica, diabete mellito neonatale in caso di omozigosi                                       |
| <i>HNF1a</i> (MODY3)                  | Mancato sviluppo pancreas esocrino, aumentata sensibilità alle sulfaniluree, glicosinuria                            |
| <i>IPF1</i> (MODY4)                   | Agenesia del pancreas  |
| <i>HNF1<math>\beta</math></i> (MODY5) | Anomalie congenite del tratto urinario, agenesia pancreatica.  |
| <i>NEUROD1</i> (MODY6)                | Anomalie pancreatiche  |
| <i>KLF11</i> (MODY7)                  | Tumori pancreatici   |
| <i>CEL</i> (MODY8)                    | Mancato sviluppo pancreas end. Es.   |
| <i>PAX4</i> (MODY9)                   | Diabete mellito  |

E' inoltre possibile delineare le caratteristiche che possono aiutare nell'identificazione del MODY confrontandole con le più frequenti forme di DT1 e DT2 (Tab.4)

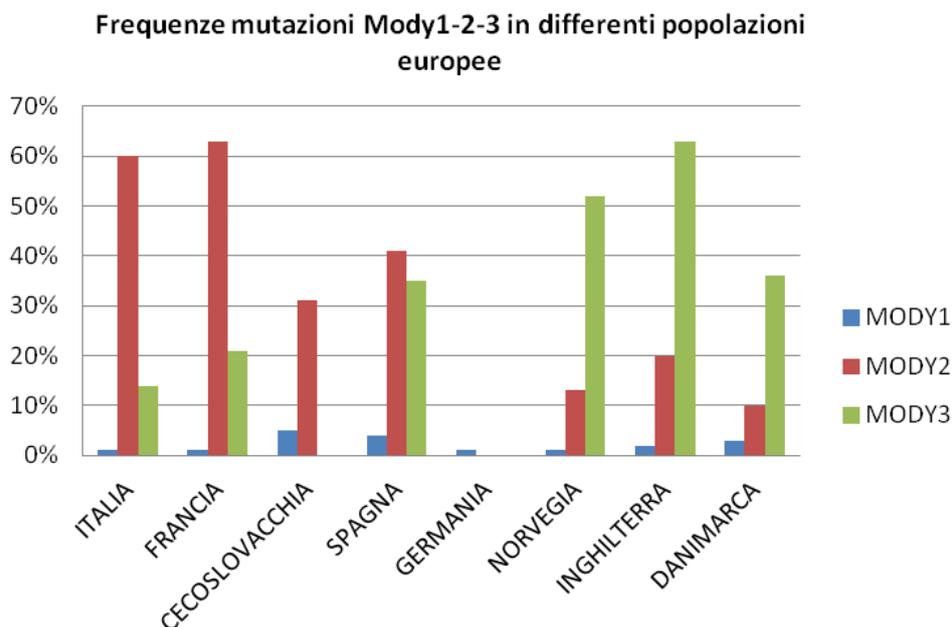
**Tabella 4:caratteristiche cliniche DT1, DT2, MODY**

|                            | <b>DT1</b>                  | <b>DT2</b>                                  | <b>MODY</b>                          |
|----------------------------|-----------------------------|---|--------------------------------------|
| <b>Frequenza</b>           | Comune                      | Aumentata                                   | 2-5% diabeti non-insulino dipendenti |
| <b>Genetica</b>            | Poligenica                  | Poligenica                                  | AD, monogenica                       |
| <b>Storia familiare</b>    | <15%                        | >50%  | 100%                                 |
| <b>Gruppo etnico</b>       | Varia                       | Asiatica, polinesiana, indigeni australiani | Varia                                |
| <b>Età esordio</b>         | Infanzia                    | Dopo la pubertà                             | <25 anni                             |
| <b>Gravità all'esordio</b> | Acuta e grave               | Lieve                                       | Lieve/asintomatica                   |
| <b>Chetoacidosi</b>        | Comune                      | Non comune                                  | Rara                                 |
| <b>Obesità</b>             | +/-                         | >90%  | +/-                                  |
| <b>Acantosi Nigricans</b>  | No                          | Comune                                      | Rara                                 |
| <b>Sindrome metabolica</b> | Assente                     | Comune                                      | Assente                              |
| <b>Autoimmunità</b>        | Positiva                    | Negativa                                    | Negativa                             |
| <b>Fisiopatologia</b>      | Distruzione cellule $\beta$ | Resistenza all'insulina e insulinopenia     | Disfunzione cellule $\beta$          |

## Frequenza

La frequenza dei casi MODY è di circa il 2-5%, e rappresenta circa 1-2% di casi di diabete in Europa. Le mutazioni più frequenti sono a carico dei geni *GCK* e *HNF1α* (14) (20) (21).

La fig.4 mostra le diverse frequenze di MODY nelle diverse popolazioni europee (22).



**Figura 4: Frequenze di MODY 1-2-3 in differenti popolazioni Europee (le percentuali sono relative al totale dei casi MODY diagnosticati nei diversi centri/nazioni) (22)**

La frequenza di MODY in Italia è di circa 100-150000 casi. Questi dati risultano essere sotto stimati a causa delle errate diagnosi in cui i casi MODY vengono scambiati con DT2.

Le mutazioni più frequenti sono a carico del gene *GCK* (40%), mentre sono rare quelle a carico del gene *HNF1α*(7%) le quali rappresentano la maggioranza nei paesi del nord Europa (Fig.4) (22).

### Casistica AOU Meyer

Nel laboratorio di Genetica Medica dell’Azienda Ospedaliera Universitaria “Anna Meyer” sono stati analizzati finora 266 pazienti pediatrici che presentavano caratteristiche cliniche generali riconducibili a diabete di tipo MODY. Le analisi sono state condotte mediante tecnica di sequenziamento Sanger.

L’analisi molecolare è stata effettuata per i geni *GCK*, *HNF1 $\alpha$* , *HNF4 $\alpha$* , *IPF1* e *HNF1 $\beta$*  mediante sequenziamento diretto delle regioni codificanti, dei tratti intronici fiancheggiati e delle regioni regolatrici.

L’analisi molecolare della coorte di pazienti ha evidenziato la presenza di mutazioni in geni responsabili di MODY in 116 casi, 44%(Fig.5a). Il 56% dei casi risultano essere negativi, in linea con i dati presenti in letteratura (Fig.5b).

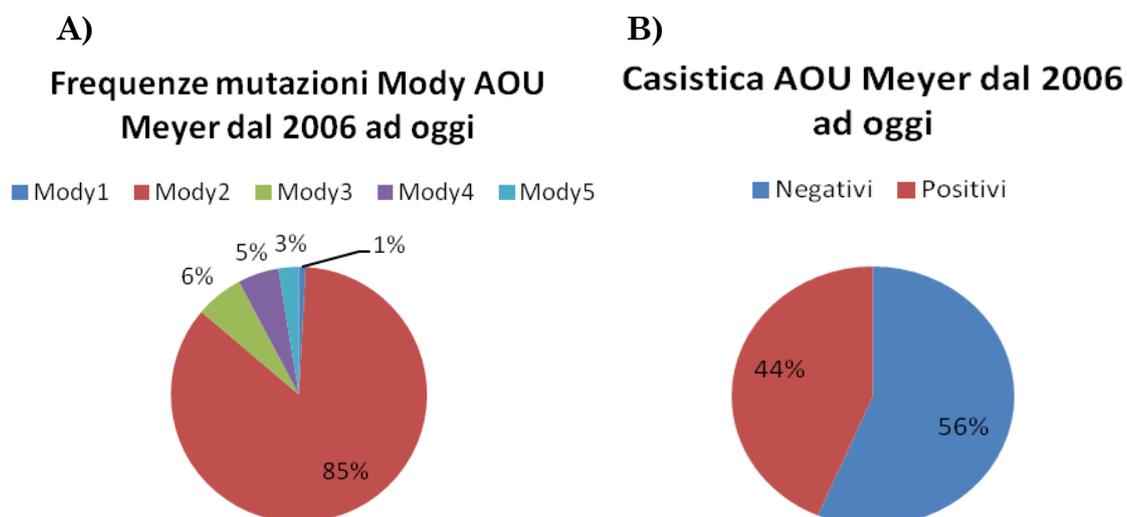
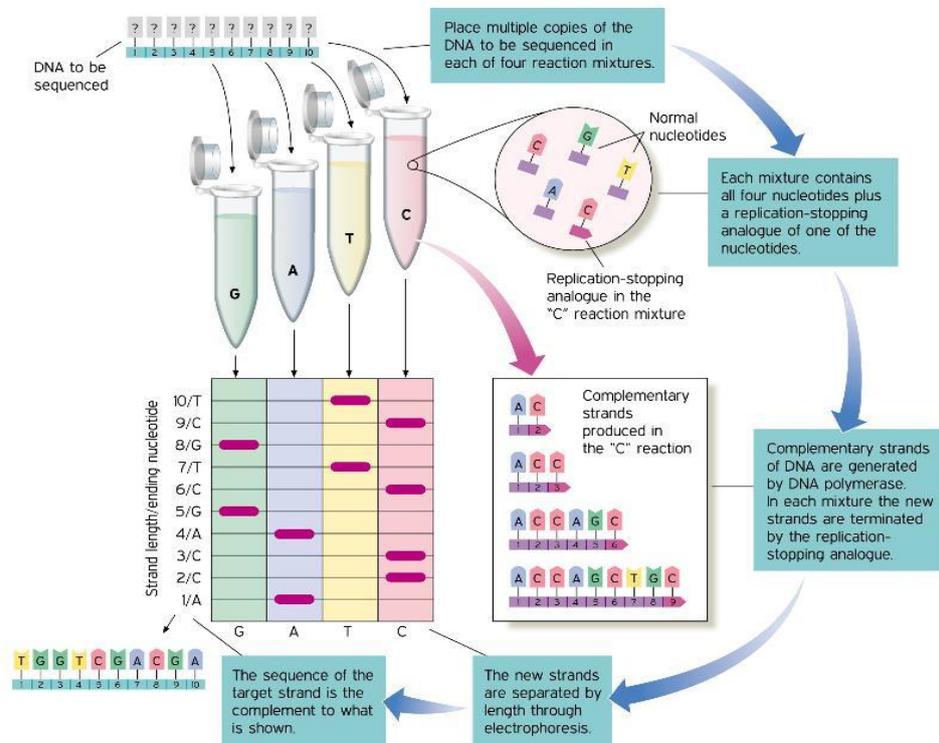


Figura 5: (A)Casi analizzati presso il laboratorio di Genetica Medica dell’ AOU Meyer Firenze dal 2006 a Marzo 2013. (B) Percentuale casi negativi e positivi analizzati fino a Marzo 2013.

## 1.4. Next Generation Sequencing (NGS)

Per oltre 30 anni la tecnologia ideata da Frederick Sanger è stata considerata il metodo “gold standard” per il sequenziamento DNA. Tale tipologia di sequenziamento nasce dall’unione della tecnica ideata da Allan Maxam, Walter Gilbert e Frederick Sanger (23) (24) (Fig.6).



**Figura 6: Rappresentazione schematica del metodo di sequenziamento Sanger.**  
 Il metodo che si basa sulla terminazione della catena. Questa tecnica prevede l'utilizzo dinucleotidi modificati (dideostrifosfato, ddNTPs) per interrompere la reazione di sintesi in posizioni specifiche lungo la sequenza. Il sequenziamento tramite il *metodo Sanger* risulta oggi giorno permette di sequenziare frammenti fino a 1000 basi e l'automazione ha reso possibile la corsa in parallelo di 384 reazioni contemporaneamente (24).

Da allora questa tecnologia ha subito molteplici cambiamenti mirati a migliorarne la processività e facilità di analisi, fino ad arrivare all'automazione.

La metodica Sanger automatizzata è ancora largamente utilizzata, è definita come la tecnologia di “prima generazione”.

Tale tecnologia è stata utilizzata per il primo sequenziamento completo del genoma umano concluso nel 2003, un progetto durato 13 anni costato circa 2.7 bilioni di dollari (25).

Da allora sono stati fatti numerosi passi avanti per migliorare le tecnologie di sequenziamento, volti ad aumentare la processività della tecnica ed abbattere i costi. Infatti qualche anno più tardi, nel 2008, il genoma è stato risequenziato in circa 5 mesi, al costo di 1,5 milioni di dollari (25).

Le nuove piattaforme NGS sono infatti caratterizzate dal sequenziamento massivo parallelo per l'amplificazione clonale di una singola molecola di DNA, le quali sono separate spazialmente in una *flow cell*. Tecnologia di gran lunga differente rispetto al metodo Sanger basato sulla separazione elettroforetica delle *chain-termination* (25).

Diverse sono le nuove tecnologie di sequenziamento, che negli ultimi anni sono entrate sul mercato, in grado di generare dati di 3- 4 ordini di grandezza, megabasi o gigabasi, maggiori rispetto al metodo Sanger (26).

Le tecnologie NGS rappresentano la metodica *High-Throughput* per il sequenziamento DNA, tanto da essere stato selezionato come metodo dell'anno nel 2007 da Nature Methods (27).

Le piattaforme in commercio maggiormente diffuse sono:

- Roche/454 Life Sciences,
- Illumina/Solexa,
- Applied Biosystems/Solid

Le caratteristiche principali che differenziano le varie piattaforme sono riassunte nella tabella 5.

**Tabella 5: caratteristiche piattaforme NGS (25).**

|                                      | <b>Roche 454 GS-FLX</b> | <b>Illumina Genome Analyzer</b> | <b>Applied Biosystems SOLiD</b> | <b>Sanger</b>       |
|--------------------------------------|-------------------------|---------------------------------|---------------------------------|---------------------|
| <b>Metodo di sequenziamento</b>      | Pirosequenziamento      | Terminatori reversibili marcati | Sequenziamento per ligazione    | Terminatori marcati |
| <b>Lunghezza frammenti (reads)</b>   | 400 bp                  | 36 bp                           | 35 bp                           | 800 bp              |
| <b>Tempo di corsa delle sequenze</b> | 10 h                    | 2.5 giorni                      | 6 giorni                        | 3 h                 |
| <b>Basi totali per corsa</b>         | A partire da 500 Mb     | 1.5 Gb                          | 4 Gb                            | 800 bp              |

### **Caratteristiche generali delle piattaforme NGS**

Le piattaforme NGS condividono tecniche comuni di sequenziamento massivo, ad alta processività, di milioni di paia di basi di DNA in un singolo esperimento. Le fasi principali sono (28):

- preparazione dei templati
- sequenziamento
- imaging
- analisi dei dati

Le librerie di sequenziamento del DNA sono amplificate in vitro in maniera clonale, in minor tempo e ridotta laboriosità rispetto

alla clonazione di una genoteca clonale batterica. I metodi attuali per ottenere una library, generalmente prevedono la nebulizzazione del DNA genomico in frammenti più piccoli da cui vengono creati i *templates*. Una caratteristica comune delle tecnologie NGS è l'immobilizzazione su una superficie, o supporto solido, dei *templates*. L'immobilizzazione dei *templates* e la loro separazione fisica, permette di eseguire migliaia di miliardi di reazioni di sequenziamento contemporaneamente (29) (28).

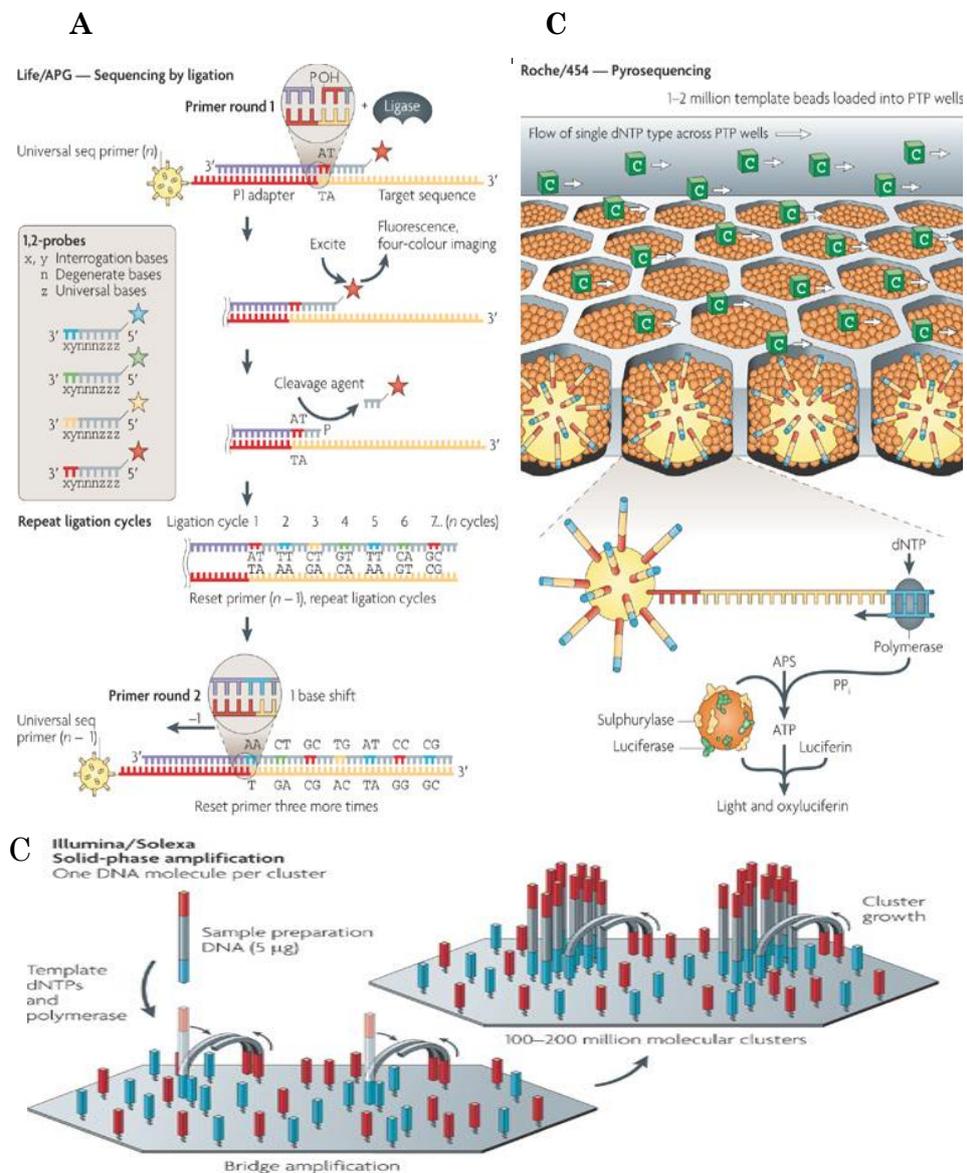


Figura 7: Strategie di immobilizzazione e sequenziamento delle varie piattaforme NGS. (A) Life/APG (B) Illumina/Solera (C) Roche 454.

Il sequenziamento con la NGS avviene per sintesi, ovvero la sequenza di DNA viene prodotta aggiungendo nucleotidi al filamento complementare, piuttosto che attraverso la chimica della terminazione della catena.

Ciò che varia maggiormente tra le diverse piattaforme è la modalità di sequenziamento.

Le strategie di sequenziamento maggiormente utilizzate sono:

- terminazione ciclica reversibile (CRT)
- sequenziamento per ligazione
- pirosequenziamento

La piattaforma utilizzata all'interno del laboratorio di Genetica Medica dell'ospedale pediatrico "A. Meyer" per gli esperimenti di sequenziamento, è la Roche/454 life sciences.

### **Roche/454 life science**

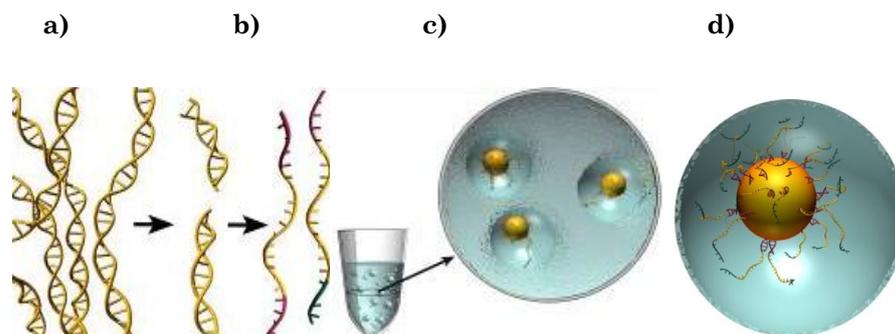
La tecnologia 454 è derivata dall'unione della PCR in emulsione e il pirosequenziamento. Nel 1993, Nyren et al. descrivono una metodica basata sul rilevamento dei gruppi fosfato, rilasciati a seguito dell'incorporazione di ogni nuovo nucleotide (dNTP), durante la fase di sintesi del filamento complementare a quello da sequenziare(25).

Contemporaneamente Tawfik e Griffiths descrivono l'amplificazione PCR, di una singola molecola di DNA, all'interno di micro compartimenti formati dall'emulsione dell'acqua nell'olio (25). La piattaforma GS 20 entra sul mercato nel 2005. La

combinazione delle due tecnologie ha permesso di completare il sequenziamento del genoma *Mycoplasma genitalia* con il 96% di copertura e il 99.96% di accuratezza in una singola GC 20 *run*. Successivamente la 454 Life Science viene acquisita dalla Roche Applied Science, la quale propone una nuova versione dello strumento, GS FLX, inserendo l'utilizzo di un vetrino in fibra ottica, il *PicoTiter*, per lo step di sequenziamento (25).

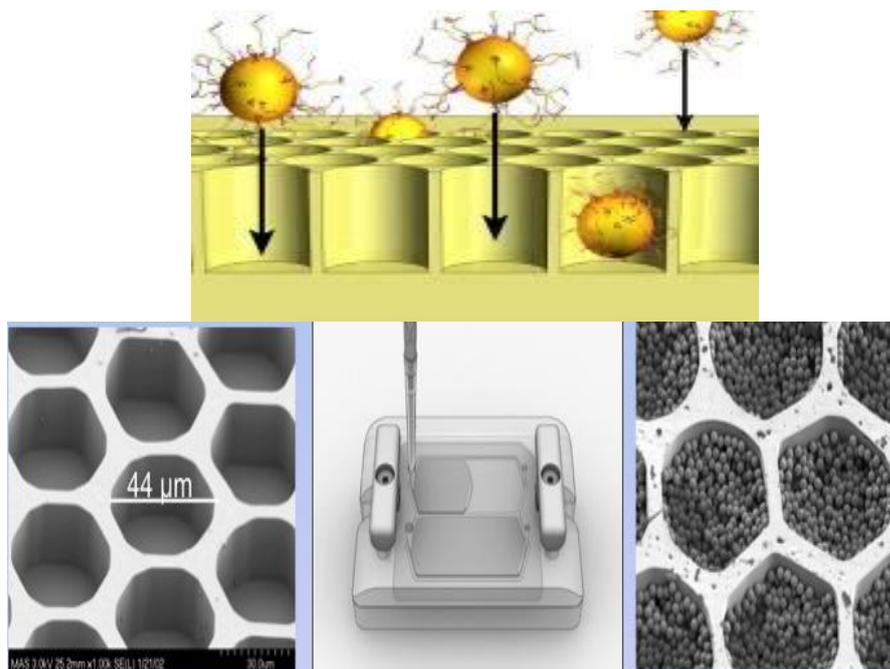
La tecnologia Roche 454 consente l'utilizzo di vari campioni di partenza, tra cui: DNA genomico, prodotti di PCR, BACs e cDNA.

Questa tecnica prevede la creazione di una libreria di DNA attraverso la nebulizzazione, la quale produce tanti frammenti di DNA genomico più piccoli, nell'ordine delle 300-800bp, i quali vengono amplificati attraverso una normale PCR. Ad ogni frammento di DNA, successivamente, vengono aggiunti dei piccoli adattatori oligonucleotidici, i quali serviranno per gli *steps* di purificazione, amplificazione e sequenziamento. I frammenti generati a singolo filamento (sstDNA) contenenti gli adattatori rappresentano le librerie. Le librerie così ottenute vengono poi denaturate ed ibridate con delle biglie, contenenti le sequenze complementari agli adattatori oligonucleotidici. Tali adattatori serviranno da supporto per l'amplificazione mediante PCR in emulsione (emPCR). La emPCR avviene all'interno di moltissimi microreattori creati dall'emulsione acquosa, in ogni micro reattore avviene un'unica reazione di PCR, come mostrato nella Fig.8 (29).



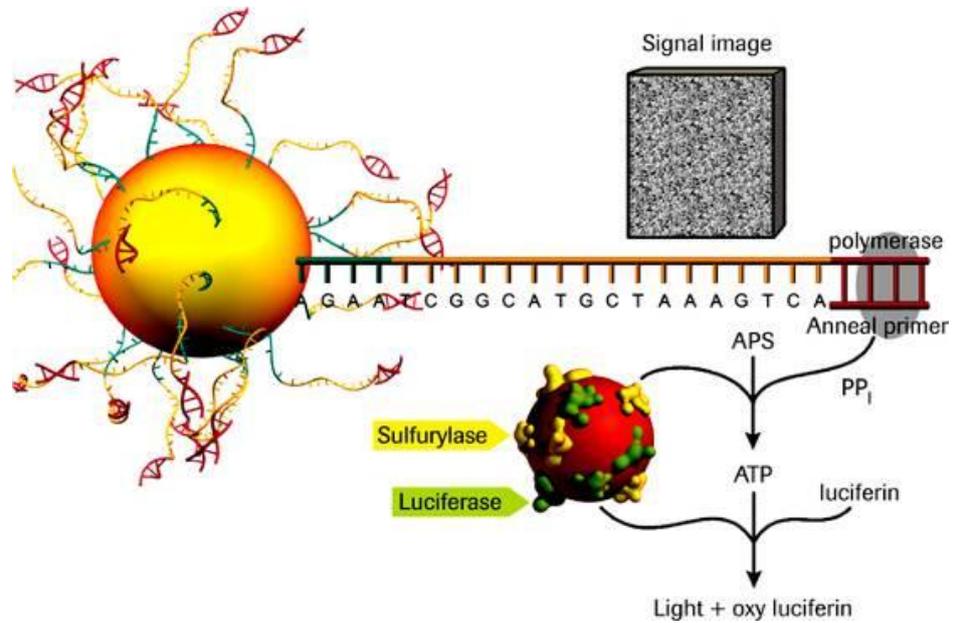
**Figura 8: Step di preparazione emPCR per il sequenziamento Roche 454 i quali prevedono la creazione di frammenti di DNA (a), ai quali vengono aggiunti degli adattatori (b). I frammenti vengono addizionati a delle biglie (c) La reazione di PCR avviene all'interno di microreattori creati dall'emulsione acquosa (d).**

Il DNA legato alle biglie preamplificato viene trasferito nei pozzetti della piastra, 454 *PicoTiterPlatet<sup>TM</sup>*, con gli enzimi necessari per produrre luce in presenza di ATP. La piastra, grande 70mm x 75mm è composta da fibra ottica, e contiene 1.6milioni di pozzetti dal diametro di 44µm ciascuno; ognuno di questi pozzetti può contenere solamente una libreria legata alla propria biglia, le quali vi rimangono all'interno per tutta la durata del sequenziamento (Fig.9). Questo meccanismo permette la massimizzazione del processo, e ne facilita la lettura.



**Figura 9: Caricamento delle microparticelle contenenti DNA amplificato nei singoli pozzetti del PicoTiterPlate (PTP)**

La metodica di sequenziamento adottata dalla Roche 454 è il pirosequenziamento. Il pirosequenziamento è una metodica basata sul rilevamento dei gruppi fosfato rilasciati dall'incorporazione di ogni nuovo nucleotide durante la sintesi del filamento complementare a quello da sequenziare. Questa tecnica prevede che il DNA a singolo filamento sia legato a delle biglie e amplificato attraverso una PCR in emulsione, consentendo quindi la creazione di moltissime microcelle di PCR in soluzione acquosa separate da olio. Il DNA legato alle biglie preamplificato, viene trasferito in pozzetti di una piastra con fibre ottiche e gli enzimi necessari per produrre luce in presenza di ATP. La chemiluminescenza è generata quando i nucleotidi si legano alla base complementare e rilasciano il pirofosfato. Aggiungendo i quattro nucleotidi uno alla volta in sequenza (A, C, G, T) è possibile ricostruire la sequenza fino a 250-400 nucleotidi (Fig.10) (25) (28) (30).



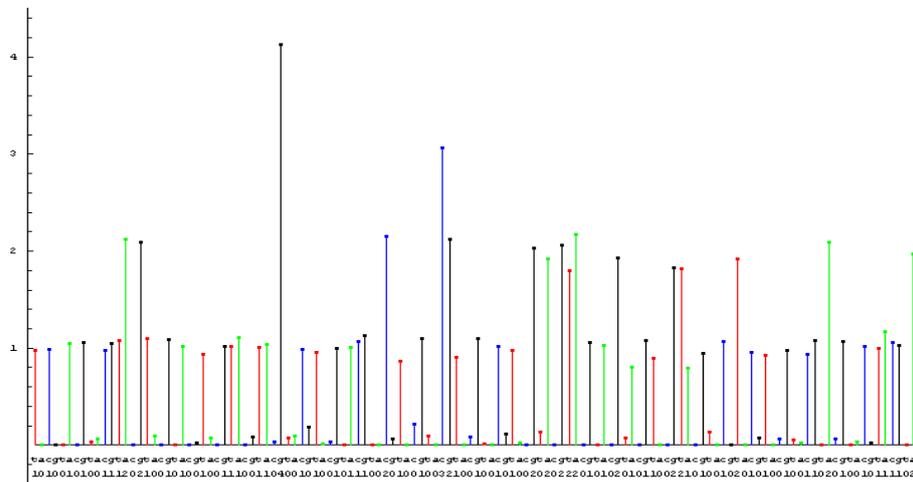
**Figura10: Pirosequenziamento utilizzando la piattaforma Titanium Roche/454 .**  
 Dopo aver caricato le microparticelle contenenti DNA amplificato nei singoli pozzetti del PicoTiterPlate (PTP), vengono aggiunte biglie contenenti solforilasi e luciferasi.

Si ottengono così le cosiddette *reads*, ovvero milioni di brevi frammenti di sequenza, che vengono allineati ad una sequenza di riferimento conosciuta o assemblati *de novo* (25) (28) (30).

Il sistema Roche 454 ha un'accuratezza di ogni singola lettura del 99.5%, gli errori di sostituzione sono infatti estremamente rari. Inoltre il punto di forza di questa tecnologia rispetto alle altre piattaforme è l'uso, per l'analisi dei dati, di algoritmi i quali filtrano in maniera indipendente dalla presenza o meno di sequenze di riferimento (30).

## Elaborazione dei dati

L'enorme quantità di dati di sequenza generata dal campione di un singolo paziente crea nuove problematiche all'interno di un laboratorio, in quanto richiede significativi investimenti in infrastrutture bioinformatiche e in personale qualificato. Anche se ogni piattaforma di sequenziamento ha un metodo unico di trattamento dei dati, vengono comunque impiegate strategie simili per la trasformazione di dati 'grezzi' delle sequenze in una forma interpretabile.



**Figura 11: Rappresentazione dei dati mediante il flogramma 454GS FLX Titanium (Roche)**

Poiché avvengono milioni di reazioni di sequenziamento in parallelo, bisogna analizzare le metriche di performance globali della singola corsa per assicurarsi che lo strumento sia calibrato. Per fare questo, molti strumenti di sequenziamento NGS includono sequenze di controllo standard all'interno della corsa. Successivamente, ogni *read* sequenziata individualmente deve essere sottoposta ad una valutazione di qualità specifica per il sequenziamento. Per esempio, per quanto riguarda il sequenziamento con 454 sono stati sviluppati appositi algoritmi per definire i criteri per l'identificazione di delezioni o errori di

inserzione, che si verificano nelle regioni omopolimero durante pirosequenziamento (31).

Dopo il controllo di qualità, la sequenza genomica deve essere "ri-generata" attraverso l'allineamento con un genoma di riferimento o con un assemblaggio *de novo*. Sono stati sviluppati una grande varietà di metodi di calcolo per eseguire l'allineamento efficace dei dati di sequenza *short-read* con un genoma di riferimento. Le due strategie più comuni sono convertire i dati di sequenza (o il genoma di riferimento) in una serie unica di valori interi o eseguire una 'trasformata di Burrows-Wheeler' per costruire una matrice di tutte le rotazioni possibili di una data sequenza (32).

Con la tecnologia di Sanger, il frammento di sequenza ottenuto permette l'assemblaggio della sequenza in base al grado di sovrapposizione tra i frammenti sequenziati. Tuttavia, questo approccio non è realizzabile per le *reads* prodotte dai sistemi di sequenziamento *next generation*. Per risolvere questo problema, sono stati sviluppati nuovi algoritmi che analizzano i dati di sequenza come piccole sotto-sequenze a lunghezza fissa (33).

Con l'avvento di *reads* più lunghe nelle tecnologie NGS ad alta produttività, la potenza di calcolo necessaria per eseguire il l'assemblaggio di un genoma *de novo* probabilmente diminuirà con un miglioramento concomitante nel rilevamento delle varianti (34).

## **Applicazioni NGS**

Le applicazioni delle tecnologie NGS sono molteplici (35) (25):

- Target resequencing
- Sequenziamento e comparazione dei genomi batterici
- Sequenziamento degli smallRNAs
- Paleobiologia e studio del DNA preistorico
- Metagenomica e malattie infettive
- Struttura del genoma
- Sequenziamento trascrittoma
- Analisi dei siti di legame delle proteine al DNA ed alla cromatina

## 2. Scopo della tesi

---

La diagnosi genetica per il diabete monofattoriale (DM), che rappresenta circa il 2% delle così dette “forme rare” di diabete, è ormai diventata una procedura standard che permette non solo di caratterizzare individui affetti, ma anche di identificare familiari a rischio, al fine di indirizzarli verso una terapia personalizzata e consentire una diagnosi differenziale dal diabete di tipo 2 (DT2).

In questi ultimi 10 anni si è assistito a un’esplosione di nuove informazioni circa i meccanismi che regolano la secrezione di insulina e le basi genetiche della malattia, enfatizzando l’importanza di come alterazioni delle funzioni delle cellule  $\beta$  del pancreas siano determinanti nello sviluppo sia delle forme monofattoriali che poligeniche del diabete.

L’obiettivo di questo lavoro è stato quindi quello di individuare e caratterizzare da un punto di vista genetico pazienti diabetici con un quadro clinico attribuibile a MODY.

Mediante la tecnologia NGS 454GS FLX Titanium (Roche), già presente nella AOU Meyer, abbiamo analizzato un gruppo di 14 pazienti con sospetto clinico MODY risultati negativi alla prima analisi di sequenza dei geni MODY noti mediante sequenziamento Sanger. Per l’allestimento dell’esperimento sono stati selezionati un set di 103 geni: geni noti per DM, geni del *pathway* delle cellule  $\beta$  pancreatiche, geni candidati per il DT2 e geni causativi di diabete in modelli murini.

### 3. Materiali e metodi

---

### *3.1. Casistica analizzata*

L'indagine molecolare per il MODY1, 2, 3, 4 e 5 viene effettuata nel Centro Ospedaliero Meyer su campioni di DNA provenienti da leucociti periferici di pazienti con sospetto MODY. I pazienti provengono principalmente dall'Unità di Diabetologia dell'Ospedale Pediatrico "A. Meyer" e da altre strutture ospedaliere italiane, in cui lo studio genetico della patologia, non viene eseguito.

L'esecuzione dell'indagine genetica richiede l'espressione, con apposito modulo firmato, del proprio consenso informato. Le informazioni riguardanti i pazienti sono state raccolte in cartelle cliniche personali, in cui sono stati riportati i dati anamnestici, i risultati delle analisi mediche effettuate e l'albero genealogico. I campioni sono stati anonimizzati e contrassegnati con codici progressivi.

### *3.2. Estrazione del DNA da sangue periferico*

L'estrazione del DNA è stata effettuata su leucociti da sangue periferico utilizzando il kit QIAamp DNA Mini kit (Qiagen). In un tubo da microcentrifuga da 1.5 ml sono stati trasferiti 20 µl di Proteasi Qiagen (o Proteinasi K), 200 µl di campione e 200 µl di Buffer AL, la mix così preparata è stata passata 15s al vortex e quindi incubata a 56° C per 10'. A questo punto sono stati aggiunti 200 µl di etanolo (96-100%) e la mix è stata trasferita in una colonnina da microestrazione QIAamp spin e sottoposta a centrifugazione a 8000 rpm per 1', dopo l'eliminazione del

materiale di scarto sono stati aggiunti 500  $\mu$ l di Buffer AW1 e nuovamente centrifugato a 8000 rpm per 1', quindi lo stesso passaggio è stato ripetuto utilizzando 500  $\mu$ l di Buffer AW2 e centrifugando a 14000 rpm per 3'. La colonnina da microestrazione QIAamp spin è stata inserita in un nuovo tubo da microcentrifuga sterile, il filtro contenete il DNA è stato idratato con 150  $\mu$ l di Buffer AE e dopo una incubazione di 1' a temperatura ambiente centrifugato a 8000 rpm per 1' per eluire; il DNA così ottenuto è stato conservato a -20 °C fino al momento dell'utilizzo.

### 3.3. *Determinazione della concentrazione del DNA*

La concentrazione e la qualità del DNA ottenuto dal processo di estrazione, sono state valutate attraverso l'uso dello strumento Nanodrop<sup>™</sup> 2000 (Thermo scientific), uno spettrofotometro UV-visibile che consente la misurazione accurata e rapida di piccole quantità di campione (0.5  $\mu$ l - 2.0  $\mu$ l), sfruttando la tensione superficiale per mantenere il campione tra due fibre ottiche. Per ogni campione è stato posizionato 1  $\mu$ l direttamente sopra la superficie di rilevamento della fibra ottica. Abbassando il braccio dello strumento, un secondo cavo ottico viene messo in contatto col primo. Si crea così una colonna di campione a ponte tra le due fibre ottiche. Una lampada allo Xeno fornisce la sorgente luminosa e uno spettrometro analizza la luce che ha attraversato il campione. Lo strumento fornisce tre valori: dall'assorbanza a 260 nm (lunghezza d'onda di maggiore assorbimento del DNA), sulla base della legge di *Lambert-Beer*, si ottiene la

concentrazione del DNA del campione in analisi, dal rapporto A260/A280 il suo grado di purezza (a 280 nm si ha il maggior assorbimento delle proteine) e dal rapporto A260/A230 l'eventuale contaminazione da reagenti (l'assorbanza a 230 nm riflette contaminazioni dovute a sostanze come carboidrati, fenoli, peptidi o composti aromatici).

### 3.4. *Analisi molecolare dei geni implicati nel MODY*

L'analisi molecolare dei geni *HNF4 $\alpha$* , *GCK*, *HNF1 $\alpha$* , *IPF1* e *HNF1 $\beta$*  è stata effettuata mediante amplificazione degli esoni e delle regioni introniche fiancheggianti. I primers utilizzati per la PCR sono stati disegnati utilizzando il programma Primer3 (primer3\_www.cgi)

**Tabella 6: Primers impiegati per l'amplificazione del gene *HNF4 $\alpha$*  (MODY 1)**

| PRIMER FORWARD                | ESONE        | PRIMER REVERSE                |
|-------------------------------|--------------|-------------------------------|
| 5'- CGCCTGTATGCAACTCC -3'     | <b>P2-1D</b> | 5'- CAAAGCTGACCGCAGTC -3'     |
| 5'- CAAGGATCCAGAAGATTGGC -3'  | <b>P1</b>    | 5'- GGTCCAGTGACAGCACTGTAG -3' |
| 5'- TCATGATGCCTGCCTTGAC -3'   | <b>1A-1C</b> | 5'- CTCTCTGAACCTAGGCCCAA -3'  |
| 5'- CATGCGGACCTGTTGGAGTG -3'  | <b>1B</b>    | 5'- CACAAGATTCCAGACCTCAG -3'  |
| 5'- AGGTGATGGAGTGGGAACAG -3'  | <b>2</b>     | 5'-TGTCACTGGGACAGTCCTCC-3'    |
| 5'- CGGGATGAAGAGATGAGAGC -3'  | <b>3</b>     | 5'- TCTGGGACCTACCCACTCAG -3'  |
| 5'- GGGCAGTTGTGATATGTGA -3'   | <b>4</b>     | 5'- CAAACTGGGCCATGTGAAAC -3'  |
| 5'- ACCCTGAGCTTCCTCAGAG -3'   | <b>5</b>     | 5'- CCACGGCTATATCCAGGT -3'    |
| 5'- TCAGGCAGGTAGAGGCAGAG -3'  | <b>6</b>     | 5'- CATGCCACCATGTGAATCTC -3'  |
| 5'- CCACAGGCACCAGCTATCTT 3'   | <b>7</b>     | 5'- CCTGGCTGTCGTTAACATGA -3'  |
| 5'- TTGTTGAGGTCCCTGAATCCTT-3' | <b>8</b>     | 5'ACAGATATCCACGCATCCATACA -3' |
| 5'- TATTGGATGGGCTGGTTGAT-3'   | <b>9</b>     | 5'-ACCCTGGAACCCAGAAAACT-3'    |
| 5'-AGGGTGGGAGGGGAGAA-3'       | <b>10</b>    | 5'-GTTCTCGGGCCCTTCATC-3'      |

**Tabella 7: Primers impiegati per l'amplificazione del gene *GCK* (*MODY 2*)**

| PRIMER FORWARD             | ESONE | PRIMER REVERSE                |
|----------------------------|-------|-------------------------------|
| 5'-GAAGGACACTAAGCCCCACA-3' | 1A    | 5'-GGCAAGCAAACACTCCCAGAATG-3' |
| 5'-GGCTGAGGAATGGAAGATGG-3' | 1B    | 5'-TCTCCAGTGCAAAGTCCCTAAC-3'  |
| 5'-TCCCCTGCTATCCAACCTGT-3' | 1C    | 5'-GCGTCAGGCTGTCTAACA-3'      |
| 5'-CGGGGTCAGAAGACAGAAGG-3' | 2     | 5'-TGGCTGTGAGTCTGGGAGTG-3'    |
| 5'-GCTCAGTCACCTGGGGCC-3'   | 3     | 5'-ACAGGTGGCACCTCCCGTC-3'     |
| 5'-AGTGGCCAGGTGTTGCAGTG-3' | 4     | 5'-GAAGGCAGAGTTCCTCTGGG-3'    |
| 5'-CAGGAGGTAGTGACAGGCC-3'  | 5     | 5'-AAGGAGAAAGGCAGGCAGTG-3'    |
| 5'-CAGCACTGCCTGCCTTTCTC-3' | 6     | 5'-CCTGTTGTACACAGGGAGCC-3'    |
| 5'-GGCAGGAACCAGGCCCTAC-3'  | 7     | 5'-CTCCCATCTGCCGTGCACC-3'     |
| 5'-CCC GGCTTCCACCTGCATG-3' | 8     | 5'-GCCCTGAGACCAAGTCTGC-3'     |
| 5'-CTGTGCGGAGCGACACTCAG-3' | 9     | 5'-GGTTGGGAACCGCAAGGAAC-3'    |
| 5'-GTCGACTGCGTGCAGGGCGC-3' | 10    | 5'-TGTGGCATCCTCCCTGCGCT-3'    |

**Tabella 8: Primers impiegati per l'amplificazione del gene *HNF1a* (*MODY 3*)**

| PRIMER FORWARD             | ESONE | PRIMER REVERSE             |
|----------------------------|-------|----------------------------|
| 5'-AGAGCCTCGATTTCCTCCTC-3' | Pr.   | 5'-CCTAAACCCCACTAGCCAAG-3' |
| 5'-AGGCAAACGCAACCCAC-3'    | 1     | 5'-GGCTCGTTAGGAGCTGAGG-3'  |
| 5'-ACAAGGTTCCAGCACCCAG-3'  | 2     | 5'-TGGCTGTGAGTCTGGGAGTG-3' |
| 5'-AATCAAGGGCAAGTCAAGG-3'  | 3     | 5'-CAGGACCTAGAGTCACTTCC-3' |
| 5'-AGGACAGGGTTCCTTGAGC-3'  | 4     | 5'-TGTCAGTGGGACAGTCTCC-3'  |
| 5'-TTTGAAGTGCTGAGGGCTG-3'  | 5     | 5'-CCAGAATCTCCCTGCCAAG-3'  |
| 5'-CTAGGGAGGCCCTGTGG-3'    | 6     | 5'-CTCTTCCAGCTCCTGGATTG-3' |
| 5'-GGGTGGGATATAACTGGGG-3'  | 7     | 5'-ATCCATTGACAGCCAACCTC-3' |
| 5'-CTGTCAGGCACGTCTGC-3'    | 8     | 5'-GGGTGAGGGGCTCTGTC-3'    |
| 5'-CCTGTGACAGAGCCCTC-3'    | 9     | 5'-GGTGGACCAACATGGAAG-3'   |
| 5'-CTAGGGACAGGCAGGTGG-3'   | 10    | 5'-CTGCTGCCCTCATCACC-3'    |

**Tabella 9: Primers impiegati per l'amplificazione del gene *IPF1*(MODY 4)**

| PRIMER FORWARD                | ESONE | PRIMER REVERSE                |
|-------------------------------|-------|-------------------------------|
| 5'- AACGCCACACAGTGCCAAATC -3' | 1     | 5'- TTAGTCCGACCCGGGATAATC -3' |
| 5'- GCCTACACTAGGCGCTGAAA -3'  | 2     | 5'- GTCCCTCTAGGCCCTCTGCT -3'  |

**Tabella 10: Primers impiegati per l'amplificazione del gene *HNF1β* (MODY5)**

| PRIMER FORWARD                | ESONE | PRIMER REVERSE               |
|-------------------------------|-------|------------------------------|
| 5'-GGGTGGAGGGGTTCTGGAT-3'     | 1     | 5'-CGGGCGCAGTGTCACTCAGG-3'   |
| 5'-CTCCCACTAGTACCCTAACCC-3'   | 2     | 5'-GAGAGGGCAAGGTCACTTCAG -3' |
| 5'-AGTGAAGGCTACAGACCCCTATC-3' | 3     | 5'-TTCTGGGTCTGTGTACTTGC-3'   |
| 5'-TGTGTTTTTGGCCAAGCACCA-3'   | 4     | 5'-AACCAGATAAGATCCGTGGC-3'   |
| 5'-TGCCGAGTCATTGTTCAGG-3'     | 5     | 5'-CCTCTTATCTTATCAGCTCCAG-3' |
| 5'-CTGCTCTTTGTGGTCCAAGTCC-3'  | 6     | 5'-GAGTTTGAAGGAGACCTACAG-3'  |
| 5'-ATCCACCTCTCCTTATCCAG-3'    | 7     | 5'-ACTTCCGAGAAAGTTCAGACC-3'  |
| 5'-TTTGCTGTGTATGCACCTTG-3'    | 8     | 5'-GCCGAGTCCATGCTTGCCAC-3'   |
| 5'-CTTGCTGGTTGAGTTGGGC -3'    | 9     | 5'-TTCCATGACAGCTGCCCAGAG-3'  |

L'amplificazione è stata effettuata secondo il protocollo del manuale *HotStarTaq® PCR* in un volume finale di 25 µl contenente.

|                             |         |
|-----------------------------|---------|
| HotStar Taq (5 U/µl)        | 0,1 µl  |
| PCR Buffer 10X              | 2,5 µl  |
| dNTP (1.25 mM)              | 2 µl    |
| DNA stampo (50-100 ng/µl)   | 1 µl    |
| Primer Forward (10 pmol/µl) | 1 µl    |
| Primer Reverse (10 pmol/µl) | 1 µl    |
| Acqua sterile               | 17,4 µl |

**Il ciclo di PCR ha previsto:**

|                   |          |          |
|-------------------|----------|----------|
|                   | 95 15'   |          |
| Denaturazione     | 94°C 1'  | 30 cicli |
| <i>Annealing</i>  | * °C 1'  |          |
| Estensione        | 72°C 1'  |          |
| Estensione finale | 72°C 10' |          |

Con \* si indica la *T* di annealing specifica per ogni coppia di primers.

Tutte le reazioni di PCR sono state effettuate in un termociclatore *GeneAmp® PCR System 9700* (Applied Biosystems).

### 3.5. *Elettroforesi su gel di agarosio*

Il risultato dell'amplificazione PCR dei frammenti è stato verificato mediante elettroforesi su gel di agarosio al 2% (Agarosio standard low EEO, AB ANALITICA), in TBE 0.5%, a 140V per 15'. Nella preparazione del gel di agarosio sono stati aggiunti 6µl/100ml di Gel Red™ (Biotium, Inc.), una molecola intercalante del DNA che consente la visualizzazione dei prodotti PCR su un transilluminatore (Genenco). Un marker di peso molecolare noto è stato utilizzato per determinare la corretta dimensione dei campioni.

### 3.6. *Purificazione dei prodotti di PCR*

Il metodo di purificazione utilizzato per questo studio è quello enzimatico (EXOSAP-IT<sup>®</sup>, USB), che utilizza la combinazione di due enzimi: esonucleasi I e fosfatasi alcalina. Questa strategia permette la purificazione di frammenti di PCR anche inferiori a 100 bp e grandi fino a 20 kb, senza significativa perdita di campione.

In una microprovetta da 0.2 ml contenente 5 µl di prodotto di PCR sono stati aggiunti 2 µl di EXOSAP-IT. Il campione è stato collocato in un termociclatore *GeneAmp<sup>®</sup> PCR System 9700* (Applied Biosystems) con un ciclo di purificazione che ha previsto un'incubazione a 37°C per 15' per la degradazione dei primers e dei nucleotidi in eccesso, seguita da un'inattivazione enzimatica a 80°C per 15'.

### 3.7. *Reazione di sequenza*

Il protocollo di lavoro utilizzato, è basato sull'impiego del kit BigDye<sup>®</sup>Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems) contenente BigDye<sup>®</sup>Terminator v1.1 5X Sequencing Buffer e BigDye<sup>®</sup>Terminator v1.1 Cycle Sequencing RR-100, in cui sono inclusi un AmpliTaq<sup>®</sup> DNA Polymerase, BigDye terminators marcati con dRodamina e altri componenti necessari per la reazione di sequenza.

*L'AmpliTaq<sup>®</sup> DNA Polimerasi* è una variante della Taq polimerasi con una mutazione nel sito attivo che le permette una minore discriminazione nei confronti dei dideossinucleotidi.

In una microprovetta è stata preparata la reazione di sequenza 1X in un volume finale di 20  $\mu\text{l}$  con il seguente protocollo:

|   |                  |
|---|------------------|
| BigDye <sup>®</sup> Terminator v1.1 Cycle Requencing RR-100 | 2 $\mu\text{l}$  |
| BigDye <sup>®</sup> Terminator v1.1 5X Sequencing Buffer    | 2 $\mu\text{l}$  |
| Prodotto PCR purificato                                     | 2 $\mu\text{l}$  |
| Primer (3.2 pmoli/ $\mu\text{l}$ )                          | 1 $\mu\text{l}$  |
| Acqua sterile   | 13 $\mu\text{l}$ |

La microprovetta è stata inserita nel termociclatore *GeneAmp<sup>®</sup>PCR System 9700* (Applied Biosystems) con il ciclo di sequenza: 96°C per 10", 50°C per 5", 60°C per 2' (24 cicli).

### 3.8. *Purificazione della reazione di sequenza*

La purificazione dei campioni dopo la reazione di sequenza è stata ottenuta utilizzando una piastra da 96 pozzetti (*MultiScreen<sup>®</sup>HTS HV*, Millipore) che è stata preventivamente preparata. L'allestimento della piastra ha previsto il trasferimento in ogni pozzetto di *Sephadex G-50 Superfine* (GE Healthcare Bio-Sciences AB), un gel di destrano a legami crociati che lega selettivamente i nucleotidi terminatori marcati e ne permette la separazione dai prodotti di sequenza. In ciascun pozzetto della piastra sono stati aggiunti 300  $\mu\text{l}$  di H<sub>2</sub>O, e la piastra è stata quindi incubata per 180' a temperatura ambiente per permetterne l'idratazione. La successiva centrifugazione a 2500 rpm per 12' consentiva la raccolta del materiale liquido in una piastra da 96 pozzetti con fondo a V (*96 Well V-bottom plate*, Millipore) e quindi la sua eliminazione.

In ogni pozzetto della piastra così preparata sono stati trasferiti 20 µl della reazione di sequenza e la piastra è stata centrifugata, al fine di recuperare la sequenza purificata nella piastra di raccolta, a 2500 rpm per 4'. I campioni, una volta purificati, sono stati preparati diluendo 5 µl di reazione di sequenza in 15 µl di *Hi-Di™ Formamide* (Applied Biosystem) e caricati sul sequenziatore *ABI Prism 3130 Genetic Analyser*.

### 3.9. Interpretazione dei risultati

Il sistema utilizzato si basa sull'elettroforesi capillare (CE) (*ABI Prism 3130 Genetic Analyser*, Applied Biosystems). Il sequenziatore *ABI Prism 3130* è sotto il controllo di specifici programmi software: *ABI 3130 data collection software* e *ABIPRISM DNA Sequencing Analysis software*. A questi è attribuita la funzione di controllare i parametri elettroforetici, oltre che di gestire l'analisi e l'elaborazione dei dati. I dati di sequenza possono essere presentati sotto forma di bande o picchi cromatografici a quattro colori, oppure ancora sotto forma di sequenza, con le lettere ACGT che indicano le singole basi identificate.

L'interpretazione delle varianti identificate è stata effettuata mediante l'utilizzo dei siti di predizione e database disponibili online:

- (<http://sift.jcvi.org>)
- (<http://www.ncbi.nlm.nih.gov/Omim>)
- (<http://www.hgmd.cf.ac.uk/ac/index.php>)
- (<http://www.ensembl.org/index.html>)
- (<http://genetics.bwh.harvard.edu/pph2/>)
- (<http://mmb2.pcb.ub.es:8080/PMut/>)

### 3.10. *Analisi molecolare mediante sistema Roche/454*

Come precedentemente accennato, questa tecnica prevede 6 fasi principali:

- Selezione geni target
- Preparazione *library*
- Ibridazione
- emPCR: amplificazione clonale in emulsione
- sequenziamento con *GS FLX Titanium 454 Roche*
- Analisi dei dati

#### **Selezione geni target**

Per l'allestimento dell'esperimento sono stati selezionati un set di 103 geni: geni noti per DM, geni del *pathway* delle cellule  $\beta$  pancreatiche e geni candidati per il DT2; inoltre abbiamo selezionato geni causativi di diabete in esperimenti condotti su topi. Alla luce degli studi effettuati sono pertanto stati selezionati i seguenti geni per il vetrino in fase liquida: *HNF4A, GCK, HNF1A, PDX1, HNF1B, NEUROD1, KLF11, CEL, PAX4, INS, BLK, WFS1, PPARG, KCNJ11, TCF7L2, CDKAL1, SLC30A8, IGF2BP2, NOTCH2, HHEX, ADAMTS9, JAZF1, IRS1, FTO, GCKR, ADCY5, PROX1, DGKB, MAFA, TRPM5, PAX6, NEUROG3, EIF2AK3, FOXP3, PTF1A, RFX6, NKX2-2, NKX6-1, MYT1, INSR, ABCC8, SLC2A4, SLC2A2, RETN, PTPN1, IAPP, SH2B1, ZFP57, CAPN10, CDKN2A, CDKN2B, CDK5, KCNQ1, PTPRD, SRR, YYIAP1, CDC123, CAMK1D, TSPAN8, LGR5, THADA, RBMS1, MTNR1B, BCL11A, TP53INP1, TLE4, ARAP1, HMGA2, ZFAND6, NCR3, AIF1, PRC1, CRY2, ADRA2A, GLIS3, C2CD4B, IGF1, TMEM195, FADS1, GIPR, MC4R, TMEM18, GNPDA2, MTCH2, NEGR1, BCDIN3D, FAIM2, PPARA, LEP, ADAM30, ETV5, DGKG, KLF14, GLP1R, GLP2R, IFIH1, GPR120, GATA2, SYNI, SYNII, CENTD2, SREBF1, DUSP9*

## Preparazione delle *library*

Per ogni campione di DNA è stata costruita una libreria di frammenti, seguendo il protocollo *Rapid Library Preparation Method Manual* GS FLX Titanium Series 454 Roche (<http://454.com/downloads/my454/documentation/gsflx/methodmanuals/GS-FLX-Titanium-Rapid-Library-Preparation-Method-Manual-%28Jan2010%29.pdf>)

La concentrazione e qualità è stata analizzata mediante *Nanodrop<sup>™</sup> 2000* (Thermo scientific) e *Qant- iT PicoGreen dsDNA Assay Kit* :

- Rapporto A260/280 > a 1,80
- Concentrazione > 5ng/μl

La costruzione della libreria ha previsto 6 passaggi principali:

### 1. DNA fragment By Nebulizzazione

Frammentazione del DNA [500ng] tramite nebulizzazione con azoto ad alta pressione (2,1 bar/min) per ottenere frammenti di DNA delle dimensioni di 400- 1000bp.

Alla fine del processo i campioni sono stati purificati mediante *Quiagen MinElute PCR Purification Kit* per eliminare tutti i frammenti troppo piccoli.

### 2. Fragment and Repair

Dal momento che la nebulizzazione produce frammenti con estremità rovinata, queste sono riparate utilizzando una specifica *End Repair mix*, seguita da un apposito ciclo al termociclatore.

### 3. Adaptor- Ligation

Ad ogni campione viene aggiunto 1µl di *Rapid Library MID Adaptors*, i quali si legano alle due estremità di ogni frammento di DNA. Ogni adattatore consiste in una coppia di oligonucleotidi a doppio filamento: un adattatore A ed un adattatore B con sequenze complementari a quelle dei primers di amplificazione e sequenziamento. Inoltre ogni adattatore al 3' contiene una *Key region*, ovvero una regione di 4 basi non palindromica, indispensabile per la verifica della qualità della reazione di sequenza, seguita da una regione di 11 nucleotidi, disegnata in modo tale che più librerie possano essere sequenziate assieme. Per questo ad ogni campione è stato associato un RL MID diverso.

### 4. Small Fragment Removal

Successivamente i frammenti sono stati selezionati per lunghezza mediante il sistema *AMPure Beads*, in maniera tale da rimuovere i frammenti troppo piccoli, i dimeri di primers e gli adattatori.

### 5. Library Quality assessment

E' stata determinata la concentrazione e qualità mediante il sistema *Nanodrop<sup>Tm</sup> 2000 (Thermo scientific)* e *Qant- iT PicoGreen dsDNA Assay Kit* e la dimensione dei frammenti attraverso i sistemi *Agilent 2100 Bio Analyzer®*.

La quantità totale di una libreria dovrebbe essere di almeno  $7,3 \cdot 10^9$  molecole, che corrisponde a una concentrazione di  $1,46 \cdot 10^8$  mol/µl.

Indice di buona qualità è:

- lunghezza dei frammenti tra 600- 900 bp
- < del 10 % di frammenti al di sotto dei 350bp.

#### 6. Pre-Capture Library amplification by LM-PCR

Successivamente ogni libreria di DNA è stata amplificata seguendo il protocollo *NimbleGen sequence Capture assays (pre-capture Library Amplification by LM- PCR)*.

La libreria pre-capture è poi purificata tramite *Qiaquick PCR Purification Kit column*.

L'amplificato ottenuto è nuovamente valutato mediante il *Nanodrop<sup>Tm</sup> 2000 (Thermo scientific)* e l'*Agilent 2100 Bio Analyzer®*. Indice di buona qualità sono:

- Concentrazione maggiore di [2,5µg]
- Grandezza frammenti di dimensioni comprese tra 600-900bp

### **Ibridazione library**

Durante questa fase si ha la “cattura” delle regioni specifiche d'interesse del DNA genomico; in questa fase è stato seguito il protocollo *NimbleGene SeqCap EZ libraries and GS FLX Titanium Rapid MID- tagged sample libraries*. ([http://www.nimblegen.com/products/lit/06560881001\\_SeqCap\\_EZLibraryLR\\_Guide\\_v2p0.pdf](http://www.nimblegen.com/products/lit/06560881001_SeqCap_EZLibraryLR_Guide_v2p0.pdf))

Per la reazione di ibridazione è importante conoscere la concentrazione di ogni libreria, in quanto si può ibridare al massimo 1µgr del pool finale delle librerie.

La libreria post-capture è poi purificata tramite *Qiaquick PCR Purification Kit column*.

Anche dopo questo passaggio dell'amplificato ottenuto ne va valutata la concentrazione al *Nanodrop<sup>Tm</sup> 2000 (Thermo scientific)* e al grandezza dei frammenti con *Agilent 2100 Bio Analyzer®*.

- Concentrazione maggiore di [1µg ]
- dimensioni dei frammenti comprese tra 500-1500bp

### **emPCR: amplificazione clonale in emulsione**

Il protocollo seguito durante questa fase è *emPCR Method Manual –Lib- LV GS FLX Titanium Series*.

[http://454.com/downloads/my454/documentation/gs-flx-plus/emPCR-Amp-Lib-LLVMethodManual\\_XLPlus\\_May2011.pdf](http://454.com/downloads/my454/documentation/gs-flx-plus/emPCR-Amp-Lib-LLVMethodManual_XLPlus_May2011.pdf)

### **Sequenziamento con GS FLX Titanium 454**

#### **Roche**

La fase di sequenziamento ha previsto in primo luogo il lavaggio del *PicoTeiterPlate (PTP)* mediante l'utilizzo di buffers appositi e la preparazione dei vari strati che andranno posizionati su di esso.

Le biglie di DNA arricchite dopo la emPCR vengono quindi trasferite successivamente nella piastra *PicoTeiterPlate (PTP)*. In

ogni pozzetto viene caricata una biglia di DNA assieme agli enzimi necessari per la reazione di sequenziamento. Congiuntamente alle biglie contenenti il DNA da sequenziare, sono contenute le biglie di controllo, le quali serviranno per il controllo interno della reazione di sequenziamento. Il vetrino verrà così posizionato sullo strumento, il quale è dotato di una camera CCD che registrerà i segnali luminosi scaturiti dalle reazioni di pirosequenziamento.

([http://454.com/downloads/my454/documentation/gs-flx-plus/GS%20FLXPlus Sequencing Method%20Manual XLPlus Kit Aug2012.pdf](http://454.com/downloads/my454/documentation/gs-flx-plus/GS%20FLXPlus%20Sequencing%20Method%20Manual%20XLPlus%20Kit%20Aug2012.pdf))

### **Analisi dei dati**

L'analisi bioinformatica è un momento fondamentale della tecnologia *next generation*, ed avviene a più livelli. Si può procedere all'analisi della *run* di sequenziamento, attraverso il software *GS Run Browser*. Inizialmente si valuta il valore di *raw wells*, che rappresenta il numero di pozzetti che sono stati caricati, e quello di *key pass wells*, che ci dice quante delle *raw wells* sono state riconosciute attraverso la *key region* presente a livello dell'adattatore; sono indici di qualità un valore di *raw wells* pari a circa il 50% del numero di biglie caricate, ed un valore di *key pass wells* che sia circa il 90% delle *raw wells*. Altri parametri di qualità della *run* sono la percentuale di *passed filter wells*, ovvero la percentuale di letture che hanno passato tutti i filtri, la quale deve essere maggiore del 50%, ed il valore di *mixed + dots key pass*, che esprime il numero di letture scartate per bassa qualità e deve essere inferiore al 25%. Da questa prima analisi si ottiene anche indicazione sulla lunghezza media, in bp, delle

letture, ed è possibile visualizzare il segnale emesso dalla PTP ed anche il *flowgram* di ogni singolo pozzetto.

L'efficienza del *sequence capture* è stata valutata tramite analisi di *coverage* delle sequenze *target* ad una profondità pari a 1 (almeno una lettura per base). Per ogni campione è stato valutato il numero totale di letture, il numero totale di basi lette, la lunghezza media dei frammenti e la profondità media del sequenziamento.

L'analisi delle varianti è stata effettuata tramite il software *GS Reference Mapper* (Roche): ogni sequenza letta è così controllata con la sequenza genomica di riferimento (assemblaggio hg19); da questa analisi si genera una lista di varianti ad alta confidenza (HC), le quali sono inizialmente controllate come possibili SNPs attraverso il dbSNP build 135. Di ogni variante nucleotidica identificata dal *GS Referenc Mapper* sono state annotate le coordinate cromosomiche, il gene, il nucleotide *wild-type* e quello mutato, l'eventuale cambio aminoacidico, la profondità di lettura per quella base ed la percentuale di letture che contengono la variante. Per l'identificazione di possibili varianti ai siti di splicing è stata calcolata la distanza di ogni variante dalla più vicina giunzione introne-esone: sono state selezionate le varianti di *splicing* entro i 10 nucleotidi dalla giunzione.

## **Validazione varianti tramite sequenziamento**

### **Sanger**

I dati ottenuti mediante l'analisi bioinformatica vengono validati con metodo Sanger e studiati con i programmi di predizione: *PolyPhen* (<http://genetics.bwh.harvard.edu/pph2/>), *SIFT*

(<http://sift.jcvi.org/>) e *PMUT* ([mmb.pcb.ub.es/PMut](http://mmb.pcb.ub.es/PMut)) . Inoltre è stato valutato il grado di conservazione dell'aminoacido *wild type* nelle diverse specie utilizzando il software CLUSTALW (<http://www.ch.embnet.org/software/ClustalW.html>).

## 4. Risultati

---

Dall'analisi mediante sequenziamento NGS sono state identificate moltissime varianti di sequenza per ogni campione analizzato. Le varianti sono state filtrate con la versione disponibile del database degli SNP (dbSNP135) ed analizzate in base alla loro posizione all'interno del gene.

La tabella 11 mostra i risultati ottenuti dal sequenziamento.

Tabella 9: Risultati campioni analizzati

| Codice pz | n° varianti totali trovate | Rs (SNPs) | rs(SNPs) OMIM associati | Varianti codificanti |
|-----------|----------------------------|-----------|-------------------------|----------------------|
| A/343     | 552                        | 125       | 3                       | 0                    |
| A/436     | 614                        | 135       | 6                       | 2                    |
| A/530     | 555                        | 120       | 4                       | 1                    |
| A/808     | 553                        | 122       | 5                       | 0                    |
| A/818     | 597                        | 122       | 4                       | 1                    |
| A/885     | /                          | /         | /                       | 1                    |
| A/886     | 654                        | 135       | 3                       | 0                    |
| A/1257    | 1356                       | 194       | 6                       | 2                    |
| B/117     | 1151                       | 170       | 2                       | 0                    |
| B/446     | 578                        | 136       | 4                       | 2                    |
| B/1349    | 1003                       | 169       | 2                       | 0                    |
| C/1066    | 1160                       | 173       | 3                       | 0                    |
| C/1070    | 1032                       | 182       | 3                       | 1                    |
| C1/071    | 1153                       | 181       | 4                       | 0                    |

Dopo l'analisi, mediante l'utilizzo di software bioinformatici, è stato considerato un numero ridotto di varianti codificanti. Al fine di validarle è stata seguita la strategia sperimentale illustrata nella figura 12.

In presenza di varianti non-sinonime o varianti nei siti di splicing, abbiamo proceduto con lo screening su HGMD®

(Listened in the Human Genome Mutation Database); è regola generale che nel caso in cui ci sia un corrispondenza della stessa, si proceda con la refertazione. Nel caso in cui non si riscontrino, l'indagine procede con la ricerca su dbSNP (build 135); nel caso in cui la variante sia stata segnalata come SNP, si procede con la valutazione della sua frequenza nella popolazione generale (MAF-*Minor Allele Frequency*). Se questa è rara, cioè <1%, procediamo con l'analisi mediante i software di predizione per valutarne la patogenicità, nel caso in cui sia una variante compresa tra 1-5%, abbiamo la necessità di valutarne il suo significato, cioè dobbiamo considerare se questa possa avere un ruolo "causativo" o "modificatore" sul fenotipo clinico. Se la ricerca su dbSNP (build 135) risulta negativa, allora continuiamo la ricerca della stessa in letteratura e sui vari database ad oggi disponibili (100 genomi, Kaviar, etc...); anche in questo caso, se il risultato dovesse essere positivo, si procede con l'analisi MAF, mentre se negativo si conclude valutando il ruolo della mutazione sui vari siti di predizione.

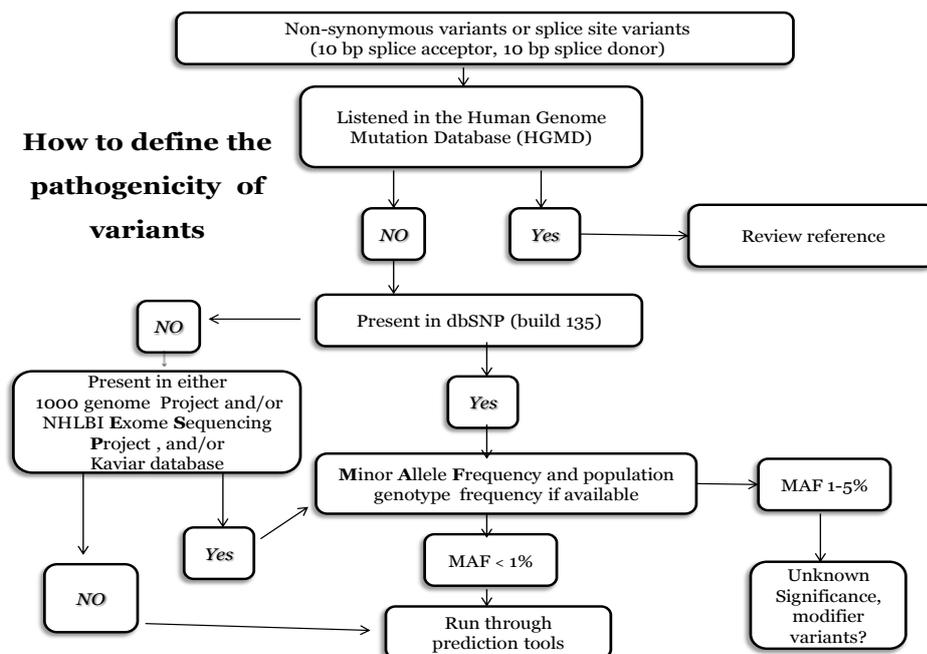


Figura 12: Strategia sperimentale utilizzata per la validazione dei risultati.

Le varianti che ci sono sembrate significative e verosimilmente patogenetiche sono state per prima cosa confermate mediante sequenziamento Sanger e successivamente ampiamente studiate (Tab.12). Inoltre abbiamo valutato anche gli SNPs OMIM (*Online Mendelian Inheritance in Man*) associati, cioè le varianti di malattia già note riportate nel database OMIM. Questo database che rappresenta la versione online del catalogo dei fenotipi mendeliani umani, contiene infatti informazioni su > 13000 geni e su tutte le malattie e i fenotipi mendeliani noti.

**Tabella 10: Riepilogo varianti significative individuate significative e potenzialmente patogenetiche. Nella colonna centrale abbiamo indicato gli SNPs OMIM associati**

| Paziente | SNP sOMIM associato   | Variazione rara patogenetica  |
|----------|---|---|
| A/343    | rs1169305 ( <i>HNFI1A</i> c.1720A>G p.Ser574Gly)<br>rs1260326 ( <i>GCKR</i> c.1403C>T p.Pro446Leu)<br>rs1801282 ( <i>PPARG</i> c.34C>G p.Pro115Gln)   |   |
| A/436    | rs5219 ( <i>KCNJ11</i> c.67G>A p.Glu23Lys)<br>rs1169288 ( <i>HNFI1A</i> c.79A>C p.Ile27Leu)<br>rs1169305 ( <i>HNFI1A</i> c.1720A>G p.Ser574Gly)<br>rs1260326 ( <i>GCKR</i> c.1403C>T p.Pro446Leu)<br>rs3856806 ( <i>PPARG</i> c.1347C>T p.His449His)<br>rs10010131 ( <i>WFS1</i> IVS4-9G>A) | <b>ABCC8</b> (NM_000353),<br>c.4500C>A,p.Ser1500Arg<br><br><b>RFX6</b> (NM_173560),<br>c.1558A>T, p.Ser520Thr |
| A/530    | rs5219 ( <i>KCNJ11</i> c.67G>A p.Glu23Lys)<br>rs1169288 ( <i>HNFI1A</i> c.79A>C p.Ile27Leu)<br>rs1169305 ( <i>HNFI1A</i> c.1720A>G p.Ser574Gly)<br>rs10010131 ( <i>WFS1</i> IVS4-9G>A)  | <b>NOTCH2</b> (NM_024408)<br>c.1315G>A, p.A439T   |
| A/808    | rs5219 ( <i>KCNJ11</i> c.67G>A p.Glu23Lys)<br>rs1169305 ( <i>HNFI1A</i> c.1720A>G p.Ser574Gly)<br>rs1260326 ( <i>GCKR</i> c.1403C>T p.Pro446Leu)<br>rs10010131 ( <i>WFS1</i> IVS4-9G>A)<br>rs13266634 ( <i>SLC30A8</i> c.826C>Trp.Arg276Trp)  |   |
| A/818    | rs5219 ( <i>KCNJ11</i> c.67G>A p.Glu23Lys)<br>rs1169305 ( <i>HNFI1A</i> c.1720A>G p.Ser574Gly)<br>rs1260326 ( <i>GCKR</i> c.1403C>T p.Pro446Leu)<br>rs10010131 ( <i>WFS1</i> IVS4-9G>A)   | <b>HNFI4A</b> (NM_175914)<br>c.287G>A, p.C96Y   |

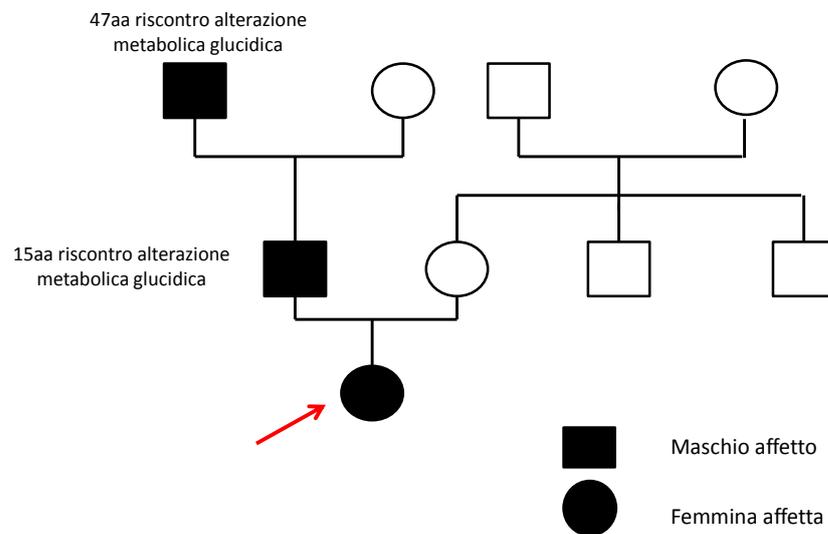
|               |   |   |
|---------------|---|---|
| <b>A/885</b>  |   | <b>MAFA</b> (NM_201589,<br>c. 604C>T, p.Hys202Tyr)  |
| <b>A/886</b>  | rs5219 ( <i>KCNJ11</i> c.67G>A p.Glu23Lys)<br>rs1169288 ( <i>HNF1A</i> c.79A>C p.Ile27Leu)<br>rs1169305 ( <i>HNF1A</i> c.1720A>G p.Ser574Gly)   |   |
| <b>A/1257</b> | rs1169288 ( <i>HNF1A</i> c.79A>C p.Ile27Leu)<br>rs1169305 ( <i>HNF1A</i> c.1720A>G p.Ser574Gly)<br>rs1260326 ( <i>GCKR</i> c.1403C>T (p.Pro446Leu)<br>rs10010131 ( <i>WFS1</i> IVS4-9G>A)<br>rs1801278 ( <i>IRS1</i> c.2911G>A p.Gly971Arg)<br>rs5219 ( <i>KCNJ11</i> c.67G>A p.Glu23Lys) | <b>GIPR</b> (NM_000164),<br>c. 983 G>C p.Arg328Pro<br><br><b>IFIH1</b> (NM_022168),<br>c. 2027 T>A p.Leu676His                  |
| <b>B/1117</b> | rs1169305 ( <i>HNF1A</i> c.1720A>G p.Ser574Gly)<br>rs5219 ( <i>KCNJ11</i> c.67G>A p.Glu23Lys)   | <b>HNF-1<math>\alpha</math></b> (MN_000545)<br>c.1720A>G (p.Ser574Gly)<br><br><b>KCNJ11</b> (NM_000525)<br>c.67G>A (p.Glu23Lys) |
| <b>B/446</b>  | rs5219 ( <i>KCNJ11</i> c.67G>A p.Glu23Lys)<br>rs1169288 ( <i>HNF1A</i> c.79A>C p.Ile27Leu)<br>rs1169305 ( <i>HNF1A</i> c.1720A>G p.Ser574Gly)<br>rs1260326 ( <i>GCKR</i> c.1403C>T (p.Pro446Leu)  | <b>WFS1</b> (NM_0065005),<br>c.2054 G>A, p.Arg685His<br><br><b>RFX6</b> (NM_173560),<br>c.1678G>A, p. Asp560Asn                 |
| <b>B/1349</b> | rs1169305 ( <i>HNF1A</i> c.1720A>G p.Ser574Gly)<br>rs5219 ( <i>KCNJ11</i> c.67G>A p.Glu23Lys)   |   |
| <b>C/1066</b> | rs1169288 ( <i>HNF1A</i> c.79A>C p.Ile27Leu)<br>rs1169305 ( <i>HNF1A</i> c.1720A>G p.Ser574Gly)<br>rs1801278 ( <i>IRS1</i> c.2911G>A p.Gly971Arg)   |   |
| <b>C/1070</b> | rs1169288 ( <i>HNF1A</i> c.79A>C p.Ile27Leu)<br>rs1169305 ( <i>HNF1A</i> c.1720A>G p.Ser574Gly)<br>rs5219 ( <i>KCNJ11</i> c.67G>A p.Glu23Lys)   | <b>RFX6</b> (NM_173560),<br>c.1865C>A, p. Thr622Lys   |
| <b>C/1071</b> | rs1169288 ( <i>HNF1A</i> c.79A>C p.Ile27Leu)<br>rs1169305 ( <i>HNF1A</i> c.1720A>G p.Ser574Gly)<br>rs1800206 ( <i>PPARA</i> c.484C>G p.Leu162Val)<br>rs5219 ( <i>KCNJ11</i> c.67G>A p.Glu23Lys)   |   |

Abbiamo quindi deciso di considerare, in questa prima parte del lavoro, i casi che presentavano varianti rare in almeno due geni e che spesso avevano un quadro clinico che sebbene ponesse il sospetto per un diabete tipo MODY, non era perfettamente inquadrabile e spesso resistente alle classiche terapie adottate nei diabete monofattoriali. Di seguito abbiamo descritto questi casi.

**Paziente A/436**

Storia clinica: il primo riscontro di iperglicemia è avvenuto ad 1 anno, con valori di glicemia a digiuno pari a 290 mg/dl.

Storia familiare: familiarità per DT2, come mostra la Fig.13



**Figura 13: Albero genealogico paziente A/436. In nero sono indicati i pazienti affetti. La freccia rossa indica il paziente esaminato.**

Terapia insulinica: GARGINE- MUTIALOG.

La prima analisi ha previsto la ricerca di mutazioni a carico del gene *GCK* (*MODY2*) e *HNF1α* (*MODY3*) le quali sono risultate essere negative.

Abbiamo quindi proceduto con l'analisi mediante NGS che ha evidenziato la presenza di due variazioni nucleotidiche. A carico del gene *ABCC8* (*NM\_000352*) è stata identificata in eterozigosi la variante c.4500C>A. Tale sostituzione determina un cambiamento amminoacidico della Serina in posizione 1500 con una Arginina (p.Ser1500Arg), ad oggi non riportata in letteratura.

L'analisi è stata estesa anche ai familiari. La stessa variante è stata riscontrata nel padre.

La seconda variante individuata, c.1558A>T in eterozigosi, interessa il gene *RFX6* (*NM\_173560*). Si tratta di una sostituzione che causa il cambiamento amminoacidico da Serina in Treonina in posizione 520 (p.Ser520Thr). Anche questa sostituzione non è stata riportata precedentemente in letteratura. La stessa variante è presente anche nella madre.

Quindi il nostro paziente, che presentava un quadro di diabete monofattoriale, tipo *MODY*, ma con una certa resistenza alle “classiche terapie” con ipoglicemizzanti orali, presenta due mutazioni, ereditate da ciascun genitore.

Inoltre sono stati riscontrati diversi polimorfismi (Tab.12) riportati in letteratura come associati ad un'aumentata suscettibilità per *DM2*:

- c.67G>A (p.Glu23Lys) in omozigosi a livello del gene *KCNJ11* (*NM\_000525*)
- c.1720A>G (p.Ser574Gly) in omozigosi a livello del gene *HNF-1α* (*MN\_000545*)

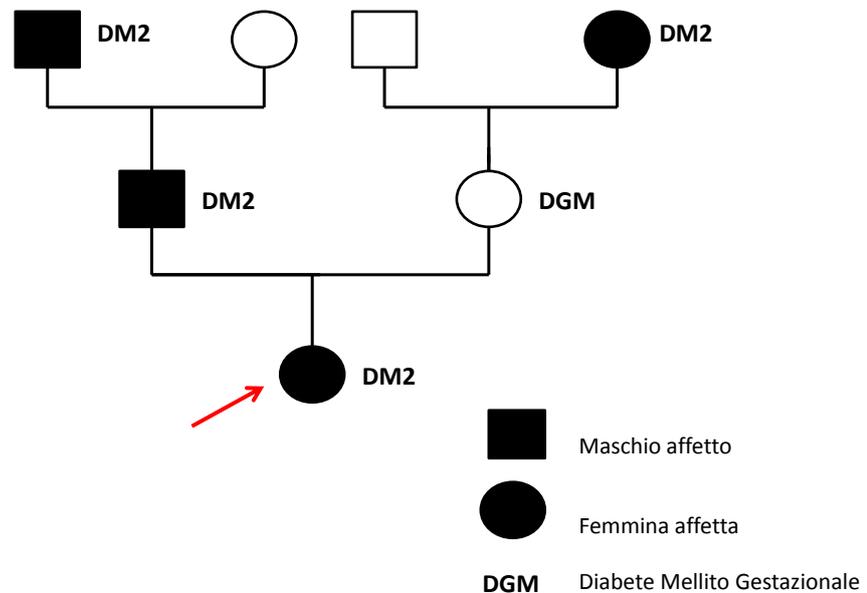
- c.1403C>T (p.Pro446Leu) in eterozigosi, a livello del gene *GCKR* (NM\_001486)

### Paziente B/446

Storia clinica: esordio con poliuria, iperglicemia a digiuno lieve-moderata, esami ICA, GAD, IA2 e anti-ins negativi.

Terapia: EUGLUCON 1C\*3, NOVONORM.

Storia familiare: familiarità per DT2 come mostra la Fig.14



**Figura 14: Albero genealogico paziente B/446. In nero sono indicati i familiari affetti da Diabete mellito di tipo 2(DM2), il paziente esaminato è indicato dalla freccia rossa .**

Le prime indagini molecolari sono state eseguite per sospetto MODY2 e MODY3, le quali sono risultate negative.

La successiva analisi mediante NGS ha evidenziato la presenza di due varianti in eterozigosi. A carico del gene *RFX6*

(NM\_173560) è stata individuata la sostituzione c.1678G>A in eterozigosi, con il conseguente cambiamento amminoacidico in posizione 560 di un Acido Aspartico con una Asparagina. Tale variante è stata riscontrata anche nella madre.

La seconda variante riscontrata è c.2054G>A sul gene *WFS1* (NM\_0065005) e causa la sostituzione amminoacidica da Arginina ad Istidina in posizione 685 (p.Arg685His). La variante è stata già riportata in letteratura a carico dello stesso codone ma con una diversa sostituzione amminoacidica, la quale, oltre la predisposizione al diabete, determina sordità neuro sensoriale (38).

La variante è presente anche nella madre.

### **Paziente A/1257**

Storia clinica: diabete gestazionale (DMG) durante la seconda gravidanza e familiarità per DT2. Terapia insulinica durante la seconda e terza gravidanza(Fig.15).

Storia familiare: familiarità per DT2 (Fig.15) e sospetto MODY2.

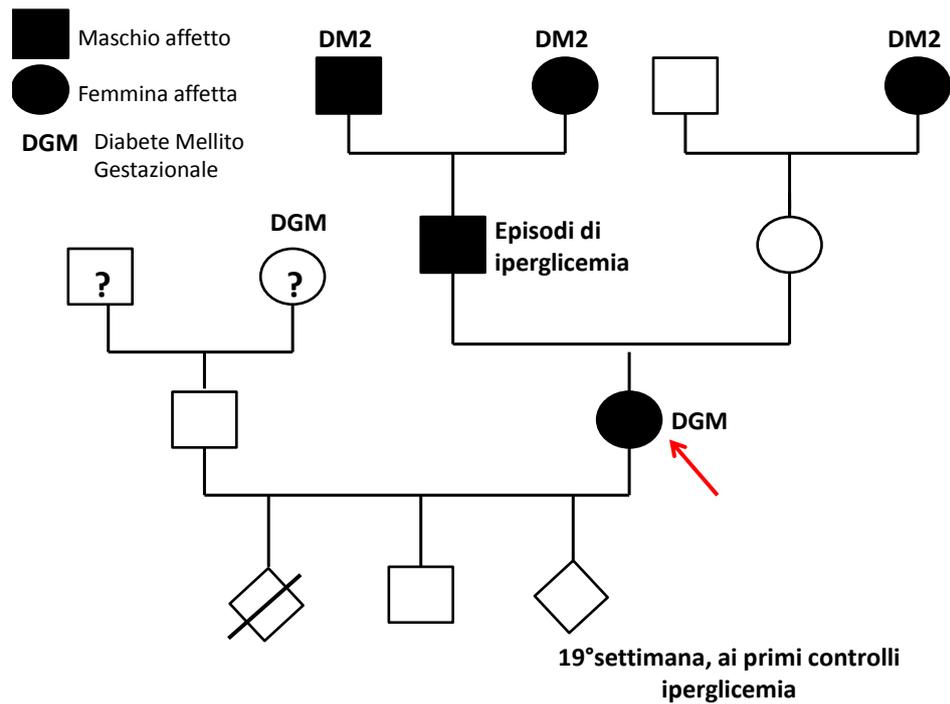


Figura 15: Albero genealogico paziente A/1257, indicato dalla freccia rossa, il quale è affetto da Diabete Mellito Gestazionale (DGM).

Le prime indagini molecolari sono state eseguite per i geni *GCK* e *IPF1*, le quali sono risultati negative.

L'analisi mediante NGS ha mostrato la presenza di due sostituzioni aminoacidiche in eterozigosi. A carico del gene *GIPR* (NM\_000164) è stata riscontrata la variante c.983G>C che determina una sostituzione amminoacidica da Arginina in Prolina in posizione 328 (p.Arg328Pro). Questa sostituzione non è stata mai riportata in letteratura.

L'altra variante individuata è c.2027T>A a livello del gene *IFIH1*(NM\_0221868), la quale è responsabile della sostituzione amminoacidica della Lisina in posizione 676 con Istidina (p.Leu676His). La variante non è stata mai riportata in letteratura.

Sono stati individuati anche diversi polimorfismi (Tab.12), i più significativi sono:

- IVS4-9G>A a livello del gene *WFS1* (NM\_0065005)
- c.1403C>T (p.Pro446Leu) a carico del gene *GCKR* (NM\_001686)
- c.1720A>G (p.Ser574Gly) a livello del gene *HNF-1α* (NM\_000545)

### **Paziente C/1070**

Storia clinica: glicemia a digiuno 131 mg/dl

Indagini per sospetto MODY

Storia familiare: la madre ha mostrato intolleranza glucidica con valori >200 mg/dl.

L'analisi mediante NGS ha mostrato la presenza di due sostituzioni amminoacidiche. A carico del gene *RFX6* (NM\_173560) è stata trovata la variante c.1865C>A in eterozigosi la quale determina il cambiamento dell'aminoacido Treonina 622 in Lisina (p.Thr622Lys). Tale variante non è mai stata riportata in letteratura.

La seconda variante identificata è c.67G>A in omozigosi a carico del gene *KCNJ11* (NM\_000525) che determina la sostituzione dell'amminoacido Acido Glutammico 23 in Lisina (p.Glu23Lys).

Inoltre è stata evidenziata la presenza in omozigosi della variante c.1720A>G a carico dell'esone 9 del gene *HNF-1α* (NM\_000545).

**Paziente B/1117**

Storia clinica: frequenti ricoveri per tachicardia ventricolare a seguito di sforzi o stress psicofisico. Durante il corso di alcune indagini metaboliche viene riscontrata l'ipoglicemia.

Trattamento:  $\beta$ -bloccanti per la tachicardia.

Dall'analisi NGS sono stati riscontrati la presenza di due polimorfismi a carico dei geni:

- Variante in omozigosi c.1720A>G (p.Ser574Gly) a carico del gene *HNF-1a* (NM\_000545).
- Una sostituzione c.67G>A sul gene *KCNJ11* (NM\_000525) in omozigosi, la quale causa un cambiamento amminoacidico da Acido Aspartico a Lisina in posizione 23 (p.Glu23Lys).

## 5. Discussione

---

Il lavoro di questa tesi è stato finalizzato ad un nuovo approccio di screening mutazionale, attraverso il sequenziamento massivo del DNA, di pazienti giunti alla nostra osservazione con diagnosi di diabete monofattoriale tipo MODY. La sfida è stata quella di analizzare un numero crescente di geni candidati associati alla malattia in una coorte di soggetti affetti, risultati negativi ad una prima analisi dei geni *GCK*, *HNF1 $\alpha$* , *HNF4 $\alpha$* , *HNF1 $\beta$*  e *IPF1* e quindi di rivalutarli da un punto di vista molecolare mediante NGS (*Next Generation Sequencing*), che rappresenta ormai una metodica di elezione per lo studio di ampie coorti di pazienti e per l'individuazione di nuovi geni malattia.

Questi ultimi anni hanno visto un'ampia accelerazione delle pubblicazioni in cui il sequenziamento di *next generation* viene applicato. Applicazioni importanti includono:

- risequenziamento dell'intero genoma o sequenziamenti più mirati in modo da evidenziare mutazioni o polimorfismi;
- mappatura di riarrangiamenti strutturali, che possono includere variazioni in numero di copie del DNA (CNVs), breakpoint di traslocazioni bilanciate e inversioni cromosomiche;
- 'RNA-Seq', per individuare l'espressione di differenti geni da librerie derivate da mRNA o piccoli RNA che possono essere sequenziali in profondità;
- analisi della metilazione del DNA su larga scala, mediante profondo sequenziamento di DNA trattato con bisolfito
- 'ChIP-Seq', o mappatura dell'intero genoma e delle interazioni proteina-DNA, mediante sequenziamento di frammenti di DNA annullato mediante immunoprecipitazione della cromatina (36).

Attualmente, l'applicazione di NGS più utilizzata in studi genetici di tratti complessi coinvolge la scoperta di SNPs e la genotipizzazione. Ma soprattutto, a differenza dei precedenti studi di Genome Wide (GWS) applicabili alle malattie complesse, permette la scoperta sistematica di varianti a bassa frequenza, in migliaia di campioni e l'associazione di queste varianti con i fenotipi di interesse (37). Esperimenti su larga scala, come i “1000 genomi” (<http://www.1000genomes.org/page.php>) ed i progetti UK10k (<http://www.uk10k.org/>) stanno fornendo un eccellente risorsa di dati di sequenza in modo da ricavare informazioni di varianti presenti con una certa frequenza nella popolazione generale, di utile riscontro per la comunità scientifica. Strategie di sequenziamento per l'identificazione di loci associati a malattie e tratti complessi possono variare notevolmente e sono dotati di diverse sensibilità ed esigenze di specificità. Uno studio sul diabete di tipo 1, ad esempio, ha fornito uno dei primi esempi di identificazione di locus malattia grazie all'applicazione di NGS. Nejentsev et al., utilizzando la piattaforma 454, hanno identificato quattro SNPs rari a bassa frequenza significativamente associati a diabete tipo 1 nel gene *IFIH1* dopo risequenziamento di 480 casi e 480 controlli, identificando *IFIH1* appunto come il gene-malattia più probabile (38).

Sulla base di queste considerazioni abbiamo deciso di usare una simile strategia anche per lo studio dei nostri casi che erano risultati negativi alle prime indagini molecolari.

Il diabete tipo MODY, spesso in presenza di quadri clinici sovrapponibili al DT2, presenta un'ampia eterogeneità genetica e allelica, con considerevoli sovrapposizioni fenotipiche, che quindi richiedono notevole sforzo per l'analisi di mutazione, al fine di identificarne l'eziologia molecolare.

Per il nostro studio abbiamo usato la piattaforma 454GS FLX Titanium (Roche), presente presso l'Unità di Genetica Medica dell'AOU Meyer. Attraverso l'uso di questa tecnologia è stato possibile identificare una o più varianti in diversi geni candidati o causativi di malattia, abbiamo quindi determinato l'origine parentale delle stesse, e cercato di individuare interazioni gene/gene e gene/ambiente, per ciascun campione esaminato.

Il nostro esperimento si è basato su sequenziare una Megabase circa di DNA, nella quale erano incluse porzioni codificanti e non di 103 geni coinvolti in forme di diabete monofattoriale tipo MODY o che dopo studi Genome Wide (GWS) erano considerati predisponenti a DT2. Alcuni dei geni che abbiamo scelto, ad oggi, non sono mai stati riportati mutati ma, sulla base di esperimenti su modelli animali che determinavano quadri associabili a diabete, abbiamo pensato che sarebbero potuti essere causativi di malattia anche nell'uomo.

In questa prima parte del lavoro, l'analisi NGS è stata effettuata su 14 campioni selezionati sulla base di un'alta familiarità per diabete e con una storia clinica ascrivibile a MODY.

Abbiamo focalizzato la nostra attenzione sulle varianti associate a cambiamenti aminoacidici, e tutte queste sono state avvalorate con la tecnica di sequenziamento Sanger classico.

Una volta che i risultati sono stati riconfermati, le varianti più indicative sono state ampiamente studiate. Tra queste sono state valutate differenti mutazioni localizzate in porzioni codificanti di alcuni geni, che abbiamo riportato nei Risultati (Tabella 12).

In generale, per cercare di chiarire il potenziale significato patogenetico delle rare mutazioni trovate sono state compiute

delle analisi *in silico* con i più diffusi *software* di bioinformatica: SIFT, PolyPhen e pMut. Il programma SIFT (*Sorting Intolerant From Tolerant Aminoacid Sostitution*) fa ricorso ad un approccio di tipo evolucionistico basandosi sull'assunzione che aminoacidi funzionalmente e strutturalmente importanti tendono ad essere conservati tra le specie. In base a tale considerazione e alle caratteristiche fisico-chimiche degli aminoacidi, il programma assegna una probabilità di sostituzione per ogni possibile cambio aminoacidico. Tale parametro fornisce un valore della significatività della sostituzione aminoacidica esaminata permettendo di considerare una variante come tollerata o meno. Il *software* disponibile in rete Polyphen permette invece di valutare l'impatto di una sostituzione nucleotidica sulla struttura e funzione di una proteina attraverso valutazioni strutturali e comparative. pMut invece permette la scansione veloce di hot spot mutazionali lungo l'intera sequenza di riferimento, aiutando ad individuare le regioni in cui le mutazioni sembrano avere un maggiore impatto patologico.

In questa prima parte del lavoro abbiamo deciso di considerare in modo particolare quei soggetti in cui si sono riscontrate mutazioni rare di più geni.

Relativamente al caso A/436, è stata trovata, sul gene *ABCC8*, la mutazione c.4500C>A che determina una sostituzione Ser1500Arg. La Serina, polare, è dunque sostituita da un aminoacido basico (Arginina) che può comportare un cambiamento nella struttura secondaria della proteina. Infatti nella porzione terminale di questa sono presenti i domini di legame all'ATP (AAA+ATPasi domain e ABC transporter-like domain). Una mutazione a questo livello potrebbe dunque interferire con l'attività ATPasica della proteina. Questa è

implicata soprattutto nei processi biologici del trasporto di ioni potassio.

Il gene *ABCC8* è noto essere associato ad una grave forma di diabete neonatale (Familial Hyperinsulinemic Hypoglycemia - HHH1) che si trasmette con meccanismo autosomico recessivo o, quando la mutazione è presente in eterozigosi, a forme di iperglicemia, riconosciute nell'adulto come associate a DT2 (39). Inoltre le varianti in eterozigosi sono state riportate in rare forme di MODY12 (18). Il gene *ABCC8* codifica per le (*sulfonylurea receptor 1*) della pompa potassio ATP dipendente (K-ATP) che è un etero ottamero formato da 4 subunità *KIR 6.2* e 4 subunità *SUR 1*. Il canale è implicato nell'attività elettrica della membrana delle cellule  $\beta$ , per questo gioca un ruolo fondamentale nella secrezione insulinica (18) (19). La variante individuata nel nostro caso non è stata mai riportata in letteratura.

In questo stesso soggetto abbiamo anche rilevato la sostituzione c.1558A>T sul gene *RFX6* (Tab.12), anche questa al momento non ancora riportata in letteratura. Il gene *RFX6* è situato sul cromosoma 6q22.1 e codifica per un fattore trascrizionale che è stato dimostrato avere un ruolo nello sviluppo delle cellule pancreatiche, sia nel topo che nell'uomo (40) (41).

Mutazioni in omozigosi di questo gene sono state riportate in rari pazienti che presentano diabete neonatale, atresia intestinale e agenesia della vescica (41) (42). I genitori dei due soggetti riportati in letteratura presentavano iperglicemia e diagnosi di DT2. Recentemente è stato dimostrato, attraverso studi sia nell'uomo che nel topo, che *RFX6* è essenziale per lo sviluppo del pancreas endocrino. Questo gene mantiene una specifica espressione nella produzione di insulina nelle cellule del pancreas nel periodo postnatale. Pertanto, *RFX6* sembra avere un ruolo

fondamentale nel mantenimento del fenotipo delle beta-cellule in aggiunta al loro sviluppo (43).

Sono state poi ritrovate altre 2 varianti diverse su due differenti campioni a carico di questo stesso gene (Tab.12), ovvero, c.1678G>A e c.1865C>A, nessuna delle quale risulta essere presente in letteratura. Quest'ultima si presenta isolata in un soggetto con diagnosi di diabete tipo MODY, mentre l'altra è associata ad un'altra rara variante sul gene *WFS1* (c.2054 G>A).

Varianti nel gene *WFS1* sono state riportate in diversi lavori della letteratura associate alla sindrome di Wolfram, un disordine autosomico recessivo caratterizzato da una forma monogenica di diabete e patologia neurodegenerativa, associata a diabete ad esordio infantile ed atrofia ottica. Diabete e progressiva atrofia ottica si manifestano nella prima decade di vita, sordità neurosensoriale e diabete insipido nella seconda decade, mentre vescica neurogena e complicazioni neurologiche, tra cui atassia e sintomi psichiatrici, nella terza e quarta decade di vita. I soggetti eterozigoti per varianti in questo gene presentano prevalenza di DT2, associato o meno a sordità, che compare generalmente nell'età adulta (44). La wolframina (*WFS1*) è una proteina transmembrana del reticolo endoplasmatico (ER), che si localizza ubiquitariamente, ma soprattutto ad alti livelli nelle cellule  $\beta$  pancreatiche e neuroni specifici del sistema nervoso centrale (45). Controlla un ciclo di feedback negativo sulla regolamentazione della rete del ER dopo segnalazione allo stress, e funge lei stessa da soppressore di stress nel ER. Segnalazione dello stress sul ER è strettamente regolata nelle cellule beta del pancreas al fine di produrre una adeguata quantità di insulina in risposta alle frequenti fluttuazioni dei livelli di glucosio nel sangue (46). *WFS1* è sottoregolata durante la secrezione di insulina e svolge un ruolo

importante nel mantenimento dell'omeostasi del ER e dell'integrità delle cellule  $\beta$  pancreatiche. L'inattivazione del gene *WFS1* nelle cellule  $\beta$  di roditori è causa di stress sul ER e di morte delle cellule beta per accelerata apoptosi (47). Se è vero che esistono forti associazioni tra alcuni SNPs e varianti di questo gene con DT2, a maggior ragione varianti più rare potrebbero avere un effettivo ruolo causativo sull'insorgenza del diabete; se queste vengono trasmesse in modo dominante, il fenotipo clinico può essere descritto come diabete monofattoriale in senso lato.

Relativamente al caso A/1257, abbiamo riscontrato due nuove varianti a livello dei geni *GIPR* (c.983G>C) e *IFIH1* (c.2027T>A). *GIPR* codifica per il recettore polipeptide insulinotopico (GIP) glucosio-dipendente che è un ormone incretinico. GIP interagendo con il suo recettore sulle cellule  $\beta$ -porta ad un aumento cAMP e  $Ca^{2+}$  intracellulare che provoca esocitosi dei granuli di insulina in risposta tardiva al glucosio orale (20-120 min.) (48). Il gene *IFIH1* (cromosoma 2q24) codifica per l'Interferone-indotto con elicasi C dominio 1 (IFIH1), ed è in grado di attivare la risposta di tipo 1 ad interferone e citochine pro-infiammatorie da domini CARD, dopo il riconoscimento del doppio filamento di RNA di virus, come ad esempio Enterovirus, nel citoplasma di cellule infettate (49). Abbiamo già visto che è stato riportato essere uno dei geni responsabili del diabete tipo 1, in quanto mutazioni nel dominio proteico dell'elicasi potrebbero essere associate ad un migliore riconoscimento da parte di virus a RNA, determinando un maggior rischio di sviluppare il diabete (38).

In effetti il fenotipo clinico della nostra paziente non è immediatamente ascrivibile ad un tipico diabete monofattoriale tipo MODY, sebbene alcune mutazioni del gene *CGK* siano associate a diabete gestazionale. La storia familiare non è però

compatibile con una classica storia di mutazioni da glucochinasi e il riscontro di questo assetto genomico ci spinge a concentrarci sul fatto che è sempre più verosimile che quadri non “classici” di diabete tipo MODY/DT2 siano ascrivibili ad un background genomico ancora in via di definizione.

La paziente B/1117 ha una storia differente rispetto ai casi che generalmente giungono alla nostra osservazione. Ha presentato frequenti ricoveri per tachicardia ventricolare a seguito di sforzi o stress psicofisico, e solo durante il corso di alcune indagini viene riscontrata l'ipoglicemia. Dall'analisi sono stati riscontrati la presenza di due polimorfismi a carico dei geni *HNF-1α* (c.1720A>G) e *KCNJ11* (c.67G>A). Quest'ultimo gene codifica per un canale del potassio e questi tipi di proteine sono presenti ubiquitariamente nelle cellule di mammifero, partecipando a una vasta gamma di risposte fisiologiche. La proteina codificata da questo gene è una proteina integrale di membrana che soprattutto ha la funzione di far entrare il K all'interno di una cellula piuttosto che permetterne l'uscita; è controllata da proteine G e si trova associata al recettore SUR della sulfonilurea. Mutazioni in questo gene sono causa di ipoglicemia persistente iperinsulemica familiare dell'infanzia (PHHI), una malattia autosomica recessiva caratterizzata da alterazione nella secrezione di insulina. Inoltre difetti di questo gene possono anche contribuire anche alle forme dominanti DT2 non-insulino-dipendente (NIDDM), diabete mellito neonatale transitorio di tipo 3 (TNDM3) e diabete mellito neonatale permanente (PNDM).

La variante individuata nella paziente in omozigosi giustifica quindi sia la comparsa di diabete che il quadro cardiaco in quanto anche in altre situazioni riportate nella letteratura è stata descritta un'associazione con aritmia e “alterazioni cardiache” in

risposta allo stress. In un recente ampio studio, il polimorfismo Glu23Lys è stato implicato come fattore di rischio per il “rimodellamento cardiaco subclinico non adattativo” tra individui con maggiore carico di stress, ma il suo significato in una specifica coorte di soggetti con insufficienza cardiaca rimane ancora da definire con esattezza (50). Inoltre, la mancanza di un modello atomico per subunità di canali  $K_{ATP}$  limita la comprensione dei meccanismi molecolari di suscettibilità alla malattia favorita da questo polimorfismo. Sicuramente pazienti omozigoti per questa variante hanno una risposta anomala all’esercizio, come dimostrato da un aumento della frequenza cardiaca durante le prove da sforzo al tapis roulant. Modelli *in silico* suggeriscono che la sostituzione dell’amminoacido Glicina con la Lisina comprometterebbe l’integrità strutturale della proteina, influenzando l’apertura del canale (51).

## 6. Conclusioni

---

Lo scopo del nostro lavoro è stato quindi quello di identificare varianti rare e polimorfismi associati a diabete al fine di, una volta stabilito l'assetto genomico dei soggetti in studio, e individuata la possibile interazione tra i diversi geni, proporre una "terapia personalizzata". Infatti la maggior parte dei nostri soggetti spesso è risultata resistente o "poor responded" alle comuni terapie somministrate nelle forme più "classiche" di diabete monofattoriale.

La medicina personalizzata nel diabete si riferisce al conoscere le caratteristiche specifiche di un paziente per dirigere strategie diagnostiche o trattamenti che siano più efficaci per quell'individuo. Lo spettro di informazioni che possono guidare le decisioni personalizzate sulla cura del diabete include singole caratteristiche cliniche e fenotipiche, risultati clinici standard di laboratorio, e le sequenze di geni e di altri marcatori molecolari. Dato il notevole progresso nel corso degli ultimi anni nella caratterizzazione di sequenze di geni umani e nuove tecnologie molecolari, ora c'è particolare interesse per la possibilità di utilizzare i singoli biomarcatori molecolari per le decisioni dirette di pazienti sulla gestione del diabete. Soprattutto in pazienti che sono inclusi nella vasta categoria di DT2, vi è una sostanziale variabilità dei fattori di rischio genetici che stanno alla base dei meccanismi patogenetici e quindi al fenotipo clinico. Tuttavia, i pazienti con diabete di tipo 2 sono spesso trattati in modo analogo, con scarsa considerazione delle caratteristiche individuali che possono influenzare l'esito clinico e la risposta terapeutica (52).

Il DT2, come abbiamo precedentemente illustrato, è tipicamente caratterizzato da una combinazione di anomalie sia nella secrezione di insulina che nella reattività, oltre a un più graduale e meno estesa perdita della capacità secretoria da parte

delle  $\beta$ -cellule. Per questo motivo, le opzioni disponibili per la gestione della glicemia, in questa tipologia di diabete, includono non solo insulina esogena, ma anche uno spettro di agenti farmacologici con azioni che includono l'aumento della sensibilità all'insulina, la stimolazione di secrezione di insulina, e il rallentando nell'assorbimento intestinale del glucosio. GWS finora hanno identificato diversi geni con variazioni di sequenza associati a T2D in popolazioni multiple, e molti altri geni sono stati collegati a T2D in studi su singole popolazioni. Tuttavia, ciascuna di queste varianti geniche identificate da sole o in combinazione con altre varianti genetiche ha il potenziale per dirigere decisioni individualizzate sulla terapia del diabete.

Sulla base di queste considerazioni e sul fatto che quadri di MODY e DT2 spesso si sovrappongono, riteniamo che lo studio di varianti di tutti i geni associati al diabete e specialmente combinazioni di queste varianti saranno il contesto essenziale per la terapia personalizzata per ogni paziente.

## **Bibliografia**

1. **Laura Dean, MD and Jo McEntyre, PhD.** *The genetic Landscape of Diabetes.* s.l. : National Center for Biotechnology Information (US), 2004.

2. *Nobel Prize for the discovery of insuline*  
([www.nobel.se/medicine/laureates/1923/index.html](http://www.nobel.se/medicine/laureates/1923/index.html)).

3. *Developement of insulin*  
(<http://digital.library.utoronto.ca/insulin/>), university of Toronto.

4. **D. Nelson, M. Cox.** *I principi di biochimica di Lehninger.* s.l. : Zanichelli, 2006.

5. **Molina, Patricia E.** *Fisiologia endocrina.* Milano : McGraw-Hill, 2004.

6. *Epidemiologia del diabete.* **Vespasiani G, Nicolucci A, Giorda.** s.l. : Annali della Sanità Pubblica, 2005.

7. *Type 2 diabetes across generations: from Pathophysiology to prevention and management.* **Cristopher J Nolan, Peter Damm, Marc Prentki.** s.l. : Lancet, 2011, Lancet, Vol. 378.

8. *Epidemiology of Diabetes and Diabetes- Related Complications.* **Anjali D Deshpande, Mercie Harris- Hayes, Mario Schootman.,** s.l. : Physical Therapy, 2008, Vol. 88. n°11

9. *Lifetime risk for diabetes mellitus in the United States.* **Narayan KM, Boyle JP, Thompson TJ.** s.l. : JAMA, 2003.

10. *Prevalence and trends in obesity among US adults.* **Flegal KM, Carrol MD, Ogden CL.** s.l. : JAMA, 2000.

11. *Diabetes mellitus: a "thrifty" genotype rendered detrimental by "progress"?*. **JV, Neel.** s.l. : JAMA, 1962.

12. *Maturity-onset Diabetes of the Young (MODY) Genes: Literature Review.* **Attiya K., Sahar F.** s.l. : Clinical Practice, 2012, Vol. 1. n°1

13. *Clinical implications of a molecular genetic classification of monogenic b- cell diabetes.* **Rinki Murphy, Sian Ellard, Andrew T Hattersley.** s.l. : Nature, 2008, Vol. 4 n°4.

14. *Investigating maturity onset diabetes of the young.* **Ohn Nyunt, Joice Y Wu, Ivan N McGown.** s.l. : Clin Biochem Rev, 2009, Vol. 30.

15. *Mutations in the genes encoding the transcription factors hepatocyte nuclear factor 1 alpha (HNF1A) and 4 alpha (HNF4A) in maturity- onset diabetes of the young.* **Sian Ellard, Kavin Colclough.,** s.l. : Human Mutation, 2006, Vol. 27 n°9.

16. *Update on mutation in glucokinase (GCK), which cause maturity- onset diabetes of the young, permanent neonatal diabetes, and Hyperinsulinemic hypoglycemia.* **Kara K. Osbak, Kevin Colclough, Cecile Saint- Martin.,** s.l. : HGVS, 2009, Vol. 30 n°11

17. *Glucokinase (GCK) mutations and their characterization in MODY2 children of southern Italy.* **Marina Capuano, Carmen Maria Garcia -Herrero, Nadia Tinto,** s.l. : PLoS ONE, 2012, Vol. 7 n°6.

18. *Whole- exome sequencing and high throughput genotyping identified KCNJ11 as the thirteenth MODY gene.* **Bonnefond A, Philippe J, Durand E.,** s.l. : PLoS ONE, 2012, Vol. 7, n°6.

19. *Heterozygous ABCC8 mutations are a cause of MODY*. **P. Bowman, S. E. Flanagan, E. L. Edghill**. s.l. : Springer, 2011.
20. *Clinical features and treatment of maturity onset diabetes of the young (MODY)*. **Daphne SL, Gardner E, Shyong Tai**. s.l. : DovePress, 2012, Vol. 5.
21. *Best practice guidelines for the molecular genetic diagnosis of maturity-onset diabetes of the young*. **S. Ellard, C. Bellannè-Chantelot, A. T. Hattersley**. s.l. : Springer, 2008, Vol. 51: 546-553.
22. *Il Maturity Onset Diabetes of the Young (MODY)*. **F. Barbetti**., s.l. : Andrea Mosca, 2006, Vol. 30. 5-6
23. *New method for sequencing DNA*. **Maxam AM, Gilibert W.A.**, s.l. : Proc Natl Acad USA, 1977, Vol. 74. n°4
24. *DNA Sequencing with chain-terminating inhibitors*. **Sanger F, Nicklen S, Coulson AR.**, s.l. : Proc Natl Acad Sci USA, 1977, Vol. 74 n° 7
25. *Next- Generation Sequencing: From Basic Reserch to Diagnostics*. **Karl V. Voelkerding, Shale A. Dames, Jacob D. Durtschi**., s.l. : Clinica Chemistry, 2009, Vol. 55, n°4
26. *Evalutation of next generation sequencing platforms for population targeted sequencing studies*. **Harismmendy O., Pauline C Ng, Strausberg R.**, s.l. : Genome Biology, 2009, Vol. 10, n° r32
27. *Method of the year, next-generation DNA sequencing. Functional genomics and medical applications*. **S.C. Schuster et al**, s.l. : Nat. Methods, 2008, Vol. 5.

28. *Sequencing technologies -The next generation.* **Metzker, Michael L.** s.l. : Macmillan, 2010, Vol. 11.

29. *Next-generation DNA sequencing techniques.* **Ansorge, Wilhem J.**, s.l. : ELSEVIER New Biotechnology, 2009, Vol. 25 n°4.

30. *The Genome Sequencer FLX TM System- Longer reads, more applications, straight forward bioinformatics and more complete ata sets.* **Marcus Droege, Brendon Hill.**, s.l. : ELSELVIER, 2008, Vol. 3 n°10.

31. *Accuracy and quality of massively parallel DNA pyrosequencing.* **Susan M Huse, Julie A Huber, Hilary G Morrison.**, s.l. : Genome Biol, 2007, Vol. 8 n°7.

32. *Sense from sequence reads: methods for alignment and assembly.* **Flicek P, Birney.** s.l. : Nat. Methods, 2009, Vol. 6.

33. *A new algorithm for DNA sequence assembly.* **Idury R, Waterman M.** s.l. : J. Comput. Biol., 1995, Vol. 2.

34. *ABYSS: a parallel assembler for short read sequence data.* **Simpson J, Wong K, Jackman, S.** s.l. : Genome Res., 2009, Vol. 19.

35. *The development and impact of 454 sequencing.* **Jonathan M. Rothberg, John H. Leamon.**, s.l. : Nature Biotechnology, 2008, Vol. 26 n°10.

36. *Clinical integration of next-generation sequencing technology.* **Gullapalli RR, Lyons-Weiler M, Petrosko P, Dhir R, Becich MJ, LaFramboise WA.**, s.l. : Clin Lab Med, 2012, Vol. 32 n°4.

37. *Next-generation association studies for complex traits* .  
**Zeggini, Eleftheria.**, s.l. : Nat Genet, 2011, Vol. 43 n°4.

38. *Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes.* **Nejentsev S, Walker N, Riches D, Egholm M, Todd JA.** s.l. : Science, 2009.

39. *Inactivation of the first nucleotide-binding fold of the sulfonylurea receptor, and familial persistent hyperinsulinemic hypoglycemia of infancy.* **Thomas PM, Wohllk N, Huang E, Kuhnle U, Rabl W, Gagel RF, Cote GJ.**, s.l. : Am. J. Hum. Genet., 1996, Vol. 59 n°3.

40. *Rfx6 directs islet formation and insulin production in mice and humans.* **Smith SB.** s.l. : Nature, 2010.

41. *Glucose intolerance caused by a defect in the entero-insular axis: A study in gastric inhibitory polypeptide receptor knockout mice.* **Kazumasa Miyawaki, Yuichiro Yamada, Hideki Yano.** s.l. : Proc Natl Acad Sci U S A, 1999.

42. *A further example of a distinctive autosomal recessive syndrome comprising neonatal diabetes mellitus, intestinal atresias and gall bladder agenesis.* **Chappell L, et al.** 13 , s.l. : Am J Med Genet A, 2008, Vol. 146A.

43. *RFX6 is needed for the development and maintenance of the  $\beta$ -cell phenotype.* **Taleb N, Polychronakos C.**, s.l. : Epub, 2011 , Vol. 3 n°5.

44. *Decreased insulin secretion and increased risk of type 2 diabetes associated with allelic variations of the WFS1 gene: the Data from Epidemiological Study on the Insulin Resistance Syndrome (DESIR) prospective study.* **Cheurfa N, Brenner GM,**

**Reis AF, Dubois-Laforgue D.**, s.l. : Diabetologia, 2011, Vol. 54 n°3.

45. *WFS1 (Wolfram syndrome 1) gene product: predominant subcellular localization to endoplasmic reticulum in cultured cells and neuronal expression in rat brain.* **Takeda K, Inoue H, Tanizawa Y et al.**, s.l. : Hum Mol Genet , 2001, Vol. 10 n°5.

46. *Endoplasmic reticulum stress signaling in pancreatic beta-cells.* . **Fonseca SG, Lipson KL, Urano F.**, s.l. : Antioxid Redox Signal, 2007, Vol. 9 n°12.

47. *Mice conditionally lacking the Wolfram gene in pancreatic islet beta cells exhibit diabetes as a result of enhanced endoplasmic reticulum stress and apoptosis.* **Riggs AC, Bernal-Mizrachi E, Ohsugi M et al.** s.l. : Diabetologia , 2005, Vol. 48.

48. *Role of incretin hormones in the regulation of insulin secretion in diabetic and nondiabetic humans.* . **Holst JJ, Gromada J.** s.l. : Am J Physiol Endocrinol Metab, 2004.

49. *Interferon induced with helicase C domain 1 (IFIH1) and virus-induced autoimmunity: a review.* **Chistiakov, D.A.**, s.l. : Viral Immunol, 2010, Vol. 23 n°1.

50. *KATP channel polymorphism is associated with left ventricular size in hypertensive individuals: a large-scale community-based study.* **Reyes S, Terzic A, Mahoney DW, Redfield MM.**, s.l. : Hum Genet , 2008, Vol. 123 n°6.

51. *KATP channel Kir6.2 E23K variant overrepresented in human heart failure is associated with impaired exercise stress response.* . **Reyes S, Park S, Johnson BD, Terzic A, Olson TM.**, s.l. : Hum Genet., 2009, Vol. 126 n°6.

52. *Individualizing therapies in type 2 diabetes mellitus based on patient characteristics: what we know and what we need to know.* . **Nathan DM, Arslanian SA, Groop L, Rizza RA, Rotter JL.**, s.l. : J Clin Endocrinol Metab, 2010, Vol. 95 4.