

UNIVERSITÀ DEGLI STUDI DI CATANIA

FACOLTA' DI MEDICINA E CHIRURGIA

**DOTTORATO DI RICERCA IN PATOLOGIA ED EMATOLOGIA
CLINICA, SPERIMENTALE E COMPUTAZIONALE (EX EMBRIOLOGIA
MEDICA E PATOLOGIA ED EMATOLOGIA SPERIMENTALE)**

SEBASTIANO DI BELLA

**RNA-interference e Farmacogenomica: dall'analisi
dell'RNA agli effetti dei farmaci**

TESI DI DOTTORATO

Relatore:

Chiar.mo Prof. Alfredo Ferro

Correlatore:

Prof.ssa Rosalba Giugno

Prof. Alfredo Pulvirenti

DOTTORATO XXV CICLO

SOMMARIO

1. MicroRNA	8
1.1. Introduzione	8
1.2. Organizzazione dei microRNA nel genoma umano	9
1.3. Biogenesi dei microRNA	10
1.4. Regolazione post-trascrizionale mediata dai microRNA	13
1.5. Regolazione dell'espressione dei microRNA	14
1.6. Predizione Bioinformatica dei target molecolari dei microRNA	15
1.7. MicroRNA circolanti	16
1.7.1. miRandola	17
1.7.2. miRandola BackEnd	17
1.7.3. miRandola FrontEnd	19
1.7.4. miRandola – miRò	23
1.7.5. Utilizzo di miRandola	28
1.7.6. Prospettive Future	29
1.8. I microRNA e le malattie	30
1.8.1. microRNA nel Cancro	30
1.8.2. microRNA nell'Alzheimer	33
1.8.3. microRNA nel Diabete	34
1.8.4. microRNA nella Sclerosi Multipla	36
1.9. microRNA, un futuro approccio terapeutico	37
2. Genotipo/Fenotipo	38
2.1. Genotipo Definizioni	38
2.1.1. Geni e Proteine	39
2.1.2. Annotazione Funzionale: Gene Ontology (GO)	39
2.2. Fenotipo Definizioni	40
2.2.1. Mammalian Phenotype Ontology (MP)	41
2.3. Relazione Genotipo + Fenotipo	43
2.3.1. Fenotipi Mendeliani	43
2.3.2. Trattati Complessi	43
2.3.3. Associazione Genotipo – Fenotipo	44

2.4. Analisi dei Fenotipi: Dati e Tool	44
2.4.1. PhenoGO	47
2.4.1.1. Metodi	48
2.4.2. PhenicDB	49
2.4.2.1. Phenocustering	50
2.4.3. CAESAR	51
2.4.3.1. Text-Mining	51
2.4.3.2. Data-Mining	52
2.4.3.3. Data-Integration	52
2.4.4. MimMiner	54
2.4.4.1. Metodi	54
3. Drug	57
3.1. Drug Discovery	57
3.1.1. System Biology in Drug Discovery	57
3.1.2. Approccio corrente nella Drug Discovery	59
3.1.3. Drug Database	61
3.2. Side Effect	62
3.2.1. SIDER 2: Side Effect Resource	64
3.3. Systems Pharmacology	65
3.4. Network Analysis	70
3.5. Problemi e Ricerca	74
3.6. ESHMUN	75
3.6.1. Introduzione	75
3.6.2. ESHMUN BackEnd	77
4. ESHMUN – TSC	81
4.1. TSC1/TSC2	81
4.2. Funzioni principali delle proteina amartina e tuberina	82
4.3. Difetti genici, mutazioni e quadri clinici	84
4.4. Fenotipo Sclerosi Tuberosa	85
4.4.1. Cute	86
4.4.2. Occhio	87
4.4.3. Lesioni Orali	87
4.4.4. Cuore	87
4.4.5. Sistema Vascolare	88
4.4.6. Rene	88

4.4.7. Sistema Nervoso Centrale	89
4.4.7.1. Manifestazioni neurologiche e psichiatriche	89
4.4.8. Apparato Respiratorio	90
4.5. Terapia	90
4.5.1. Rapamicina	90
4.5.2. Vigabatrin	93
4.6. ESHMUN e caso studio	95
4.7. Conclusioni	114
Bibliografia	116

SOMMARIO FIGURE

Figure 1: organizzazione genomica e struttura dei geni miRNA. (a) miRNA intronico in una unità trascrizionale codificante proteina (tu). i triangoli verdi indicano la locazione di uno stem-loop e gli esoni sono mostrati in giallo. (b) miRNA intronici in un trascritto non codificante. (c) la struttura del miRNA esonico in trascritti non codificanti. [11]	9
Figure 2: modello per la biogenesi e per l'attività di soppressione trascrizionale dei microrna [13].....	11
Figure 3: rappresentazione di alcuni dei possibili meccanismi d'azione del complesso risc indotto dai mirna. il messaggero target può essere deadelinato e degradato, oppure può essere inibita la sua traduzione, in diversi modi qui rappresentati [13]	14
Figure 4: Schema della table miRandola relative ai mirnas, samples, articles ed experiments.....	18
Figure 5: Le table mirna_converter e submission	19
Figure 6: Homepage miRandola.....	20
Figure 7: Search Page	21
Figure 8: Esempio Pagina Risultati.....	21
Figure 9: Pagina Advanced Search	22
Figure 10: Page Tools	23
Figure 11: link tra mirandola e mirò (diseases, functions, processes, tissues).....	24
Figure 12: pagina di mirò relativa alle patologia di hsa-mir-21	25
Figure 13: pagian di mirò relativa alle funzioni di hsa-mir-21	26
Figure 14: pagina di mirò relativa ai processi di hsa-mir-21	27
Figure 15: pagina di mirò relativa ai tessuti di hsa-mir-21	27
Figure 16: Utilizzo del db mirandola	28
Figure 17: statistiche utilizzo db mirandola	29
Figure 18: Localizzaione dei microRNA in CAGR (cancer associated genome regions) [56]	31
Figure 19: MicroRNAs che agiscono come oncogeni e oncosoppressori e relative tumori in cui sono coinvolti [56]	32
Figure 20: Pattern alterati di espressione del miRNA ne parenchyma cerebrale e fluido cerebrospianle di pazienti con l'Alzheimer.....	34
Figure 21: Molti miRs sono stati identificati essere critici in diversi processi metabolici, dis-regolazioni che portano al diabete e le sue complicazioni. [71]	35
Figure 22: Differenti concetti per un genotipo a differenti livelli di granularità della composizione genetica di un individuo	38
Figure 23: Esempio GO [91]	40
Figure 24: Pagina di HP	42
Figure 25: Pagina di MP	42

Figure 26: relazione fenotipo/genotipo	44
Figure 27: Fonti di dati fenotipici	45
Figure 28: Home Page di OMIM	46
Figure 29: Pagina di ricerca dei fenotipi in GenAtlas	46
Figure 30: Home Page di HGMD	47
Figure 31: Home page di PharmKGB	47
Figure 32: Diagramma Sistema PhenoGO	48
Figure 33: Annotazioni, per specie, contenute nel database PhenoGO	49
Figure 34: percentuale di geni con uno o più fenotipi della specie data (Ce, C. Elegans; dm, D. melanogaster; hs, H. sapiens; mm, M. musculus; Sc, s. cerevisiae; dr, d. rerio) è mostrata in blu ('direct phenotypes'), di cui uno o più fenotipi associati dalle ortologie sono mostrate in rosso ('orthologous phenotypes'), e geni con nessun fenotipo associato sono mostrati in giallo.	50
Figure 35: Caesar: a) il text-mining è usato per estrarre i simboli dei geni e i termini ontologici dall'input. b) otto tipi di informazioni funzionali sono interrogati usando i geni estratti e anatomia, fenotipi e i termini dei geni ontologici.	53
Figure 36: esempio dell'espansione del concetto usando la struttura mesh gerarchica. il concetto 'photoreceptors' espanso restituisce l'iperonimo 'retina' che a sua volta è iperonimo di 'eye'	55
FIGURE 37: ciclo dei sistemi biologici per generare nuova conoscenza	58
Figure 38: (a) Numero di target per drug (b) Numero di drug per target	60
Figure 39: sito drugbank	61
Figure 40: sito supertarget	61
Figure 41: RELAZIONE TRA L'INCIDENZA DEGLI EFFETTI COLLATERALI DELLA DRUG E LA DISTANZA NETWORK DEI TARGET DELLA DRUG E DEI GENI DELLE MALATTIE.	63
Figure 42: STATISTICHE DEL DATABASE. (A) IL NUMERO DI EFFETTI COLLATERALI PER OGNI DRUG (B) NUMERO DI DRUG PER EFFETTI COLLATERALI	64
Figure 43: ASSOCIAZIONE TRA LE CLASSI DELLE DRUG E LE CLASSI ANATOMICHE DEGLI EFFETTI COLLATERALI.....	65
Figure 44: relazioni tra il genome, proteome, diseasome e drugome	66
Figure 45: analisi multi-scala nella system pharmacology.....	67
Figure 46: le relazioni tra medicina genomica e system pharmacology. il diagramma riassume i vari aspetti della medicina genomica (in blu) e la system pharmacology (in giallo), gli aspetti in overlap dell'analisi sono in verde.....	69
Figure 47: relazioni differenti tra nodi nelle network usate nella system pharmacology. a) nodi sono proteine connesse da interazioni fisiche. b) i nodi ovali sono drug connesse da target condivisi all'interno del dasatinib, mostrato in un triangolo. c) i nodi ovali sono drug connesse condividendo un'indicazione terapeutica	70

Figure 48: geni essenziali e disease nell'interattoma. a) dei 25.000 geni umani, 2.418 sono associati a specifiche malattie, overlap tra 1.777 geni associati alla disease b) diagramma schematico delle differenze tra geni disease essenziali e non essenziali	72
Figure 49: diagramma schematico dei tre concetti di modularità, topological, functional e disease module	72
Figure 50: schema del funzionamento di eshmun	77
Figure 51: tabella mysql per la conversione input	78
Figure 52: tabella mysql dei codici hgnc e sinomini.....	78
Figure 53: connessione delle tabelle mysql di eshmun	80
Figure 54: pathway della tsc [174].....	84
Figure 55: mTOR signaling pathway	91
Figure 56: pi3k-akt signaling pathway.....	92
Figure 57: p53 signaling pathway	92
Figure 58: gabaergic synapse.....	94
Figure 59: uploAD file di lista genica	95
Figure 60: textarea per l'inserimento della lista dei geni	95
Figure 61: select pathway di kegg	96
Figure 62: sezione di output nei risultati di eshmun	99
Figure 63: esempio dei geni trovati relativi alla tsc.....	100
Figure 64: sezione delle network generate	100
Figure 65: heatmap fenotipica della tsc	101
Figure 66: network gene-gene della tsc. I nodi sono i geni gli archi la correlazione fenotipca	102
Figure 67: network geni-mirna validati (TSC). dettaglio del gene PTen e miRNA validati...	105
Figure 68: Schema dell'azione del mir-21	106
Figure 69: network geni-mirna predetti (TSC). dettaglio del gene SOS1 e mirna predetti..	107
Figure 70: network geni-drug (TSC). dettaglio del gene mtor con le relative drug	108
Figure 71: network gene-gene con archi passati sulle interazione proteina-proteina (TSC)	109
Figure 72: sezione delle clique massimali generate relative alla tsc	110
Figure 73: drug similarity DB0087 (sirolimus), maggior similarità con db00864 (tacrolimus) e db00337 (pimecrolimus)	111
Figure 74: TF dei geni (TS)	111
Figure 75: TF dei miRNA (TSC)	112
Figure 76: Sezione per il calcolo degli elementi comuni tra set di geni e calcolo pvalue..	113
Figure 77: Sezione download file	113

CAPITOLO 1:

MICRORNA

1. **MicroRNA**

1.1. **Introduzione**

I microRNA (miRNA) sono un'ampia classe di piccoli RNA non codificanti di circa 21-25 nucleotidi che regolano negativamente l'espressione genica a livello post-trascrizionale, inducendo la degradazione di specifici RNA messaggeri (mRNA), o impedendone la traduzione in proteina [1] [2] [3]. In particolare, i miRNA agiscono mediante riconoscimento di specifici mRNA *target* al fine di determinare la degradazione o la repressione della traduzione.

Da un punto di vista funzionale esistono numerosi miRNA capaci di riconoscere più *target*, molti di questi possono essere regolati da diversi miRNA. Tali affermazioni considerano i singoli miRNA come piccoli elementi di controllo di più complesse *pathway* regolatrici, che stanno alla base di numerose e fondamentali funzioni, da quelle prettamente cellulari, tra cui la regolazione del ciclo cellulare, la proliferazione cellulare, il differenziamento e l'apoptosi, per poi coinvolgere alcuni processi che riguardano l'intero organismo, comprendenti lo sviluppo embrionale e la risposta immunitaria, il metabolismo e la cariogenesi, lo sviluppo e funzione del sistema nervoso e immunitario [4] [5]. I miRNA svolgono le loro funzioni in modo fisiologico, poiché presentano degli specifici e dinamici *pattern* di espressione temporale e spaziale. Tuttavia nel caso in cui la loro espressione risulti essere alterata, i miRNA possono essere coinvolti in diverse malattie complesse, tra cui numerosi tumori. Inoltre i miRNA possono avere profili di espressione specifici per stadi di sviluppo, tessuti e varie patologie. Ciò comporta che ciascun tessuto è caratterizzato da uno specifico set di miRNA, il cui profilo di espressione è distintivo di quel tessuto [6].

Negli ultimi anni, numerosi studi hanno dimostrato il coinvolgimento dei miRNA in diverse patologie, cardiovascolari, come nell'ipertrofia cardiaca [7], neurologiche [8], obesità e diabete [9], ma soprattutto nel cancro [10].

1.2. Organizzazione dei microRNA nel genoma umano

Tutti i miRNA scoperti, sia mediante un approccio sperimentale sia mediante analisi computazionali, sono depositati in miRBase database (miRBase 2012) un vero e proprio registro dei miRNA, annotandone tutte le caratteristiche, che sono accessibili online (<http://www.mirbase.org/index.shtml>). All'ultima versione di miRBase, la release 18 contiene 18.226 *entry* che rappresentano i precursori hairpin dei miRNA e che esprimono 21.643 miRNA maturi in 168 specie, di cui 1.921 maturi e 1.527 precursori sono identificati nell'*homo sapiens*.

miRBase espone anche informazioni circa la genomica dei miRNA, ovvero la loro organizzazione sui cromosomi umani, infatti i geni che codificano miRNA sono distribuiti su tutti i cromosomi, eccetto che sul cromosoma Y, ed in buona parte dei casi geni per miRNA differenti sono localizzati in *loci* adiacenti, creando dei veri e propri *cluster* che molto spesso sono correlati tra loro anche dal punto di vista funzionale, come esempio il *cluster* dei miRNA hsa-let-7a-1, hsa-let-7f-1 e hsa-let-7d nel cromosoma 9. A seconda della localizzazione dei miRNA nel genoma, essi possono essere distinti in:

- o miRNA intergenici
- o miRNA intronici in trascritti codificanti
- o miRNA intronici in trascritti non codificanti
- o miRNA esonici in trascritti non codificanti

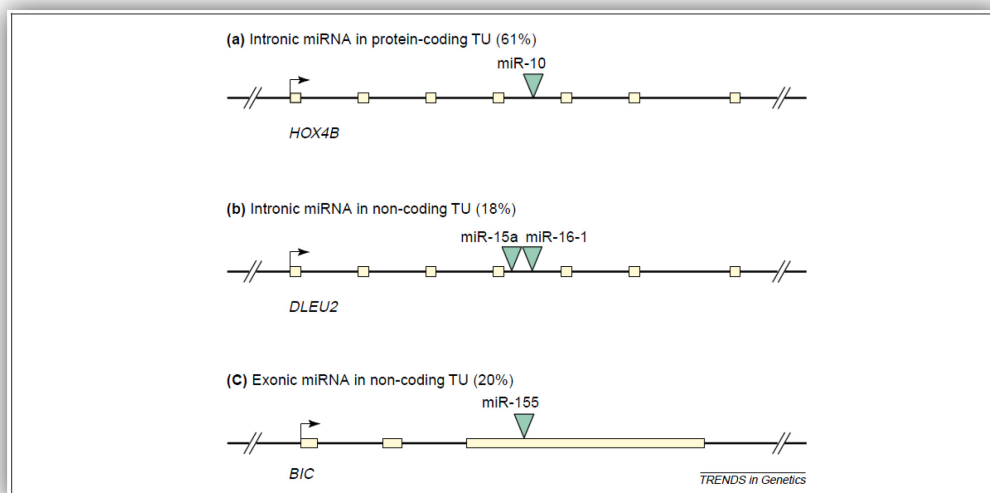


FIGURE 1: ORGANIZZAZIONE GENOMICA E STRUTTURA DEI GENI MIRNA. (A) MIRNA INTRONICO IN UNA UNITÀ TRASCRIZIONALE CODIFICANTE PROTEINA (TU). I TRIANGOLI VERDI INDICANO LA LOCAZIONE DI UNO STEM-LOOP E GLI ESONI SONO MOSTRATI IN GIALLO. (B) MIRNA INTRONICI IN UN TRASCRITTO NON CODIFICANTE. (C) LA STRUTTURA DEL MIRNA ESONICO IN TRASCRITTI NON CODIFICANTI. [11]

1.3. Biogenesi dei microRNA

I miRNA sono presenti nelle piante, negli eucarioti superiori ed in alcuni virus, e sono codificati da diversi tipi di geni, i miRNA attivi nella regolazione dei loro mRNA bersaglio (*target*) sono definiti miRNA "maturi". Questi piccoli RNA sono lunghi da 19 a 22 nucleotidi e si formano mediante un lungo processamento di sequenze ribonucleotidiche più grandi, codificate dal genoma stesso. I miRNA sono codificati da geni propri localizzati nel genoma, singolarmente o in *cluster* [12]. Generalmente, circa il 70% dei geni per i miRNA si trovano in regioni intergeniche, mentre il 30% risiede in sequenze introniche di specifici geni, definiti "ospite". Quindi i miRNA possono essere situati in unità trascrizionali indipendenti, ma comunque, un gran numero sono generati da trascritti contenenti o *cluster* di miRNA o sequenze introniche del gene ospite.

Il processo mediante il quale si giunge alla formazione dei miRNA maturi è piuttosto complesso e si compone di numerose tappe: esso ha origine nel nucleo e giunge a completamento a livello del citoplasma.

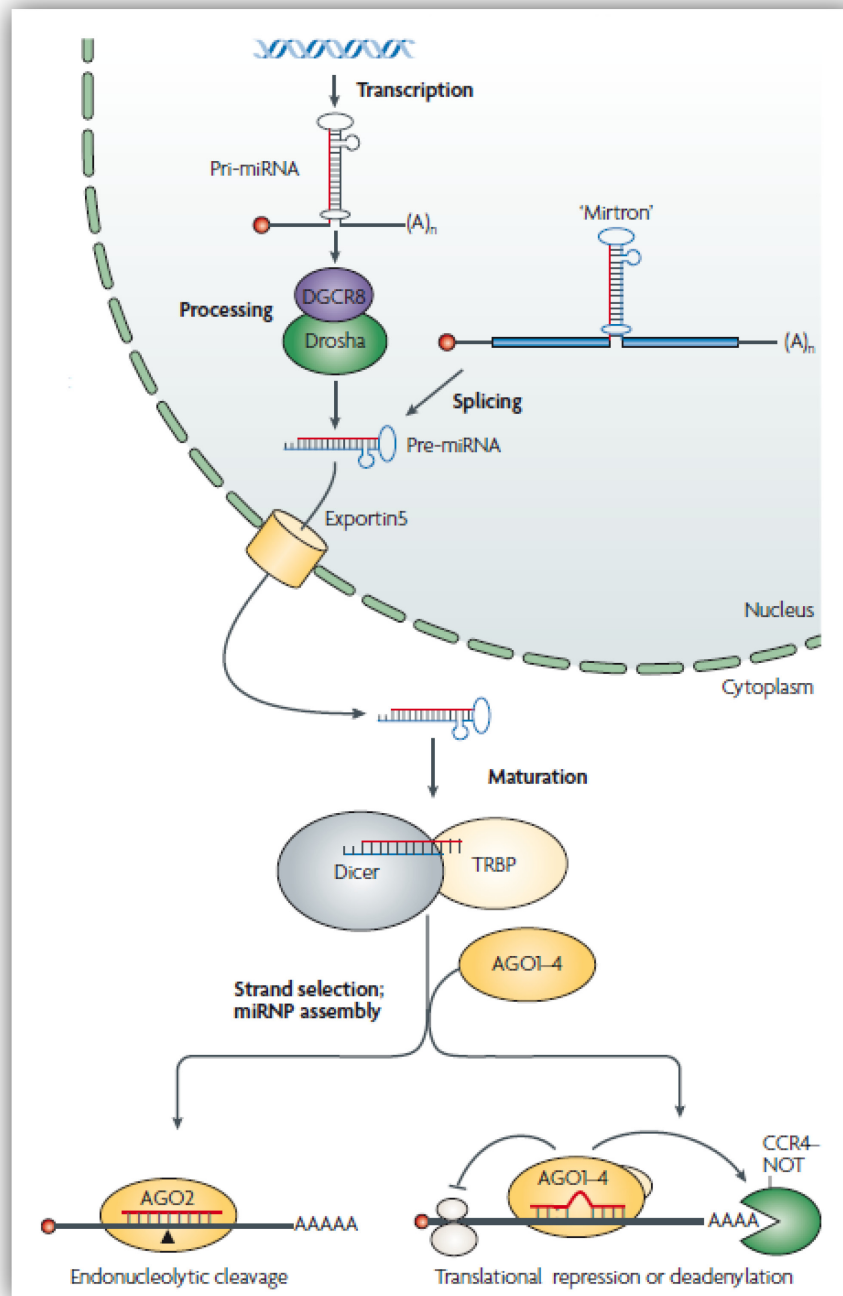


FIGURE 2: MODELLO PER LA BIOGENESI E PER L'ATTIVITÀ DI SOPPRESSIONE TRASCRIZIONALE DEI MICRORNA [13].

Questa molecola subisce due reazioni consecutive catalizzate da due endonucleasi, Drosha e Dicer, entrambe agiscono all'interno di complessi insieme a proteine contenenti domini in grado di legare molecole di RNA a doppio filamento (*dsRNA binding domains – dsRBDs*). Le proteine Drosha e Dicer possiedono degli specifici domini catalitici RNasi di tipo III (RIIDa, RRIIDb) conservati, che agiscono generando delle estremità 3' sporgenti di 2 nucleotidi.

La prima reazione avviene nel nucleo ad opera di Drosha, che forma un complesso con la proteina Pasha (DGCR8), facente parte della stessa famiglia. Il complesso Drosha/Pasha scinde il pri-miR generando una molecola con struttura a forcina di circa 70 nucleotidi, il pre-miR. Dopo la reazione iniziale da parte di Drosha il pre-miR viene esportato dal nucleo al citoplasma ad opera dell'Esportina 5 (Exp5), un trasportatore nucleo/citoplasmatico GTP-dipendente. Una volta all'interno del citoplasma, il precursore a forcina viene tagliato da Dicer con formazione di una piccola molecola di RNA duplex con appaiamenti imperfetti di dimensione variabile tra 21 e 25 nucleotidi, contenente sia il filamento maturo del miRNA che il suo filamento complementare [14].

Nei mammiferi il Dicer interagisce con le proteine Ago1-Ago4, le quali assemblandosi con il miRNA maturo formano il complesso miRISC (*microRNA Induced Silencing Complex*) che guiderà il miRNA nel riconoscimento del messaggero *target*. Le proteine Ago fanno parte della famiglia delle Argonate, che sono presenti in tutti gli eucarioti e dotate anch'esse di motivi specifici: i domini PAZ e MID per l'ancoraggio dell'RNA target al 3' e 5' e il dominio PIWI per il taglio. Nel complesso del miRISC, oltre alle proteine Ago, sono state trovate proteine della famiglia GW182 (TNRC6A, TNRC6B, TNRC6C nei mammiferi) che hanno un ruolo importante nella repressione trascrizionale mediata dai miRNA [13] e agiscono come cofattori delle Ago. Al momento in cui il duplex è stato formato in seguito all'azione di Dicer, i due filamenti presentano una stabilità differente da un punto di vista termodinamico a livello dell'estremità 5'; sebbene il miRNA maturo possa essere localizzato in uno o nell'altro filamento in un modo indistinto, esso si origina quasi sempre dal filamento che presenta l'estremità 5' più instabile, mentre l'altro filamento verrà degradato [15]. Tuttavia nei rari casi in cui le due estremità 5' presentano una stabilità paragonabile, ognuno dei due filamenti può con la stessa probabilità andare a costruire il miRNA maturo dotato di attività biologica [1].

La regolazione della biogenesi dei miRNA è chiaramente un meccanismo importante ma non ancora studiato in maniera esaustiva. Emerge tuttavia una tendenza significativa: un sorprendente numero di geni miRNA sono formati sotto il controllo dei molti *target* che regolano. Ad esempio, la trascrizione del gene miR-7 in *Drosophila* è repressa da un fattore di trascrizione chiamato Yan, la cui traduzione è a sua volta repressa da miR-7, dando luogo ad un *feedback-loop* negativo [16]. Un altro esempio si ha in *C. Elegans*, dove il miRNA let-7 inibisce la traduzione di lin-28 che a sua volta inibisce la trascrizione di let-7 [17].

La logica alla base di queste relazioni è definita dalla capacità di regolazione della biogenesi dei miRNA. L'espressione errata dei miRNA mima frequentemente il fenotipo da perdita di funzione dei loro *target*, ciò può essere prevenuto se l'espressione dei miRNA è strettamente controllata dagli stessi *target*.

1.4. **Regolazione post-trascrizionale mediata dai microRNA**

Una volta conclusa la formazione del miRISC, contenente il miR-maturo, avviene l'appaiamento di questo con il messaggero target, quindi i miRNA agiscono come degli adattatori per i complessi miRISC per riconoscere determinati mRNA *target*. I siti di legame dei miRNA nei mRNA animali si trovano nei 3'UTR e solitamente sono espressi in copie multiple, questa è una condizione necessaria affinché si abbia un'efficiente repressione della traduzione [18]. È stato tuttavia osservato *in vitro* che il riconoscimento miRNA-mRNA avverrebbe anche nelle regioni codificanti o nel 5' UTR dell'mRNA anche se questi siti di appaiamento non avrebbero sufficiente capacità di silenziamento e svolgerebbe solo un ruolo marginale [19]. La maggior parte dei miRNA animali si legano ai loro target con complementarietà imperfetta, formando *bulge* e *loop*, sebbene una caratteristica chiave del riconoscimento del *target* coinvolga l'appaiamento dei nucleotidi 2-8 del miRNA, che rappresentano la regione *seed*. Al contrario, nella maggior parte delle piante i miRNA si legano con complementarietà quasi perfetta a siti specifici presenti nella regione codificante [12].

Il grado di complementarietà miRNA/*target* è considerato un fattore chiave del meccanismo regolatorio. La complementarietà perfetta permette il taglio del filamento dell'mRNA catalizzato da Ago, mentre i *mismatch* centrali del duplex miRNA/mRNA escludono il taglio e promuovono la repressione della traduzione. Tuttavia, esperimenti suggeriscono che il meccanismo predefinito di silenziamento dell'espressione genica operata dai miRNA, sia negli animali che nelle piante, sia la repressione traduzionale, che la complementarietà perfetta siano un fatto aggiuntivo che può condurre al taglio dell'mRNA, in modo che l'effetto finale sia il risultato di entrambi i meccanismi.

Diversi studi sui miRNA animali indicano che la repressione della traduzione non è accompagnata dalla destabilizzazione dell'mRNA. Tuttavia, per alcune interazioni miRNA-*target* vi è una riduzione significativa della concentrazione dell'mRNA dovuta ad un incremento della degradazione [20] [21]. Questa degradazione non è causata dall'attività catalitica di Ago ma piuttosto da deadenilazione, decapping e digestione esonucleotidica dell'mRNA [21] [22] [23]. Al momento non è chiaro il motivo per cui alcuni *target* vengono degradati ed altri no. È stato ipotizzato che il numero, il tipo e la posizione dei *mismatch* del duplex miRNA/mRNA giochi un ruolo importante nel determinare la degradazione o l'arresto della traduzione [24].

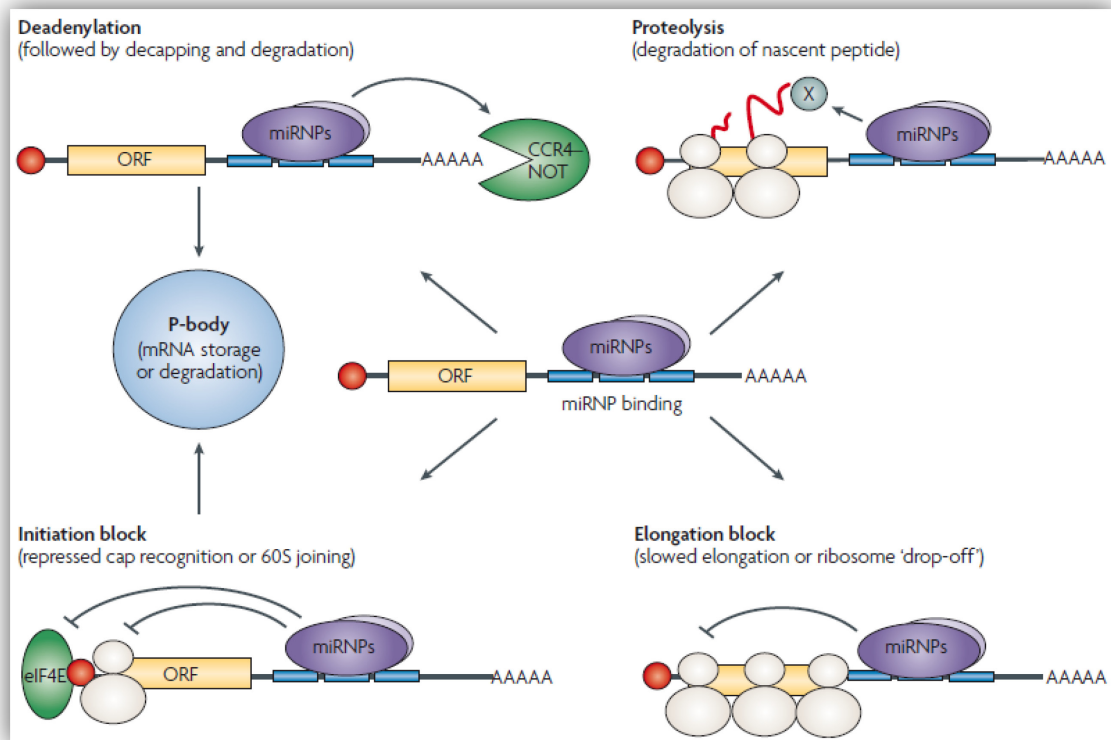


FIGURE 3: RAPPRESENTAZIONE DI ALCUNI DEI POSSIBILI MECCANISMI D'AZIONE DEL COMPLESSO RISC INDOTTO DAI MI RNA. IL MESSAGGERO TARGET PUÒ ESSERE DEADELINATO E DEGRADATO, OPPURE PUÒ ESSERE INIBITA LA SUA TRADUZIONE, IN DIVERSI MODI QUI RAPPRESENTATI [13]

1.5. Regolazione dell'espressione dei microRNA

La regolazione dell'espressione dei miRNA è fondamentale per il ruolo svolto da queste molecole ed è quindi fortemente regolata in più punti durante la loro biogenesi.

1. Regolazione della trascrizione
2. Regolazione del processamento
3. *Editing*
4. *microRNA decay*

La regolazione della trascrizione avviene come molti geni di seconda classe poiché a livello dei promotori dei miRNA sono presenti gli stessi elementi regolativi (*TATA box sequences, CpC Islands, initiation elements*) dei promotori per i geni codificanti proteine. Molti fattori di trascrizione (TF) regolano l'espressione tessuto-specifica o sviluppo-specifica, ad esempio MYC/MYCN che stimolano l'espressione del *cluster* oncogenico miR-17-5p in cellule di linfoma [25], o REST che inibisce l'espressione del miR-124 in cellule non neuronali o progenitori neuronali tramite de-acetilazione istonica e metilazione del promotore [26]. I

miRNA stessi possono a loro volta regolare l'espressione dei TF, creando così dei circuiti di regolazione positiva o negativa nei quali il preciso controllo della quantità del miRNA o del TF determina l'effetto fisiologico finale.

La regolazione del processamento dei miRNA avviene al livello del Drosha, del Dicer e delle loro proteine accessorie. Ad esempio è stato visto come alcune elicasi del topo e le proteine SMAD, trasduttori dei segnali della via del TGF- β e BMP, agiscono a livello del Drosha, controllando la produzione del pri-miR [27]. A livello del pre-miR invece è stato visto, che i livelli di Dicer vengono controllati e stabilizzati dal suo cofattore TRBP. Infatti dopo stimolazione da parte dei fattori di crescita, ERK fosforila TRBP e quindi stabilizza indirettamente il Dicer, portando a un generale aumento di tutti i pre-miR [28].

L'*editing* del pri-miR o del pre-miR da parte delle proteine ADAR (adenosine deaminase), che catalizzano la conversione dell'adenosina in inosina, altererebbe la struttura secondaria e quindi la stabilità, mentre l'*editing* del seed del miRNA maturo da parte di altre proteine altererebbe il riconoscimento del *target* [27]. Infine la regolazione della stabilità e degradazione del miRNA maturo permette di controllarne la quantità finale nella cellula e quindi l'effetto biologico. È stato osservato che i miRNA sono in generale più stabili dei messaggeri di seconda classe e hanno un'emivita che va da alcune ore a molti giorni. In generale comunque il preciso controllo del *decay* potrebbe avere un ruolo fondamentale nei meccanismi di sviluppo e di risposta di tipo *switch on-off*, ad esempio nello sviluppo della retina del topo in cui i livelli dei miR-204 e miR-211 diminuiscono rapidamente nei neuroni ma non nella glia [29]

1.6. Predizione Bioinformatica dei target molecolari dei microRNA

La scoperta dei miRNA ha introdotto un nuovo paradigma nei sistemi di regolazione genica. L'identificazione dei geni regolati dai miRNA è un punto cruciale nella comprensione del ruolo funzionale dei miRNA stessi e delle complesse reti molecolari (*network*) alla base della regolazione genica. La complessità nell'identificazione degli mRNA bersaglio dei miRNA è dovuta innanzitutto al fatto che i miRNA hanno sequenze molto corte e dall'imperfezione del loro appaiamento con il target molecolare. Negli ultimi anni, si è cercato di estrapolare dagli studi sperimentali i principi alla base di questa interazione al fine di sviluppare numerosi algoritmi matematici per la predizione *in silico* degli ipotetici mRNA bersaglio. Tali principi comprendono:

- (i) la complementarità imperfetta tra il miRNA e il 3' UTR del *target* ed il forte legame tra i 6-8 nt della regione *seed* al 5' rispetto al 3';

- (ii) la conservazione evolutiva tra le specie delle sequenze bersaglio al 3' del *target*;
- (iii) la stabilità termodinamica del *duplex* miRNA-mRNA;
- (iv) la cooperatività tra i siti multipli in stretta prossimità;
- (v) molteplicità e cooperatività dell'interazione miRNA-*target*;
- (vi) perdita della struttura secondaria del mRNA *target* al sito di legame per il miRNA.

Alcuni software, accessibili in rete, per la predizione di *target* dei miRNA sono elencati nella tabella. Una parte dei risultati ottenuti con questi algoritmi sono stati validati sperimentalmente, e ciò ha permesso di migliorare notevolmente le prestazioni per la predizione *in silico* dei *target* di miRNA.

<i>Programma</i>	<i>Sito Web</i>	<i>Ref</i>
DIANA-microT	http://www.diana.pcbi.upenn.edu/cgi-bin/micro_t.cgi	[30]
miRanda	http://www.microrna.org/	[31] [32]
MirTarget2	http://mirdb.org/	[33]
RNA22	http://cbcsrv.watson.ibm.com/rna22.html	[34]
RNAhybrid	http://bibiserv.techfak.unizbielefeld.de/rnahybrid/	[35]
TargetScan	http://www.targetscan.org/	[31] [36] [37]

1.7. **MicroRNA circolanti**

I microRNA sono inoltre presenti in fluidi corporei umani extra-cellulari come il siero, plasma, saliva e urine e questo spesso è associato con varie condizioni patologiche incluso il cancro. I microRNA circolanti sono stati trovati all'interno di vescicole chiamate *exosome* (esosomi). Tuttavia, la maggior parte sono presenti nel plasma e siero umano complessati alla proteina Argonaute2 (Ago2), piuttosto che all'interno di vescicole. Poiché i microRNA circolano nel flusso sanguigno in una forma extra-cellulare altamente stabile, essi potrebbero essere usati come biomarcatori non invasivi per patologie come il cancro [38] [39], malattie cardiovascolari [40] e malattia di Crohn pediatrica [41].

Il modello dominante per la stabilità dei miRNA circolanti è quello che i miRNA sono rilasciati dalle cellule in vescicole costituite da membrane, che li proteggono dall'attività della Rnase nel sangue. Le vescicole proposte come *carrier* di miRNA circolanti includono esosomi, che sono vescicole di 50-90-nm derivanti da corpi multi-vescicolari e rilasciate

per esocitosi [42]. Tuttavia, è stato dimostrato che una porzione significativa di miRNA circolanti nel plasma umano e nel siero è associata con Argonaute2 (Ago2) [43] [44]. Ago2 è il componente essenziale del complesso *miRNA-induced silencing complex* (miRISC) che lega direttamente i miRNA e media la repressione dell'RNA messaggero nelle cellule [45] [46]. Sebbene sia stato ipotizzato che i miRNA che si trovano negli esosomi sono coinvolti nella comunicazione intercellulare [47] [48], molti miRNA extra-cellulari potrebbero essere sotto-prodotti di cellule morte che rimangono nello spazio extra-cellulare dovuto all'alta stabilità delle proteina Ago2 e al complesso Ago2-miRNA [44]. Questi risultati recenti suggeriscono che l'analisi dei miRNA come biomarcatori dovrebbe includere tutti i tipi di miRNA circolanti trovati nei fluidi biologici.

1.7.1. **miRandola**

Nonostante lo stato dell'arte della letteratura non distingue tra le differenti forme dei miRNA circolanti, una precisa classificazione è necessaria. miRandola è un database di miRNA extra-cellulari/circolanti [49]. I miRNA sono classificati in quattro categorie, basate sulla loro forma extra-cellulare: i) miRNA-Ago2, ii) miRNA-exosome, iii) miRNA-HDL, iv) miRNA-circulating. Il database fornisce agli utenti una varietà di informazioni includendo le malattie associate, i campioni, i metodi usati per isolare i miRNA e la descrizione dell'esperimento. Le informazioni riguardo i target dei miRNA e le loro annotazioni sono fornite attraverso i link a miRò, "the miRNA knowledge base" [50]. miRò integra dati da differenti fonti per permettere l'identificazione di associazioni tra geni, processi, funzioni e malattie attraverso i target validati e predetti dei miRNA.

1.7.2. **miRandola BackEnd**

miRandola è un database contenente 89 articoli, 2132 entry e 581 unici miRNA maturi, creato in MySQL, costituito da 6 table:

- La table **mirnas**: raccoglie le informazioni relative ai miRNA, come la famiglia, il miRNA maturo ed il tipo di miRNA con i seguenti campi, "mirna_ID" la chiave primaria, "mature_mirna", "mirna_family", "mirna_type", "sample_ID", "article_ID", "experiment_ID", "mirbase_accession", "Last_Version", "AC_code", "potential biomarker"
- La table **samples**: contiene le informazioni riguardanti i campioni in cui sono stati trovati i miRNA, ha i seguenti campi, "sample_ID" la chiave primaria, "sample", "sample_source", "sample_name"

- La **table articles**: riguarda i metadati circa gli articoli da cui sono state estratte le informazioni che compongono miRandola, ha i seguenti campi, "article_ID" la chiave primaria, "author", "journal", "title", "link", "data", "database"
- La **table experiments**: fornisce dettagli circa l'esperimento condotto, ha i seguenti campi: "experiment_ID" la chiave primaria, "organism", "method", "description", "database", "link"

La **table mirnas** è legate alle **table samples**, **articles** ed **experiments** rispettivamente tramite "sample_ID", "article_ID" ed "experiment_ID" questo permette di sapere per ciascun miRNA il campione da cui è stato estratto, con una descrizione dell'esperimento circa l'organismo ed il metodo usati e le fonti da cui provengono le informazioni quali paper, titolo articolo, autori, data di pubblicazione ed il link a cui è reperibile. Maggiori dettagli sono visibili nella Figura 4.

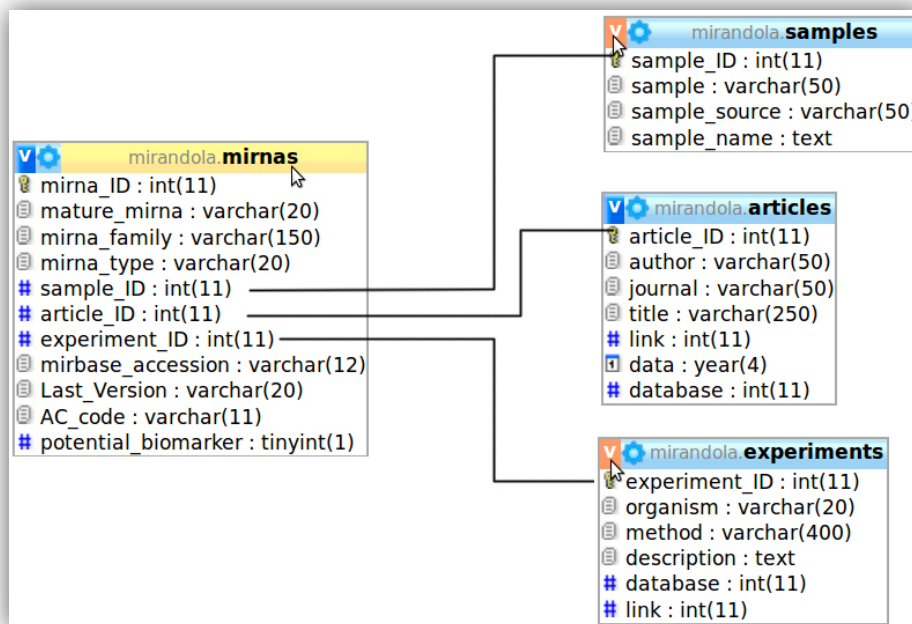


FIGURE 4: SCHEMA DELLA TABLE MIRANDOLA RELATIVE AI MIRNAS, SAMPLES, ARTICLES ED EXPERIMENTS

- La **table mirna_converter**: permette di convertire un mature miRNA qualsiasi dalla versione di miRBase 12 alla versione di miRBase 18, ha i seguenti campi, "id" la chiave primaria, "mirbase_accession", "version_12", "version_13", "version_14", "version_15", "version_16", "version_17", "version_18"
- La **table submission**: che contiene le informazioni di chi vuole inserire un suo miRNA circolante in miRandola, ha i seguenti campi, "submission_ID" e "PubMed" che insieme costituiscono la chiave primaria, "mature_mirna", "mirna_type", "sample", "sample_name", "species", "firstname", "lastname", "affiliation", "email"

mirandola.mirna_converter		mirandola.submission	
mirbase_accession	: varchar(12)	submission_ID	: int(11)
version_12	: varchar(20)	mature_mirna	: varchar(100)
version_13	: varchar(20)	mirna_type	: varchar(100)
version_14	: varchar(20)	sample	: varchar(100)
version_15	: varchar(20)	sample_name	: text
version_16	: varchar(20)	species	: varchar(200)
version_17	: varchar(20)	PubMed	: int(11)
version_18	: varchar(20)	firstname	: varchar(50)
id	: int(11)	lastname	: varchar(50)
		affiliation	: varchar(100)
		email	: varchar(100)

FIGURE 5: LE TABLE MIRNA_CONVERTER E SUBMISSION

1.7.3. *miRandola FrontEnd*

Il frontend di miRandola è realizzato usando HTML, PHP, CSS e JQuery, il sito è disponibile al link: <http://atlas.dmi.unict.it/mirandola>.

Sono disponibili diverse sezioni:

- Homepage (<http://atlas.dmi.unict.it/mirandola/index.html>): contiene una descrizione del sito, con uno schema del database e delle fonti utilizzate oltre che un piccolo box "News" che riporta le ultime notizie.

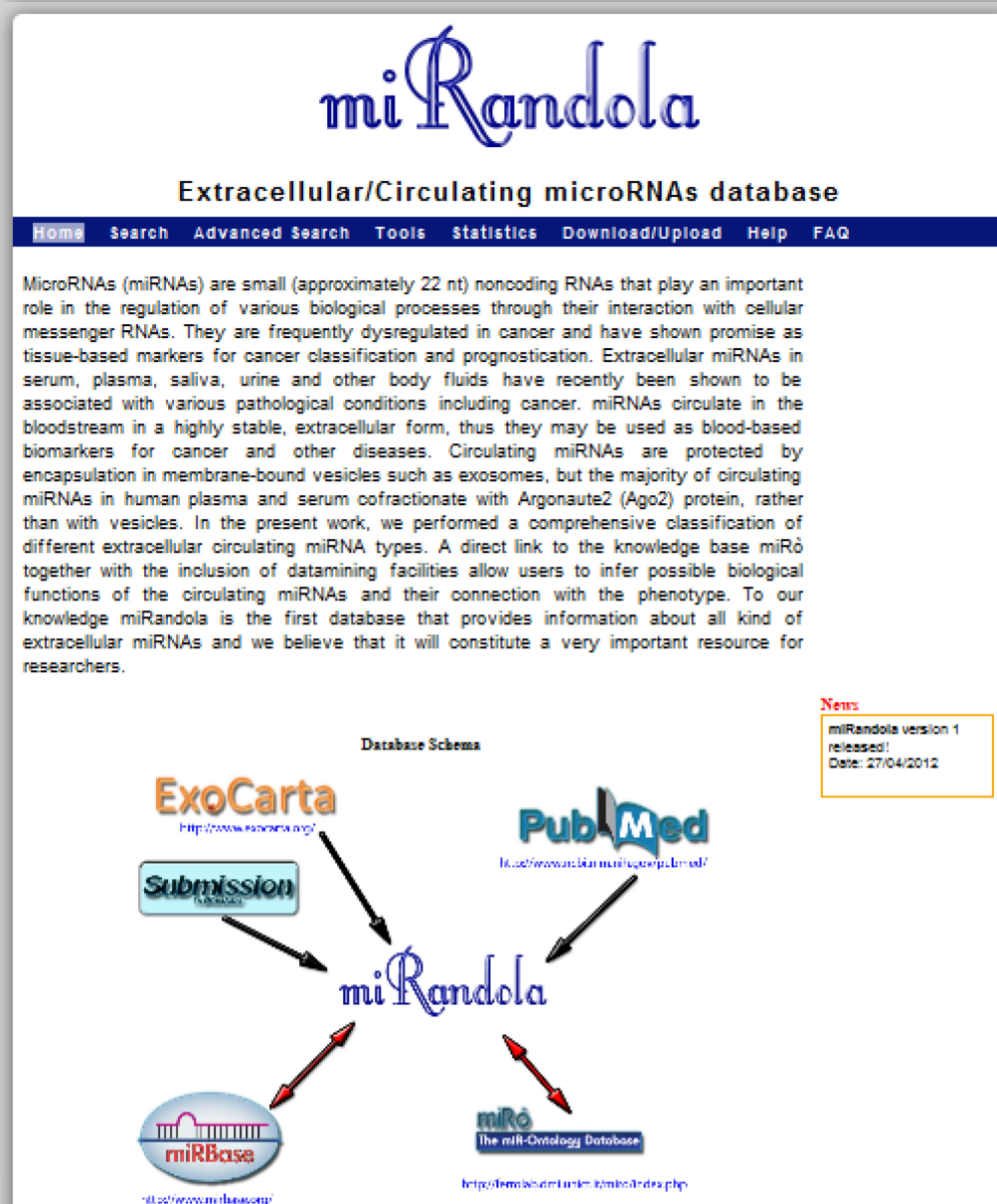


FIGURE 6: HOMPAGE MIRANDOLA

- Search (<http://atlas.dmi.unict.it/mirandola/browse.php>): è possibile cercare un miRNA per categoria: i) *Mature miRNA*, ii) *miRNA Family*, iii) *miRNA type*, iv) *Sample*, v) *Disease and Malignant Cell Lines*, vi) *Potential Biomarker*.

Per ogni categoria la ricerca può avvenire o selezionando una voce dalla *select* oppure digitando nella relativa area l'elemento da cercare (Figura 7).

Per ogni ricerca vengono forniti il numero di risultati trovati e una scheda informativa per ciascun miRNA, infine i risultati possono essere salvati in tre diversi formati pdf, txt e csv.

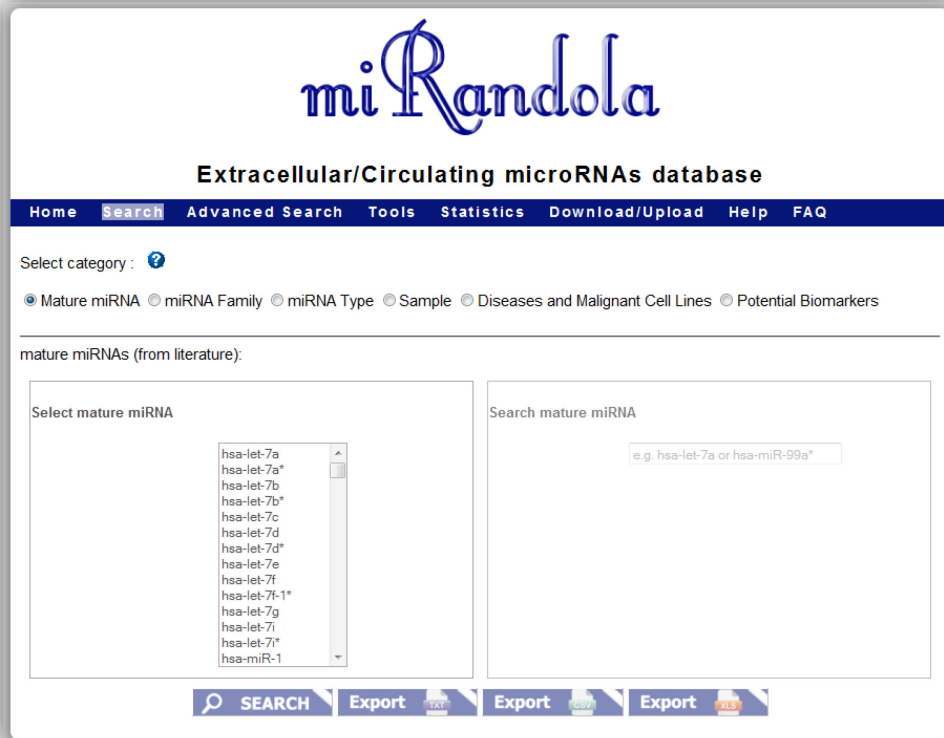


FIGURE 7: SEARCH PAGE

Results found for hsa-let-7a: #10

Each entry in miRandola provides links to miRat knowledge base
Legend:

- miRNA associated Diseases (D)
- miRNA associated Functions (F)
- miRNA associated Processes (P)
- miRNA associated Tissues (T)



Mature miRNA from literature:	hsa-let-7a (D) (F) (P) (T)
Mature miRNA from miRBase:	hsa-let-7a-5p 
miRBase Accession:	MI0000062
miRNA family:	let7
Potential Biomarker:	unknown
miRNA type:	exosome
Sample:	serum
Sample source:	ts
Diseases and Malignant Cell Lines:	papillary adenocarcinoma of ovary
First Author:	Taylor DD et al.
Journal:	Gynecol Oncol.110(1):13-21.
Title:	Microrna signatures of tumor-derived exosomes as diagnostic biomarkers of ovarian cancer.
PubMed ID:	1889210 
Date of Publication:	2008
Methods:	487microrna array
Experiment Description:	Microrna signatures of tumor-derived exosomes as diagnostic biomarkers of ovarian cancer.
Database:	ExoCarta 

FIGURE 8: ESEMPIO PAGINA RISULTATI

- Advanced Search (http://atlas.dmi.unict.it/mirandola/search_ad.php): in questa pagina è possibile fare una ricerca avanzata, dopo aver scelto tra due categorie : i) *miRNA family*, ii) *mature miRNA*, da una *select* si seleziona la voce desiderata e successivamente filtrare la ricerca per *miRNA type* e *sample*.

FIGURE 9: PAGINA ADVANCED SEARCH

- Tools (<http://atlas.dmi.unict.it/mirandola/tools.php>): in questa pagina è possibile convertire la nomenclatura del proprio miRNA maturo in quello di miRBase versione 18, la conversione avviene o usando la *select* oppure digitando nella textarea una lista di miRNA separati da “,”.

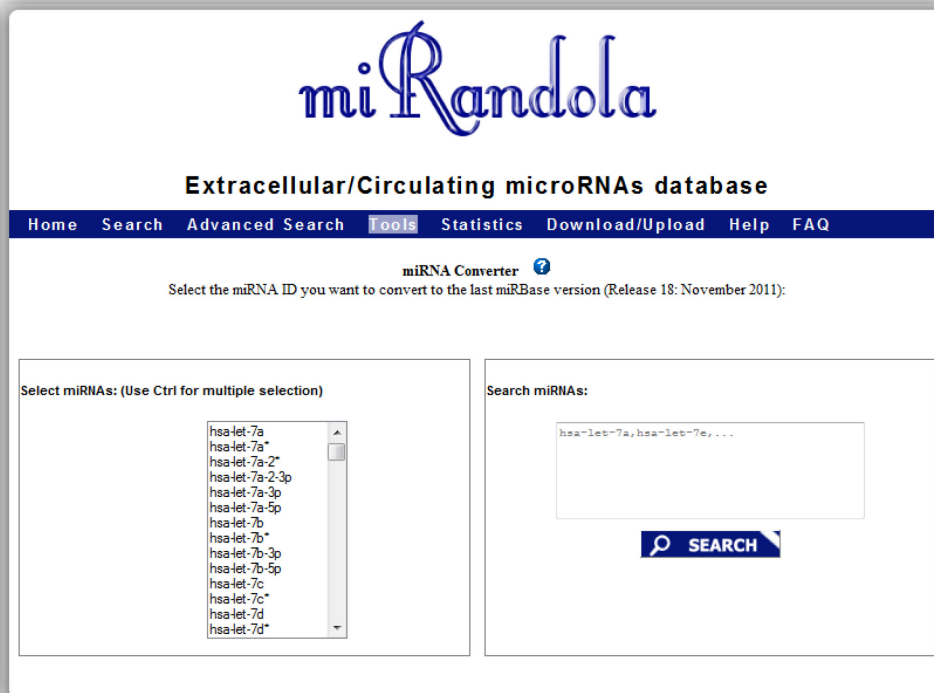


FIGURE 10: PAGE TOOLS

- Statistics (<http://atlas.dmi.unict.it/mirandola/statistics.php>): Qui vengono riportate alcune statistiche relative al database come il numero di pubblicazioni in miRandola, il numero di entry, la distribuzione dei miRNA nei campioni, una wordle dei miRNA più presenti, le riviste più frequenti in miRandola e l'elenco di tutti gli articoli curati manualmente.
- Download/Upload (http://atlas.dmi.unict.it/mirandola/down_up.php): questa pagina permette di inviare un nuovo record da poter inserire in miRandola ed inoltre è possibile fare il download dei dati già presenti nel sistema.
- Help (<http://atlas.dmi.unict.it/mirandola/help.html>): fornisce un utile tutorial sull'utilizzo di miRandola.
- FAQ (<http://atlas.dmi.unict.it/mirandola/faq.html>): questa pagina contiene le risposte alle domande più frequenti relative a miRandola.

1.7.4. miRandola – miRò

La funzione dei miRNA circolanti è ancora largamente ignota, e sono stati riportati casi in cui miRNA endogeni possono essere trasportati da High-density lipoprotein (HDL) e in

questo modo riescono ad entrare all'interno di cellule riceventi e contribuire alla repressione dei loro target [51].

Gli exosome, inoltre, sembrano svolgere un importante ruolo nello sviluppo delle metastasi, comunque la funzione dei miRNA nel targeting di siti distanti dall'organo primario da cui hanno avuto origine è ancora ignota [52].

Per tutti questi motivi, per aiutare a formulare ipotesi sulla funzione dei miRNA che vengono secreti nello spazio extracellulare, sono stati realizzati collegamenti tra miRandola e miRò, che fornisce all'utente informazioni su annotazioni funzionali attraverso i target validati e predetti dei miRNA.

Ogni entry in miRandola fornisce link a malattie, processi e funzioni in cui ogni miRNA è coinvolto e i tessuti in cui esso è espresso.

The screenshot shows a user interface for miRandola. On the left, there is a dark blue sidebar with the number '1' and the text 'Mature miRNA from literature:'. The main content area has a white background with a blue header that reads 'Each entry in miRandola provides links to miRo' knowledge base'. Below this is a legend with four items: 'miRNA associated Diseases (D)' in red, 'miRNA associated Functions (F)' in blue, 'miRNA associated Processes (P)' in green, and 'miRNA associated Tissues (T)' in black. At the bottom, there is a light blue bar containing the text 'Mature miRNA from literature: hsa-miR-21 (D) (F) (P) (T)'. The letters (D), (F), (P), and (T) are color-coded to match the legend.

FIGURE 11: LINK TRA MIRANDOLA E MIRÒ (DISEASES, FUNCTIONS, PROCESSES, TISSUES)

Dopo aver cliccato sulla lettera D (in rosso), l'utente verrà indirizzato alla pagina di miRò riferita a tutte le patologie associate allo specifico miRNA (nell'esempio hsa-miR-21):

miRò
The miR-Ontology Database

By miRNA: **hsa-miR-21** [miRò Homepage](#)

Click on miRNA name to see details

hsa-miR-21 associated Broad Phenotypes (Diseases) Order by Broad Phenotypes (Diseases) Name

Click on gene name or sources to see details Order by Gene Name

Show page

Broad Phenotypes (Diseases): 3198 entries

Broad Phenotypes (Diseases)	Target Gene	Sources			
		Validated	TargetScan	miRanda	PicTar
retinal dystrophy	RP2	✓	✓	✓	✓
retinitis pigmentosa	RP2	✓	✓	✓	✓
diabetes, type 1	SOC5	✓	✓	✓	
heart anomalies, congenital	BMPR2	✓	✓	✓	
hypertension	BMPR2	✓	✓	✓	
juvenile polyposis	BMPR2	✓	✓	✓	
pulmonary hypertension	BMPR2	✓	✓	✓	
vasoreactivity	BMPR2	✓	✓	✓	
clubfoot	APAF1	✓	✓		
melanoma	APAF1	✓	✓		
Alzheimer's Disease	FAS	✓			
arthritis lupus erythematosus	CDKN1A	✓			

FIGURE 12: PAGINA DI MIRÒ RELATIVA ALLE PATOLOGIA DI HSA-MIR-21

In questa immagine è possibile osservare il risultato restituito da miRò. Ad ogni malattia sono associati uno o più geni (i target dei miRNA) validati o predetti da diversi programmi di predizione. Se l'utente clicca, invece, sulla lettera F (in blu) verrà indirizzato alla pagina relativa alle funzioni biologiche associate al miRNA:

miRò
The miR-Ontology Database

By miRNA: **hsa-miR-21** miRò Homepage

Click on miRNA name to see details

miR-Function Order by GO Term

Click on ontology or gene name or sources to see details Order by Gene Name

Show page

Functions: 5789 entries

Function	Target Gene	Sources			
		Validated	TargetScan	miRanda	PicTar
ATP binding	RP2	✓	✓	✓	✓
binding	RP2	✓	✓	✓	✓
nucleoside diphosphate kinase activity	RP2	✓	✓	✓	✓
protein binding	PDCD4	✓	✓	✓	✓
RNA binding	PDCD4	✓	✓	✓	✓
transcription factor activity	NFIB	✓	✓	✓	✓
unfolded protein binding	RP2	✓	✓	✓	✓
protein binding	BTG2	✓	✓		✓
actin binding	CFL2	✓	✓	✓	
ATP binding	BMPR2	✓	✓	✓	
ATP binding	CDK6	✓	✓	✓	
cyclin binding	CDK6	✓	✓	✓	

FIGURE 13: PAGIAN DI MIRÒ RELATIVA ALLE FUNZIONI DI HSA-MIR-21

Cliccando sulla lettera P (in verde) verranno visualizzati tutti i processi biologici associati al miRNA:

miRò
The miR-Ontology Database

By miRNA: hsa-miR-21 miRò Homepage

Click on miRNA name to see details

miR-Process Order by GO Term

Click on ontology or gene name or sources to see details Order by Gene Name

Show page 1

Processes: 5403 entries

Process	Target Gene	Sources			
		Validated	TargetScan	miRanda	PicTar
apoptosis	PDCD4	✓	✓	✓	✓
beta-tubulin folding	RP2	✓	✓	✓	✓
cell aging	PDCD4	✓	✓	✓	✓
CTP biosynthetic process	RP2	✓	✓	✓	✓
DNA replication	NFIB	✓	✓	✓	✓
forebrain development	NFIB	✓	✓	✓	✓
GTP biosynthetic process	RP2	✓	✓	✓	✓
hindbrain development	NFIB	✓	✓	✓	✓
lung development	NFIB	✓	✓	✓	✓
negative regulation of cell cycle	PDCD4	✓	✓	✓	✓
negative regulation of cell proliferation	NFIB	✓	✓	✓	✓
negative regulation of JNK activity	PDCD4	✓	✓	✓	✓

FIGURE 14: PAGINA DI MIRÒ RELATIVA AI PROCESSI DI HSA-MIR-21

Infine, cliccando sulla lettera T (in nero) l'utente potrà visionare i livelli di espressione del miRNA in diversi tessuti:

miRò
The miR-Ontology Database

By miRNA: hsa-miR-21 miRò Homepage

Click on miRNA name to see details

hsa-miR-21 associated Tissues Order by Tissue Name

Click on gene name or sources to see details Order by Expression Value

Show page 1

Tissues: 161 entries

Tissue	Library	Expression
	hsa_USSC-d7	0.67% - 59/8820
Adeno-CA breast, epithelial like	hsa_Breast-adenocarcinoma-MCF7	0.66% - 58/8820
AML	hsa_AML1-d29	0.12% - 11/8820
AML	hsa_AML-THP1	0.15% - 13/8820
AML	hsa_AML3-d0	0.39% - 34/8820
AML	hsa_AML1-d0	0.26% - 23/8820
AML	hsa_AML2-d0	0.17% - 15/8820
AML	hsa_AML-HL60	0.01% - 1/8820
AML	hsa_AML3-d29	0.1% - 9/8820
Astroblastoma	hsa_Astroblastoma-DD040800	1.52% - 134/8820
B-ALL	hsa_B-ALL8-d43	0.05% - 4/8820
B-ALL	hsa_B-ALL2-d0	0.1% - 9/8820

FIGURE 15: PAGINA DI MIRÒ RELATIVA AI TESSUTI DI HSA-MIR-21

1.7.5. Utilizzo di miRandola

Questa immagine rappresenta la distribuzione dei visitatori del database in tutti i continenti. Come è possibile notare, dalla diversa intensità del colore, miRandola è stato prevalentemente utilizzato in Europa e i dati di seguito riportati sono riferiti all'uso del database prima della sua pubblicazione nella rivista PLoS ONE (in un periodo compreso tra fine maggio e primi giorni di settembre) e sono stati ottenuti da Google Analytics.

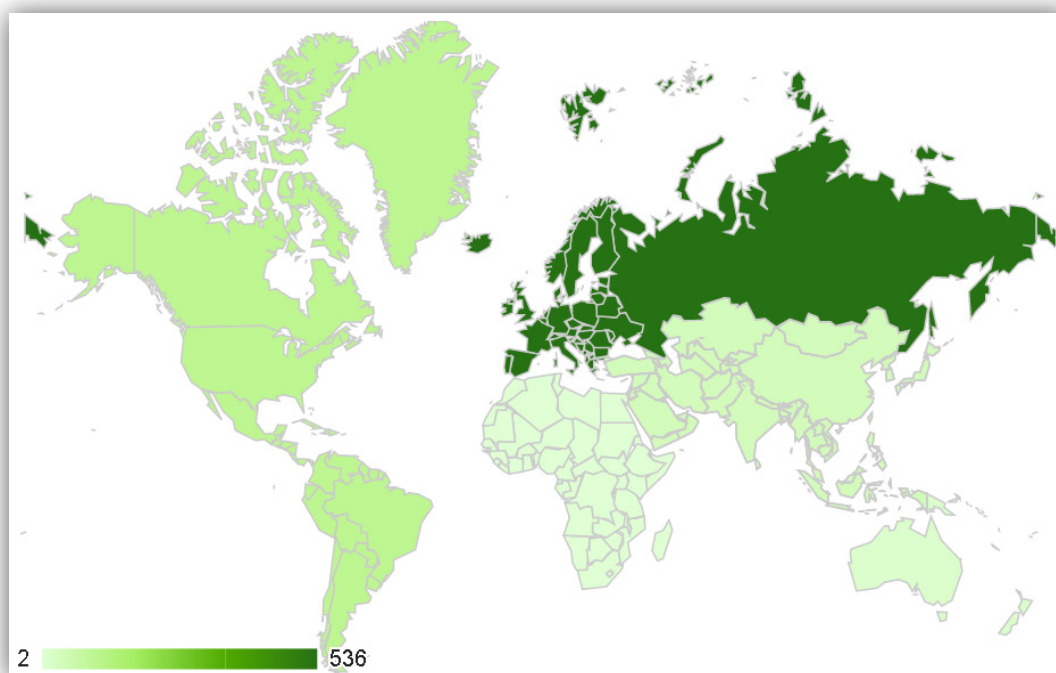


FIGURE 16: UTILIZZO DEL DB MIRANDOLA

Continente	Visite	Pagine/visita	Durata media visita	%nuove visite	Frequenza di rimbalzo
Europa	536	6,47	00:08:46	27,80%	32,84%
America	106	2,99	00:03:11	63,21%	33,96%
Asia	42	3,02	00:05:53	71,43%	35,71%
Oceania	14	3,93	00:05:12	57,14%	14,29%

Continente	Visite	Pagine/visita	Durata media visita	%nuove visite	Frequenza di rimbalzo
Africa	2	3,00	00:08:53	100,00%	0,00%

In totale il database è stato visitato da 259 visitatori unici provenienti da 39 stati e 125 città di tutto il mondo.

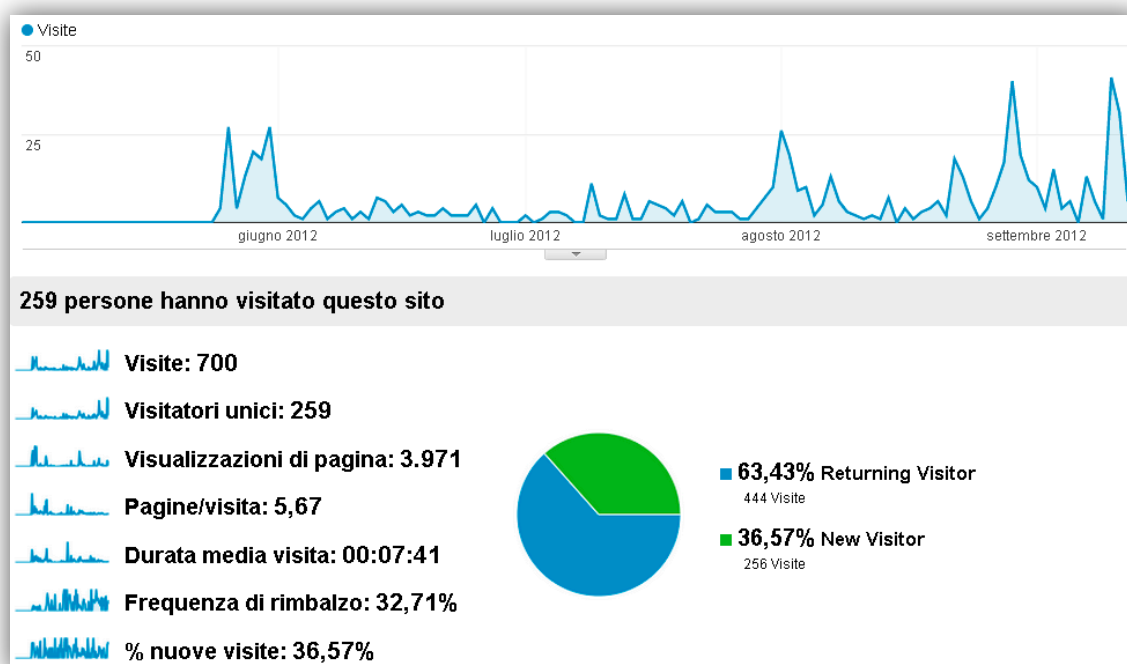


FIGURE 17: STATISTICHE UTILIZZO DB MIRANDOLA

1.7.6. Prospettive Future

miRandola è la prima risorsa online che fornisce tutti i dati disponibili su miRNA circolanti ed extracellulari in un unico ambiente. Il database rappresenta un utile punto di riferimento per chiunque sia interessato a studiare il ruolo dei miRNA extracellulari come biomarcatori e il loro ruolo fisiologico e il coinvolgimento in patologie.

L'attuale versione di miRandola è focalizzata su miRNA circolanti umani. In futuro sarà possibile introdurre informazioni anche su altre specie e differenti tipologie di miRNA circolanti, quando saranno disponibili nuovi dati.

miRandola è costantemente aggiornato dallo staff ogni volta che sono disponibili nuove informazioni da inserire nel database e l'"online submission system" è una importante caratteristica che garantisce un continuo aggiornamento del sistema.

Vi sono, inoltre, due progetti in fase di sviluppo. Uno di questi rappresenta un utile strumento che verrà inserito nel database nella sezione "Tools", che consiste nell'analisi dell'arricchimento funzionale dei miRNA circolanti ("miRto: a tool for circulating microRNAs enrichment analysis").

Questo progetto nasce dall'esigenza di valutare in modo statistico e significativo l'arricchimento funzionale dei miRNA e in particolare dei miRNA circolanti attraverso l'analisi delle funzioni (termini ontologici) condivise dai target dei miRNA.

Per raggiungere tale scopo è stato condotto un test statistico molto utilizzato nel campo dell'analisi dell'arricchimento funzionale da diversi tool come ad esempio DAVID (Nature Protocols 2009;4(1):44 & Genome Biology 2003; 4(5):P3) , ovvero il "Fisher's Exact Test". Il secondo progetto consiste nell'introduzione in miRandola di dati di espressione non ancora presenti nel database. In particolare, quando saranno disponibili, i dati di espressione derivati da "Next Generation Sequencing", verranno inseriti e rappresentati in modo grafico. I livelli di espressione verranno visualizzati per ogni tipo di campione.

1.8. I microRNA e le malattie

È ampiamente riportato che i miRNA sono associati con molte malattie umane come quelle autoimmuni, della pelle, neurologiche e psichiatriche. In aggiunta, i miRNA si stanno muovendo rapidamente al centro come chiavi che regolano lo sviluppo e la funzione neuronale e in più contribuiscono in maniera importante nei disordini neurodegenerativi come Alzheimer, Parkinson, Huntington, Tourette, Down, depressione e schizofrenia.

Molti studi hanno prodotto un ampio numero di associazioni miRNA-malattie e mostrato che i meccanismi dei miRNA coinvolti nelle malattie sono molto complessi.

1.8.1. microRNA nel Cancro

Negli ultimi anni, si è evidenziata una forte correlazione tra le disfunzioni dei miRNA e le malattie umane, esiste infatti una grande quantità di dati che dimostra il ruolo centrale dei miRNA nella patogenesi del cancro. Questo ha portato a rivisitare la lista di geni

classicamente considerati come oncosoppressori e oncogeni inserendo appunto la famiglia dei miRNA, queste molecole infatti sono in grado di regolare negativamente la stabilità e il processo di traduzione dei mRNA e risultano implicate in diversi processi cellulari quali controllo del ciclo cellulare, differenziamento ed apoptosi. La prima evidenza del coinvolgimento dei miRNA nel cancro è stata data dalla sottoespressione o delezione di miR-15 e miR-16 nella maggior parte dei pazienti affetti da leucemia linfocitica cronica (CLL) [53]. L'espressione dei miRNA è generalmente tessuto-specifica e fortemente dipendente dal livello di differenziamento della cellula, per cui la presenza di significative alterazioni nei livelli di espressione dei miRNA in un tessuto tumorale potrebbe semplicemente rappresentare delle conseguenze secondarie a processi più direttamente implicati nella tumorigenesi, piuttosto che essere un elemento causale direttamente associato alla trasformazione neoplastica della cellula stessa. Un significativo numero di geni dei miRNA sono localizzati in regioni genomiche frequentemente riarrangiate nei tumori, come siti fragili, regioni *delete* (minimal region of loss of heterozygosity, LOH) o amplificate (minimal amplicons) e regioni comuni di *break-point*, fornendo un'ulteriore evidenza del loro ruolo nella patogenesi del cancro [54] [55].



FIGURE 18: LOCALIZZAZIONE DEI MICRORNA IN CAGR (CANCER ASSOCIATED GENOME REGIONS) [56]

A seconda del tipo di funzione biologica del miRNA, determinata dalle proteina che regola, si possono distinguere miRNA oncogeni o miRNA onco-soppressori [10] [56]. Esempi noti di oncogeni sono il miR-155 amplificato in molti tumori ematologici, tumore al polmone e al seno e il cluster miR-17-92 che comprende 6 geni per miRNA sovraespressi in tumori solidi e nel linfoma diffuso a cellule larghe B.

Tra i miRNA oncosoppressori più noti ci sono i miR-15a e miR-16-1 che hanno come bersaglio la proteina-antiapoptotica BCL2, e la famiglia let-7, che reprime l'espressione dell'oncogene RAS (Figura 19)

MicroRNA	Dysregulation	Function	Validated targets	Oncogene (ONC) or tumour suppressor (TS)	Refs
miR-15a and miR-16-1	Loss in CLL, prostate cancer and multiple myeloma	Induces apoptosis and inhibits tumorigenesis	BCL2, WT1 RAB9B and MAGE83	TS	15,20,23, 30,52,69
let-7 (a, b, c, d, e, f, g and i)	Loss in lung and breast cancer and in various solid and haematopoietic malignancies	Induces apoptosis and inhibits tumorigenesis	RAS, MYC and HMGA2	TS	22,26, 42,70
miR-29 (a, b and c)	Loss in aggressive CLL, AML (11q23), MDS lung and breast cancers and cholangiocarcinoma	Induces apoptosis and inhibits tumorigenicity. Reactivates silenced tumour suppressor genes	TCL1, MCL1 and DNMTs	TS	30,64, 71,72
miR-34	Loss in pancreatic, colon, breast and liver cancers	Induces apoptosis	CDK4, CDK6, cyclin E2, EZF3 and MET	TS	56–58
miR-145	Loss in breast cancer	Inhibits proliferation and induces apoptosis of breast cancer cells	ERG	TS	31
miR-221 and miR-222	Loss in erythroblastic leukaemia	Inhibits proliferation in erythroblasts	KIT	TS	30
miR-221 and miR-222	Overexpression in aggressive CLL, thyroid carcinoma and hepatocellular carcinoma	Promotes cell proliferation and inhibits apoptosis in various solid malignancies	p27, p57, PTEN and TIMP3	ONC	43,51,73
miR-155	Upregulated in aggressive CLL, Burkitt's lymphoma and lung, breast and colon cancers	Induces cell proliferation and leukaemia or lymphoma in mice	MAF and SHIP1	ONC	32–34, 36,37
miR-17–92 cluster	Upregulated in lymphomas and in breast, lung, colon, stomach and pancreatic cancers	Induces proliferation	E2F1, BIM and PTEN	ONC	19,34,35, 40,41
miR-21	Upregulated in glioblastomas, AML (11q23), aggressive CLL and breast, colon, pancreatic, lung, prostate, liver and stomach cancers	Inhibits apoptosis and increases tumorigenicity	PTEN, PDCD4, TPM1 and TIMP3	ONC	31,37–39, 44–50
miR-372 and miR-373	Upregulated in testicular tumours	Promotes tumorigenicity in cooperation with RAS	LATS2	ONC	74

FIGURE 19: MICRORNA CHE AGISCONO COME ONCOGENI E ONCOSOPPRESSORI E RELATIVI TUMORI IN CUI SONO COINVOLTI [56]

Un'analisi completa dell'espressione dei miRNA nelle diverse fasi della carcinogenesi gastrica, ha rilevato una sovra-regolazione del cluster miR-106b-25 con conseguente effetto sulla pathway dell'oncosoppressore TGF- β , interferendo con l'espressione di p21Eaf1/Cip1 e Bim. Questi risultati suggeriscono il ruolo chiave di questo cluster nel cancro dello stomaco, causato dall'interferenza con proteine coinvolte nel ciclo cellulare dell'apoptosi [57].

1.8.2. *microRNA nell'Alzheimer*

L'*Alzheimer's disease* (AD) è la forma più comune di demenza in tutto il mondo, con un numero stimato di 25,6 milioni di persone colpite nel 2010 e si prevede saranno 65,7 milioni nel 2030. È una patologia degenerativa del cervello caratterizzata da un processo di atrofizzazione progressiva che provoca il graduale ed inesorabile declino delle funzioni intellettive, associato a un deterioramento della personalità e delle capacità di relazione.

Lo studio dei casi familiari ha portato all'identificazione di alcuni geni le cui mutazioni causano AD e di altri le cui varianti possono contribuire a determinare la suscettibilità ad ammalarsi anche per persone che non appartengono a gruppi familiari a rischio. I geni noti, le cui mutazioni causano l'Alzheimer familiare (FAD) sono: il gene della proteina amiloide (APP), il gene della presenilina 1 (PS1) e il gene della prenilina 2 (PS2). Il gene della suscettibilità maggiormente noto è invece l'Apolipoproteina E (APOE), questo si lega alla proteina amiloide. Le mutazioni dei geni APP, PS1 e PS2 causano un incremento nella produzione di β -amiloide, il costituente delle placche senili, mentre l'allele e4 dell'apolipoproteina non causa un aumento della produzione di amiloide ma piuttosto un incremento del suo deposito nello spazio extracellulare. Ricerche per un marcatore genetico per AD sporadica si sono focalizzate su polimorfismi di geni che codificano proteina riguardanti l'amiloide o le proteine tau nelle pathway metaboliche, come: (i) LRP1 (*Low density lipoprotein-related protein 1*); (ii) MAPT (*micro-tubule-associated protein tau*); (iii) BDNF (*brain-derived neurotrophic factor*); (iv) IDE (*insulin-degrading enzyme*); (v) A2M (*alpha 2-macro-globulin*), (vi) ACE (*Angiotensin I Converting Enzyme*).

Indizi crescenti suggeriscono il ruolo dei miRNA nell'AD, rispetto alla regolazione dell'enzima BACE1 (*β -amyloid precursor protein-converting enzyme 1*), in particolare miR-29a, miR-29b-1 e miR-9 possono regolare l'espressione di BACE1 *in vitro* e il cluster miR-29a/b-1 risulta decrementato significativamente nei pazienti AD che mostrano un anormale alto livello della proteina BACE1. L'espressione dei livelli di miRNA ha messo in evidenza miR-107 poiché decrementava significativamente anche in un paziente nelle prime fasi della patologia [58]. È interessante che particolari lamine delle corteccia cerebrale coinvolti nella patologia AD esibiscono un'espressione neuronale del miR-107 diminuita e che il miR-107 può essere coinvolto nella progressione della malattia attraverso la regolazione del BACE1 [58].

Mutazioni nei geni APP e PSEN causano accumulo $A\beta$ e l'AD familiare [59] [60], recentemente Hébert et al investigarono cambi nei profili di espressione di pazienti AD sporadici e trovarono che alcuni miRNA potenzialmente coinvolti nella regolazione dell'espressione di APP e BACE1 appaiono diminuiti nella malattia cerebrale [59]. Inoltre proposero che la perdita di specifici miRNA può contribuire ad incrementare BACE1 e

livelli A β in AD. In differenti studi sui profili di espressione dei miRNA del cervello affetto da AD, l'*up-regulation* del miR-125b e la *down-regulation* del miR-9 e miR-210 sono stati riportati in maniera consistente (36). L'*up-regulation* del miR-197 e la *down-regulation* dei miR-15, miR-146b, miR-181c e miR-338 sono comunemente alterati nel parenchima cerebrale e fluido cerebrospinale in AD [61] .

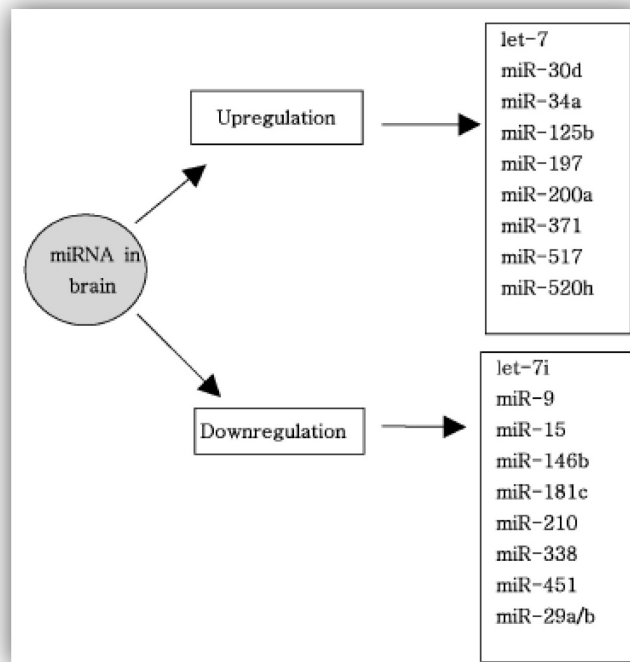


FIGURE 20: PATTERN ALTERATI DI ESPRESSIONE DEL MIRNA NEL PARENCHYMA CEREBRALE E FLUIDO CEREBROSPIANLE DI PAZIENTI CON L'ALZHEIMER.

1.8.3. *microRNA nel Diabete*

Il diabete è una delle malattie croniche più comuni al mondo, ha raggiunto proporzioni epidemiche e ci si aspetta un numero di 366 milioni nell'anno 2030 [62], causa di questa esplosione è lo sviluppo di alcuni paesi, principalmente India e Cina. Rapidi cambi nell'urbanizzazione, industrializzazione e globalizzazione hanno, da un lato, aperto nuove strade verso la prosperità socio-economica ma, dall'altro lato, hanno incrementato la tendenza verso l'inattività fisica e l'obesità donandoci una pletera di disordini metabolici.

Il diabete è causato da difetti di produzione, secrezione e segnalazione dell'insulina [63] [64], ci sono due tipi di diabete: il diabete di tipo 1 (T1D) è dovuto all'autodistruzione delle cellule beta produttrici di insulina nel pancreas, il diabete di tipo 2 (T2D) è causata da difetti di azione dell'insulina e la produzione di insulina. Molte delle complicazioni del diabete sono in origine vascolare, siano essi macrovascolari e/o micro vascolari (nefropatia, retinopatia, e microangiopatia in diversi organi), e questo pone i diabetici a

rischio di cardiopatia ischemica, insufficienza renale, ictus, amputazioni degli arti inferiori e cecità [65].

Recenti ricerche suggeriscono che i miRNA giocano un ruolo critico nella patogenesi del diabete poiché coinvolti nella secrezione dell'insulina [66], sviluppo delle isole pancreatiche [67], differenziazione delle cellule beta [68], ed indirettamente nel controllo del glucosio e nel metabolismo dei lipidi [69] e secondariamente sono coinvolti nelle complicazioni associate al diabete [70]. Il player principale che emerge come mediatore del rilascio dell'insulina e perciò dell'omeostasi dell'insulina è il microRNA specifico delle isole pancreatiche, miR-375. Il miR-375 regola negativamente la secrezione dell'insulina stimolata dal glucosio (GSIS), infatti l'inibizione del miR-375 migliora la secrezione dell'insulina, mentre la sua sovra-espressione danneggia la *pathway* secretoria dell'insulina riducendo l'espressione del myotrophin (Mtpn). Un altro microRNA che rappresenta un forte candidato è miR-9 che ha come target il fattore di trascrizione Onecut 2, che inibisce l'espressione del Granuphilin, un regolatore negativo dell'esocitosi dell'insulina, la sovra-espressione del miR-9 quindi decresce GSIS. miR-124a, oltre ad essere vitale per lo sviluppo delle cellule beta pancreatiche, modula diversi componenti del sistema di esocitosi targettando direttamente il box proteico A2 (Foxa2)- un fattore di trascrizione coinvolto nel metabolismo del glucosio e nella secrezione dell'insulina.

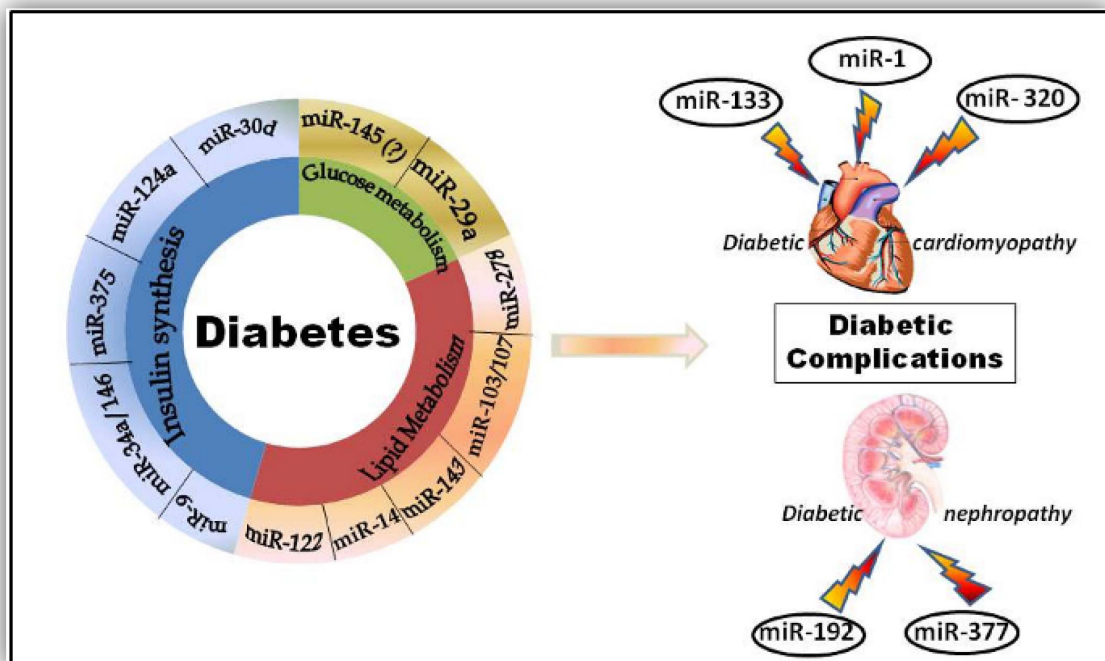


FIGURE 21: MOLTI MIRNA SONO STATI IDENTIFICATI ESSERE CRITICI IN DIVERSI PROCESSI METABOLICI, DIS-REGOLAZIONI CHE PORTANO AL DIABETE E LE SUE COMPLICAZIONI. [71]

Come sintetizza la Figura 21 i miRNA coinvolti nella sintesi dell'insulina sono miR-9, miR-34a/146, miR-375, miR-124a, e miR-30d, i miRNA coinvolti nel metabolismo del glucosio sono miR-29a e miR-145, i miRNA coinvolti nel metabolismo dei lipidi sono miR-14, miR-122, miR-143, miR-103/107 e miR-278. Per le complicazioni diabetiche, i miRNA coinvolti nella cardiomiopatia diabetica sono miR-133, miR-1, e miR-320, e i miRNA quelli coinvolti nella nefropatia diabetica sono miR-192 e miR-377.

1.8.4. *microRNA nella Sclerosi Multipla*

La sclerosi multipla (MS) è una malattia neurodegenerativa del sistema nervoso centrale (CNS), [72] [73] [74], rappresenta la principale causa di disabilità neurologica dopo una lesione cerebrale traumatica in giovani adulti con discendenza europea [75]. Tipicamente risulta essere molto più evidente nelle donne che negli uomini, è una malattia complessa in cui sia fattori genetici che ambientali giocano un ruolo importante [76]. I fattori ambientali coinvolti nella progressione della malattia possono essere patogeni come la presenza di una infezione del virus Epstein Barr che risulta in una risposta anormale o variante come manifesto di una infezione di mononucleosi (IM) quando un individuo è carente di vitamina D e MS suscettibile [77]. La malattia si presenta in un numero di differenti patologie da recidivante-remittente (RRMS) a primario progressiva ma alla fine tutti i fenotipi sembrano dar luogo ad una forma progressiva. L'identificazione di biomarcatori per MS migliorerebbe notevolmente la diagnosi ed il monitoraggio dell'attività della malattia e della risposta al trattamento. Si potrebbe anche dar luogo ad uno screen pre-sintomatico in individui geneticamente suscettibili. Inoltre, i segni e sintomi della MS usualmente condividono similarità considerabili con quelle di altre malattie infiammatorie del CNS, come la neuromielite ottica e questo porta considerevoli difficoltà di terapie e di diagnosi [78].

Uno degli aspetti più importanti della patogenesi della MS, e anche uno dei più riconosciuti soprattutto nelle fasi iniziali della malattia, è che l'autoimmunità è mediata da T-cellule CD⁺4 (87). Keller et al esaminarono i profili di espressione di 866 miRNA umani [79]. Essi analizzarono l'espressione del miRNA nelle cellule del sangue di 20 pazienti RRMS e 19 controlli sani usando *microarray* miRNA umani e la piattaforma *Genom Real Time Analyzer*. Identificarono 165 miRNA che erano significativamente sovra o sotto espressi in pazienti con RRMS paragonati con i controlli sani. Inoltre hsa-miR-145 si dimostrò essere il miglior marcatore miRNA permettendo di distinguere i pazienti MS dai controlli sani con una specificità del 89,5%, una sensibilità del 90%, ed una accuratezza del 89.7%. Inoltre 43 dei 165 miRNA deregolati nei pazienti con MS erano stati precedentemente correlati con altre malattie umane, i restanti 122 miRNA invece erano esclusivamente associati con la

MS. Ciò suggerisce che l'espressione dei miRNA può rappresentare un potenziale utile biomarcatore per la diagnosi della MS e che la deregolazione dell'espressione dei miRNA potrebbe giocare un ruolo nella complessa patogenesi della malattia.

1.9. *microRNA, un futuro approccio terapeutico*

L'identificazione di biomarcatori specifici è sempre stata la sfida per rivoluzionare prognosi e diagnosi delle malattie. Ad esempio c'è sempre stata una forte ricerca per biomarcatori clinici che potrebbero essere utilizzati per rivelare lo sviluppo e la progressione del diabete in una fase precoce per ritardare se non inibire l'insorgenza della fase tardiva della malattia. Studi recenti hanno sempre più dimostrato la presenza di RNA nel siero [80] [81], ma la sua stabilità è sempre stata una preoccupazione per i ricercatori di biomarcatori, la scoperta di miRNA è stata una svolta in questo contesto considerando la loro notevole stabilità e la presenza elevata sia nel siero che nel plasma [82].

Oltre a sfruttare miRNA circolanti come potenziali biomarker, gli alterati profili di espressione dei miRNA potrebbero essere studiate come chiave per il trattamento terapeutico in molte malattie. Ad esempio la sovra-espressione o l'inibizione di miRNA specifici potrebbe essere il tool futuro per il targeting del diabete [83].

I miRNA possono essere utilizzati anche come target terapeutici e come un potente strumento di intervento, poiché essi costituiscono una componente chiave della regolazione genica. La sovra espressione dei miRNA oncosoppressori in cellule che non li esprimono può essere indotta introducendo dall'esterno miRNA sintetici [84]. Invece, i miRNA oncogenici silenziati utilizzando sia vettori che oligonucleotidi antisenso: gli *antagomirs* (analoghi sintetici dei miRNA) [85].

I miRNA sono dei target attraenti per le *drug* poiché essi regolano l'espressione di molte proteine nella cellula e sono differentemente espresse in cellule sane e maligne. Ogni miRNA può avere anche centinaia di target mRNA nella cellula, questa regolazione *multi-target* potrebbe essere responsabile per effetti avversi o non-target in una futura terapia con i microRNA.

CAPITOLO 2:

GENOTIPO/FENOTIPO

2. Genotipo/Fenotipo

2.1. Genotipo Definizioni

Entrambi i concetti di "genotipo" e "fenotipo" sono termini vaghi per usi pratici. Per essere capaci di poter lavorare con essi, in un database o in analisi computazionali, è necessario ridurli quindi a concetti più usabili tali da poter essere entità paragonabili e tracciabili. Per meglio spiegare come il concetto di genotipo risulti vago, consideriamo la figura seguente,

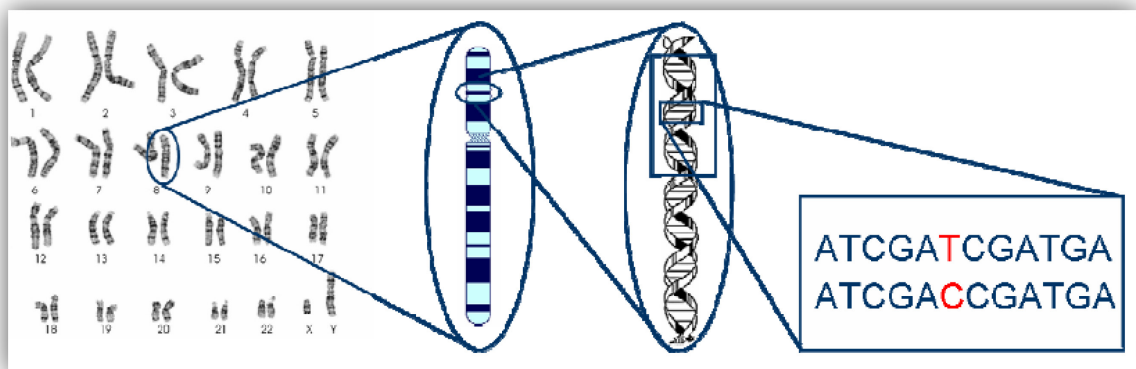


FIGURE 22: DIFFERENTI CONCETTI PER UN GENOTIPO A DIFFERENTI LIVELLI DI GRANULARITÀ DELLA COMPOSIZIONE GENETICA DI UN INDIVIDUO

da questa risulta chiaro che un genotipo è inteso come qualsiasi cosa che vada dalla somma di tutte le proprietà genetiche di un organismo ad una singola variazione in un solo nucleotide nel codice genetico dello stesso organismo. Questa eterogeneità dei concetti per un genotipo deve essere superata per un suo uso pratico. Ogni genotipo di ciascun individuo differisce da quello degli altri, anche all'interno di una stessa specie. È quindi necessario poter lavorare con strutture più pratiche rispetto al genotipo.

Un'alternativa all'uso del termine genotipo è l'uso del concetto di geni. Tuttavia, anche questo ha subito notevoli variazioni nel tempo, Thomas Morgan lo definì "*abstract entity whose existence is reflected in the way phenotypes were transmitter between generations*" [86], dopo che Watson e Crick scoprirono la struttura tridimensionale del DNA nel 1953 [87] il gene fu inserito in una visione molecolare come "*a code residing on*

nucleic acid that gives rise to a functional product" [86]. Tra gli anni del 1970 e 1980 un gene era visto come parte di una sequenza piuttosto che un *locus* genetico responsabile per un fenotipo [86]. Wain et al definirono il gene come "a DNA segment that contributes to phenotype/function. In the absence of demonstrated function, a gene may be characterized by sequence, transcription or homology" [88].

2.1.1. Geni e Proteine

Le proteine sono sintetizzate dai geni tramite la trascrizione della sequenza di DNA e la traslazione nella loro sequenza di aminoacidi. Molti geni codificano più di una proteina a causa di effetti che avvengono durante la biosintesi proteica, come le modificazioni post-trascrizionali [89].

È sufficiente sapere che le proteine sono essenziali per ogni organismo e partecipano in molti processi nelle cellule. Inoltre, le proteine hanno funzioni strutturali e meccaniche, giocano un ruolo importante anche in altre funzioni cellulari come il ciclo cellulare e la risposta immunitaria [89].

2.1.2. Annotazione Funzionale: Gene Ontology (GO)

I prodotti genici di molti eucarioti condividono molte funzioni biologiche fondamentali [90], *Gene Ontology* (GO) fu creato nel 1999 con l'obiettivo di produrre un vocabolario controllato e dinamico che possa essere applicato a tutti gli eucarioti poiché la conoscenza dei ruoli di geni e proteine nelle cellule si sta accumulando e cambiando [91]. GO è strutturato questo significa che ogni termine è organizzato in una delle tre sub-ontologie, *molecular function*, *biological process* e *cellular component*, rappresentando un grafo diretto aciclico con successori e predecessori per ogni termine. GO è ampiamente riconosciuto come il sistema di classificazione più comprensivo dal punto di vista funzionale ed è diventato uno standard internazionale per le annotazioni funzionali [92] [93]. La struttura e lo stile usati per rappresentare annotazioni funzionali e le associazioni ad altri geni all'interno di GO sono illustrati nella seguente figura:

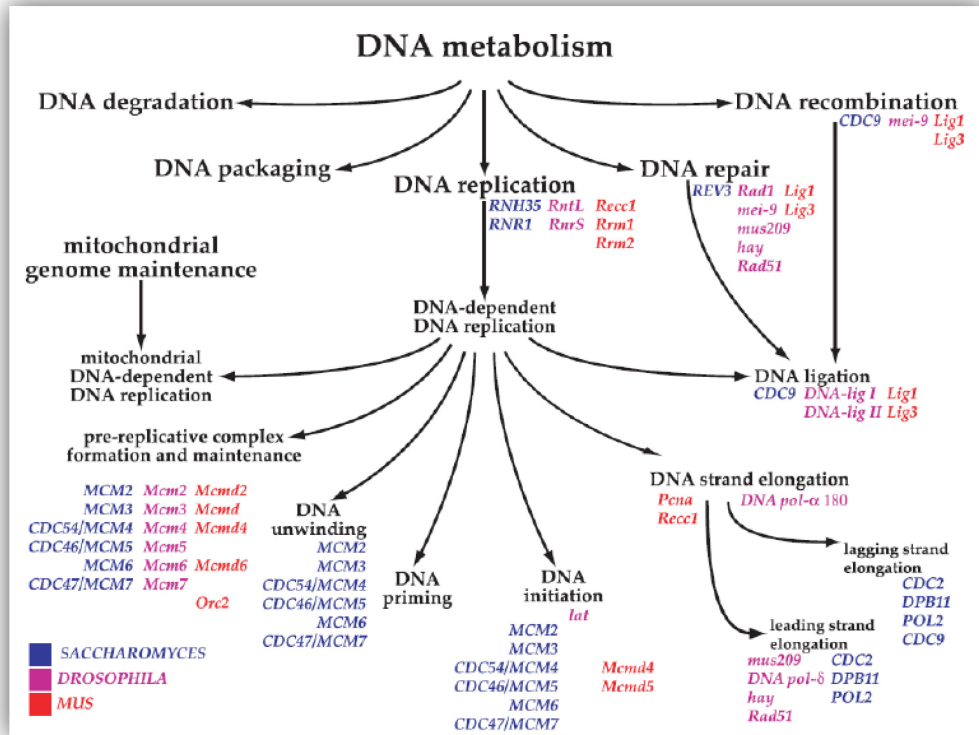


FIGURE 23: ESEMPIO GO [91]

2.2. Fenotipo Definizioni

La complessità delle interazioni e la variazione genetica inter-individuale si riflettono sull'individualità fenotipica all'interno della specie *Homo sapiens*. Quindi anche il concetto di fenotipi risulta essere difficile da comprendere come per il genotipo. Un fenotipo comprende qualsiasi caratteristica osservabile di un organismo, la descrizione di una malattia, la caratterizzazione di mutazioni naturali, ect [94]. Sfortunatamente non esiste un vocabolario per descrivere queste osservazioni, infatti i ricercatori usano vocabolari tradizionali arricchiti oppure vocabolari specifici od ancora linguaggio parlato (in inglese). A causa dell'eterogeneità che risulta nelle descrizioni, e per poter usare tutti i dati fenotipici disponibili, in questa tesi si usa il denominatore comune più ampio per un fenotipo, ovvero la sua descrizione testuale. Ciò che è considerato "dato fenotipico" varia dal livello molecolare a cartelle di pazienti clinici, da organismi a popolazioni.

La coesistenza di dati fenotipici quantitativi e qualitativi porta ad una eterogeneità che spiega perché sono stati fatti pochi tentativi per integrare i dati fenotipici, lasciando un ampio spettro di risorse. Data la pleora di database specifici per specie e locus, i database tra le specie e genome-wide sono lo step logico verso la comparazione di dati fenotipici. Progetti su ampia scala come Mouse Phenome Project [95] [96] e Eumorphia

[97] sono iniziati per coordinare gli sforzi per la generazione di dati fenotipici standardizzati pubblicati in database specifici. Recentemente anche l'Unione Europea ha finanziato con 12 milioni di euro il progetto GEN2PHEN, un progetto genotipo/fenotipo che coinvolge 19 istituti in tutta Europa [98]. Un progetto analogo nel 2003 fu *Human Phenome Project* [99].

2.2.1. Mammalian Phenotype Ontology (MP)

Mammalian Phenotype Ontology (MP) [100], conteneva a Settembre del 2008 6.182 termini fenotipici [101], che insieme alle ontologie, 8000 nel 2009, di *Human Phenotype* (HP) [102], sono ad oggi le maggiori ontologie fenotipiche di mammiferi. Anche se HP fu sviluppato per umani (si focalizza principalmente sulle anomalie anatomiche) ed MP era principalmente creato per ratti e topi, le loro potenzialità vanno oltre queste specie. Infatti MP contiene fenotipi cellulari che possono parzialmente essere applicati ad altri organismi modello come *Drosophila melanogaster* o *Caenorhabditis elegans*, aprendo nuove prospettive per questa risorsa. Per guidare gli sforzi verso la creazione di ontologie fenotipiche generale viene utilizzato un software per annotazioni fenotipiche come Phenote (disponibile on line, <http://www.phenote.org/>).

Ma perché le ontologie si prestano come strumento importante per analisi comparative? La descrizione in testo libero usata dai ricercatori per documentare le loro osservazioni fenotipiche potrebbero differire considerevolmente e permettere dei paragoni solo manualmente o tramite approcci di *text-mining*, le ontologie permettono una identificazione non ambigua di oggetti biologici quando sinonimi sono usati come termini di interrogazioni, quindi essi sono altamente specifici per paragoni tra specie.

help.

by features. by disease. by ontology. by genes.

Enter feature search. reset.

HPO-ID	Feature
HP:0010704	1-2 finger syndactyly
HP:0005767	1-2 toe complete cutaneous syndactyly
HP:0010711	1-2 toe syndactyly
HP:0010706	1-3 finger syndactyly
HP:0001459	1-3 toe syndactyly
HP:0010707	1-4 finger syndactyly
HP:0010712	1-4 toe syndactyly
HP:0006088	1-5 finger complete cutaneous syndactyly
HP:0010708	1-5 finger syndactyly
HP:0010713	1-5 toe syndactyly
HP:0000878	11 pairs of ribs
HP:0001233	2-3 finger syndactyly

Page 1 of 829 Displaying features 1 - 12 of 9942

FIGURE 24: PAGINA DI HP

```

mammalian phenotype
├─ adipose tissue phenotype +
├─ behavior/neurological phenotype +
├─ cardiovascular system phenotype +
├─ cellular phenotype +
├─ craniofacial phenotype +
├─ digestive/alimentary phenotype +
├─ embryogenesis phenotype +
├─ endocrine/exocrine gland phenotype +
├─ growth/size phenotype +
├─ hearing/vestibular/ear phenotype +
├─ hematopoietic system phenotype +
├─ homeostasis/metabolism phenotype +
├─ immune system phenotype +
├─ integument phenotype +
├─ limbs/digits/tail phenotype +
├─ liver/biliary system phenotype +
├─ mortality/aging +
├─ muscle phenotype +
├─ nervous system phenotype +
├─ no phenotypic analysis
├─ normal phenotype +
├─ other phenotype +
├─ pigmentation phenotype [MP:0001186] (1901 genotypes, 3083 annotations)
│   ├── abnormal coat/hair pigmentation +
│   ├── abnormal digit pigmentation
│   ├── abnormal extracutaneous pigmentation +
│   ├── abnormal melanocyte morphology +
│   ├── abnormal melanocyte proliferation
│   ├── abnormal melanogenesis +
│   ├── abnormal melanosome transport
│   ├── abnormal pigmentation pattern
│   ├── abnormal scrotum pigmentation
│   ├── abnormal skin pigmentation +
│   ├── hyperpigmentation +
│   ├── hypopigmentation +
│   └─ lipofuscinosis
├─ renal/urinary system phenotype +
├─ reproductive system phenotype +
├─ respiratory system phenotype +
├─ skeleton phenotype +
├─ taste/olfaction phenotype +
├─ tumorigenesis +
└─ vision/eye phenotype +

```

FIGURE 25: PAGINA DI MP

2.3. Relazione Genotipo + Fenotipo

2.3.1. Fenotipi Mendeliani

Le malattie "Mendeliane" o "monogeniche" sono dovute alle mutazioni di singoli geni, si distinguono in autosomica: i) dominante, ii) recessiva, legata al cromosoma X: i) dominante e ii) recessiva. Tramite tecniche di clonaggio sono state identificate circa 1.200 malattie causate da geni negli umani [103]. Sfortunatamente, le loro relazioni genotipo-fenotipo non sono sempre ovvie. Differenti polimorfismi di singoli nucleotidi (SNPs) dello stesso *locus* genetico causano fenotipi differenti [104], o sub-fenotipi di una malattia. Inoltre la combinazione di SNP (detti aplotipi) può "regolare" il fenotipi finale di una malattia causata da SNP o spiegare la suscettibilità a malattie diverse [103]. Ad esempio, la fenilchetonuria (PKU) è considerata una classica malattia monogenica che può essere causata da alcune mutazioni nell'enzima *phenylalanine hydroxylase* (PHA). Tuttavia anche fratelli che condividono un genotipo PHA identico mostrano fenotipi ampiamente differenti [105] [106]. Per PHA, esiste un database di mutazione PHAdb che contiene 498 mutazioni per un locus specifico [107] [108], molte delle quali si presume siano causa della malattia [106] [109]. Esistono indizi che una notevole variabilità fenotipica all'interno della stessa malattia monogenica possa essere influenzata da altri fattori genetici [110] [111]. Altro esempio della complessità di relazione genotipo-fenotipo per una malattia monogenica è la talassemia. Se il set di SNP sono ereditati insieme come aplotipi, il contributo individuale fenotipico di ciascun SNP all'interno di questo aplotipo è oscurato.

2.3.2. Tratti Complessi

I fenotipi di malattie complesse sono, paragonate alle malattie monogeniche, anche più difficili da collegare inequivocabilmente ai siti rilevanti di variazione genomica. Per malattie complesse come i diabete, Alzheimer, disordini psichiatrici o obesità il quadro completo di relazioni genotipo-fenotipo rimane ampiamente irrisolto, i contributi di un gene alla malattia sono individuati tramite studi di ampie popolazioni e sono definite come "associazione" o "suscettibilità" a causa della mancanza della comprensione del contributo. La variabilità di due fenotipi a dispetto di sequenze identiche dei geni causanti il fenotipo sono dichiarate essere una conseguenza del "differente *background* genetico" intendendo con questo effetti inaspettati dal resto del genoma a cui si unisce l'influenza di fattori esterni.

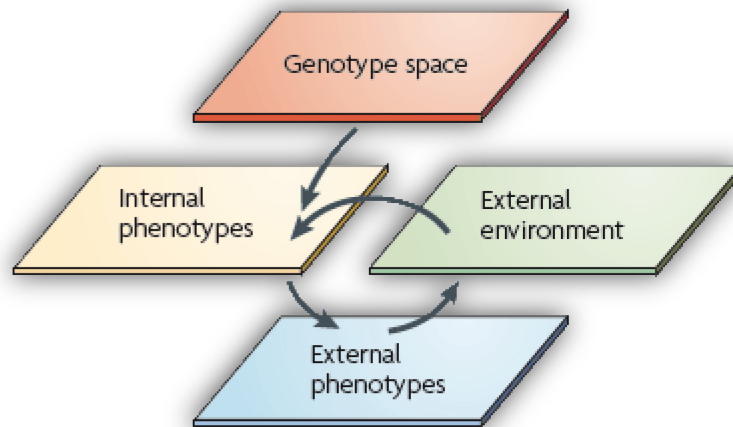


FIGURE 26: RELAZIONE FENOTIPO/GENOTIPO

2.3.3. Associazione Genotipo – Fenotipo

Nonostante le evidenti difficoltà di mappare un fenotipo sul sottostante genotipo e le sfide di descrivere fenotipi in maniera consistente e in maniera altamente standardizzata, sono stati compiuti moltissimi sforzi per raccogliere e generare dati fenotipici [112] [113] [114] [115]. Il moscerino della frutta è un buon esempio, nei decenni passati sono stati impiegate molte metodologie per collezionare e descrivere mutazioni con molti dettagli. Tali dati hanno aiutato a sviluppare le relazioni genotipo-fenotipo iniziando con lo studio di un fenotipo mutato e continuando identificando il gene mutato. Studi analoghi sono stati condotti in mammiferi superiori ma sono ostacolati da una durata di vita maggiore, mancanza di metodi sofisticati o motivi etici.

La situazione sta cambiando con l'avvento di metodi di screening di RNAi e di altri potenti strumenti di screening, come i programmi di mutagenesi che usano agenti chimici.

2.4. Analisi dei Fenotipi: Dati e Tool

Sebbene sia disponibile l'intero accesso alla maggior parte delle pubblicazioni, l'identificazione e l'estrazione automatica, tramite computer, di dati fenotipici dal corpus del testo rimane una sfida così come lo sviluppo di tool ed algoritmi per questo scopo [116] [117].

Mentre il *text-mining* per i nomi dei geni è possibile grazie alle convenzioni dei nomi (HGNC) [118], ma tuttavia rimane ancora un compito gravoso a causa di irregolarità, omonimi ect, [119] [120] [121], il *text-mining* nel campo dei fenotipi soffre della natura descrittiva

dei dati come ad esempio la mancanza di una nomenclatura standardizzata. Anche con l'esistenza delle ontologie la situazione rimane non soddisfacente, infatti è notoriamente difficile assegnare termini ontologici alla corrispondente letteratura biomedica come ha mostrato un recente tool di *text-mining* (BioCreAtivE) [122] [123]. I team per database specifici ad organismi hanno fatto un lavoro eccezionale collezionando dati fenotipici dalla letteratura, hanno fornito un accesso molto più veloce e semplice ai dati fenotipici per ogni gene per un dato organismo, vedi Figura.

Organism	Name	URL
<i>cross-species</i>	PhenomicDB -Multi-species Phenotype-Genotype Database incl. RNAi screens	http://www.phenomicdb.de
<i>Homo sapiens</i>	OMIM -Online Mendelian Inheritance in Man HGMD -Human Gene Mutation Database PharmGKB -Pharmakogenetics and Pharmakogenomics Knowledge Base GenAtlas -Gene and Phenotype database	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM http://www.hgmd.org http://www.pharmgkb.org http://www.genatlas.org
<i>Rattus norvegicus</i>	RGD -Rat Genome Database	http://rgd.mcw.edu/
<i>Mus musculus</i>	MGD -The Mouse Genome Database MPD -The Mouse Phenome Database	http://www.informatics.jax.org http://aretha.jax.org/pub/cgi/phenome/mpdcgi?rtn=docs/home
<i>Caenorhabditis elegans</i>	WormBase -Biology and Genome of <i>C. elegans</i> RNAiDB - <i>C. elegans</i> RNAi Database PhenoBank - <i>C. elegans</i> RNAi Screens	http://www.wormbase.org http://nematoda.bio.nyu.edu http://www.worm.mpi-cbg.de/phenobank2/cgi-bin/MenuPage.py
<i>Drosophila melanogaster</i>	FlyBase -Database of the <i>Drosophila</i> Genome FlyMine -Database for <i>Drosophila</i> Genomics FlyRNAi - <i>Drosophila</i> RNAi Screening Center (DRSC)	http://www.flybase.org http://www.flymine.org http://www.flyRNAi.org
<i>Saccharomyces cerevisiae</i>	PROPHECY -Profiling of Phenotypic Characteristics in Yeast SGD - <i>Saccharomyces</i> Genome Database CYGD -Comprehensive Yeast Genome Database	http://prophecy.lundberg.gu.se http://www.yeastgenome.org/ http://mips.gsf.de/genre/proj/yeast.index.jsp

FIGURE 27: DB DI DATI FENOTIPICI

Per l'*Homo sapiens* le fonti di dati fenotipici sono molto diverse, vanno da OMIM, *the Online Mendelian Inheritance in Man database* (<http://www.omim.org/>) [124] [125], un database curato manualmente in cui i geni umani sono collegati ai disordini genetici tramite descrizione in testo libero. Anche se rappresenta una buona revisione testuale la sua struttura testuale rende difficile il *parsing* automatico in elementi categorizzati. OMIM raccoglie 14.031 descrizioni di geni ed oltre 3.500 descrizione di fenotipi.



FIGURE 28: HOME PAGE DI OMIM

GenAtlas (<http://www.genatlas.org/>) [126] [127], è un ricco database di dettagli genici e di dati fenotipici con 22.689 geni, 4.965 fenotipi e 109.507 citazioni.

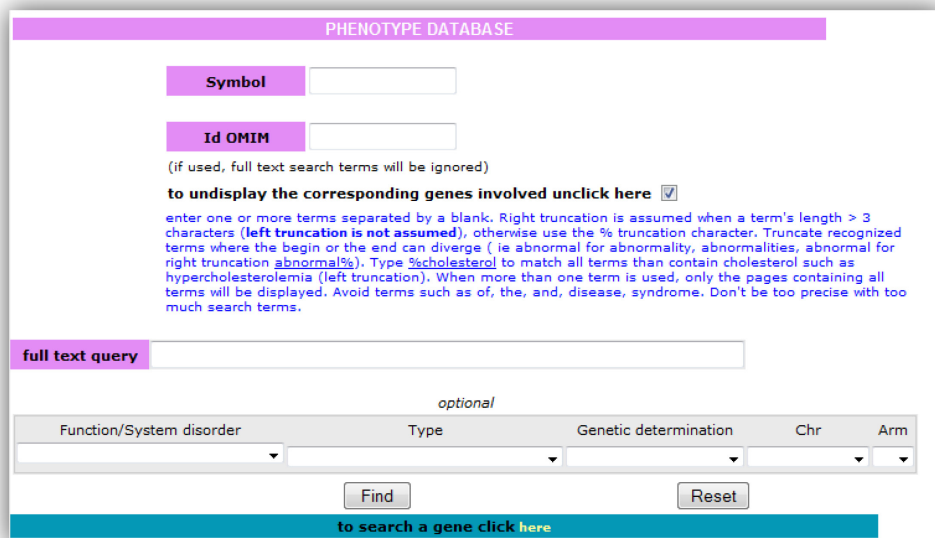


FIGURE 29: PAGINA DI RICERCA DEI FENOTIPI IN GENATLAS

HGMD, *the Human Genome Mutation Database* (<http://www.hgmd.cf.ac.uk/ac/index.php>) [128] [129] [130], è un database curato manualmente di malattie che causano mutazioni con un totale di 91.325 mutazioni e 3.498 geni.



FIGURE 30: HOME PAGE DI HGMD

PharmGKB, the *Pharmacogenetics and Pharmagenomics Knowledge Base* (<http://www.pharmgkb.org/index.jsp>) [131] [132] [133] [134], contiene dati su come la variazione genetica contribuisce a variazione nella risposta alle “drug”, il database contiene informazioni fenotipiche per oltre 400 geni.

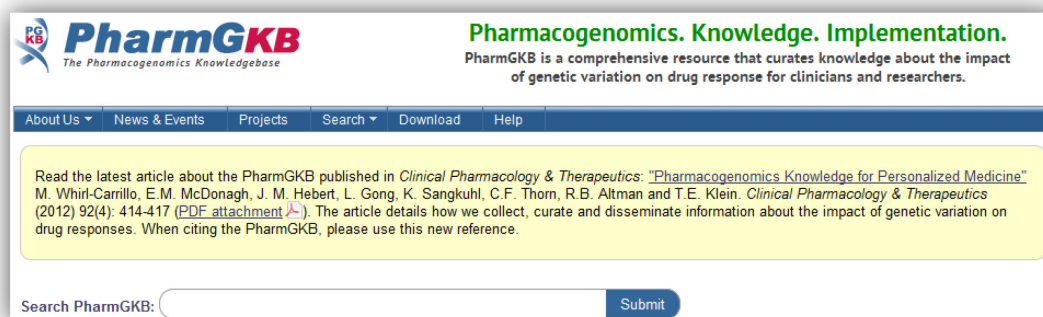


FIGURE 31: HOME PAGE DI PHARMKGB

2.4.1. PhenoGO

PhenoGO [135] utilizza un sistema esistente NLP (*Natural Language Processing*), chiamato BioMedLEE, un esistente organizzatore dei fenotipi (PhenoOS) in aggiunta con gli indici MeSH (*Medical Subject Headings*) ed ontologie mediche prestabilite. Più specificatamente PhenoGO aggiunge informazione fenotipica contestuale ad associazioni esistenti tra geni e termini GO come specificato in GOA (*Gene Ontology Annotations*). Il sistema inoltre mappa il contesto agli identificatori che sono associati con differenti ontologie biomediche, includendo *UMLS*, *Cell Ontology*, *Mouse Anatomy*, *NCBI taxonomy*, *GO*, a *Mammalian Phenotype Ontology*. In più PhenoGO valuta informazioni anatomiche e cellulari ed assegna i fenotipi al corretto GOA, i risultati mostrano che PhenoGO ha una precisione del 91%. PhenoGO è un database disponibile online al seguente link <http://www.phenoGO.org>.

2.4.1.1. Metodi

Una visione dei componenti e processi del sistema PhenoGO è illustrata nella seguente Figura.

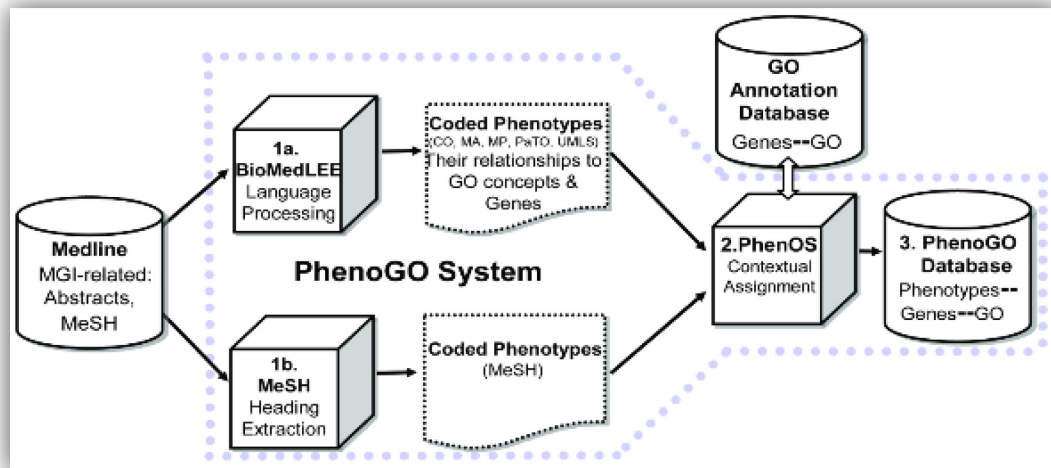


FIGURE 32: DIAGRAMMA SISTEMA PHENOGO

Per prima cosa gli *abstract* di Medline e le annotazioni GO dei geni sono identificati da GOA. Due distinti ed indipendenti processi estraggono il contesto fenotipico dagli *abstract*. PhenoGo può funzionare con uno o entrambi i processi.

Component 1a consiste del sistema **BioMedLEE NLP**, che processa il titolo e l'*abstract* e genera un *output* strutturato, che identifica geni, codici GO, fenotipi codificati (UMLS, CO; MA, MP), insieme con le relazioni gene-GO-contesto. **Component 1b**, semplicemente ottiene MeSH fenotipicamente rilevanti dagli *abstract* Medline. **Component 2**, il sistema di gestione PhenOS ottiene la coppia di termini gene-GO che è associato con l'articolo specifico nel database GOA. PhenOS (*Phenotype Organizer System*) è un sistema il cui scopo è quello di colmare il gap tra le terminologie biomediche eterogenee.

Taxon	Name	# Annotations
4896	<i>Schizosaccharomyces pombe</i>	344
4932	<i>Saccharomyces cerevisiae</i>	4,192
6239	<i>Caenorhabditis elegans</i>	12,212
7227	<i>Drosophila melanogaster</i>	91,782
7242	<i>Drosophila sp.</i>	238
7955	<i>Danio rerio</i>	3,142
9031	<i>Gallus gallus</i>	358
9606	<i>Homo sapiens</i>	102,262
9913	<i>Bos taurus</i>	804
10090	<i>Mus musculus</i>	427,275
10116	<i>Rattus norvegicus</i>	15,432
	Total	658,041

FIGURE 33: ANNOTAZIONI, PER SPECIE, CONTENUTE NEL DATABASE PHENOGO

2.4.2. PhenomicDB

PhenomicDB [136] è un database genotipo/fenotipo multi specie (liberamente disponibile al link <http://www.phenomicdb.de>) che mostra fenotipi associati con i loro corrispondenti geni e raggruppa le ortologie dei geni attraverso una varietà di specie. PhenomicDB preleva informazioni da una varietà di fonti tra cui OMIM, il *Mouse Genome Database* (MGD) [137], *WormBase* [138], *FlyBase* [139], il *Comprehensive Yeast Genome Database* (CYGD) [140], la *Zebrafish Information Network* (ZFIN) [141], e il *MIPS Arabidopsis thaliana database* (MAiDB) [142], la maggioranza di questi fenotipi è associata con geni mappati ad un comune indice, l'*Entrez Gene Index* di NCBI (*National Center for Biotechnology Information*). Geni funzionalmente equivalenti (ortologhi) da specie diverse sono raggruppate grazie al database *HomoloGene* [143] di NCBI.

PhenomicDB contiene 399.772 fenotipi, connessi a 77.400 geni di eucarioti. La percentuale di indici Entrez Gene con un fenotipo varia tra le specie. 99% per *D. melanogaster*, 79% per *C. elegans*, 21% per *Saccharomyces cerevisiae*, 16% per *Mus Musculus* ed 8% per *Homo sapiens*, l'84% di tutti i fenotipi in PhenomicDB provengono da *D. melanogaster* e *C. elegans*. Per *Homo sapiens* 2.850 geni sono collegati a 4.009 fenotipi e per 7.952 geni umani c'è almeno un "fenotipo ortologo" disponibile.

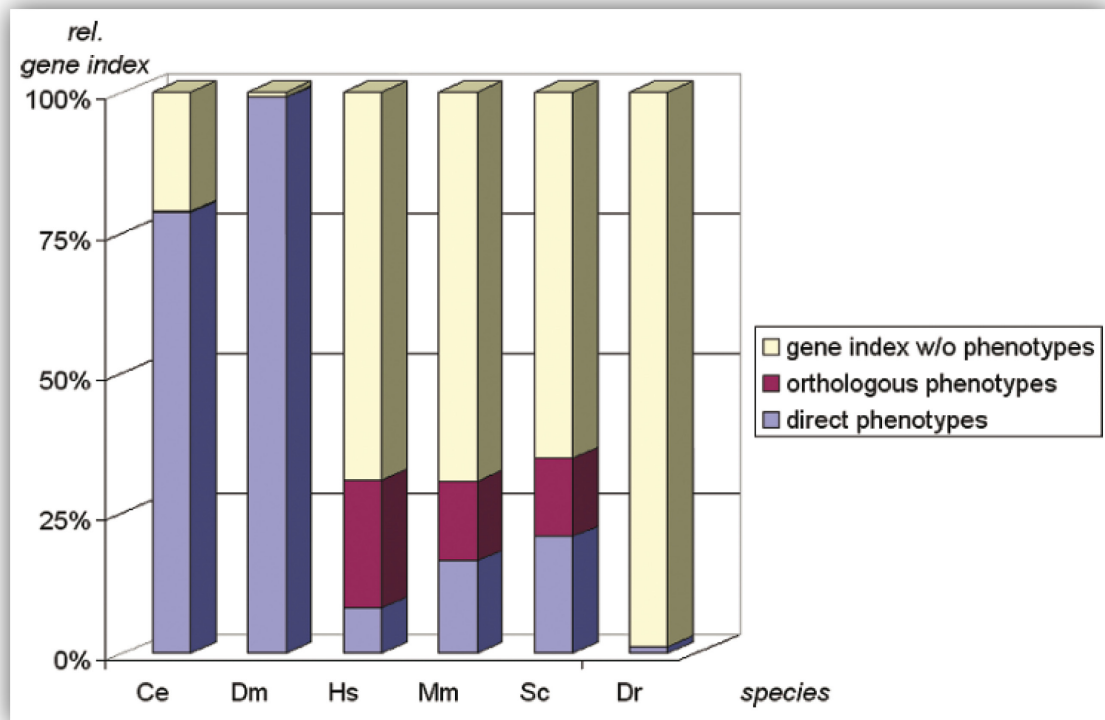


FIGURE 34: PERCENTUALE DI GENI CON UNO O PIÙ FENOTIPI DELLA SPECIE DATA (Ce, C. ELEGANS; Dm, D. MELANOGASTER; Hs, H. SAPIENS; Mm, M. MUSCULUS; Sc, S. CEREVISIAE; Dr, D. RERIO) È MOSTRATA IN BLU ('DIRECT PHENOTYPES'), DI CUI UNO O PIÙ FENOTIPI ASSOCIATI DALLE ORTOLOGIE SONO MOSTRATE IN ROSSO ('ORTHOLOGOUS PHENOTYPES'), E GENI CON NESSUN FENOTIPO ASSOCIATO SONO MOSTRATI IN GIALLO.

2.4.2.1. Phenoclustering

Il text-mining in PhenomicDb sulle descrizioni dei fenotipi sono filtrate, adattate e preparate creando quello che viene definito da Groth phenodoc, invece con phenocluster [144] ci si riferisce ad un cluster di phenodoc.

L'ortologia ci aiuta a trasferire informazioni sui fenotipi basandosi sui genotipi. Tuttavia è possibile trasferire informazioni funzionali basandosi su fenotipi simili [145]. In PhenomicDB i fenotipi simili sono clusterizzati usando la similarità testuale e mostra che i cluster emergenti sono biologicamente coerenti, permettendo una predizione della funzione dei geni con una precisione superiore al 70%. Per assicurare la significatività biologica, il sistema inoltre computa per ogni cluster un score di arricchimento in termini di interazioni proteina-proteina (PPI), similarità GO e similarità di sequenze delle proteine associate ai fenotipi. I cluster vengono creati usando l'algoritmo *vcluster* di CLUTO [146] [147]. *vcluster* è una implementazione scalabile del ben noto algoritmo k-means per dati testuali. Per calcolare lo score di similarità tra phenodoc, ognuno di essi è trasformato in una rappresentazione a vettore con gli score TFIDF per ogni parola (TF=term frequency, IDF=inverse document frequency, TFIDF=TF*IDF). Ogni phenodoc consiste dello stesso numero di feature, alcune

delle quali possono essere zero, dove la parola rappresentante la feature non è trovata nel phenodoc. Inoltre vcluster necessita di una misura per giudicare la similarità di due dati vettori phenodoc, la più popolare è la distanza Euclidea, ma per il clustering del testo, la misura del coseno è considerata più adatta [148]. La similarità tra due vettori viene determinata dal prodotto normalizzato dopo la normalizzazione della lunghezza del vettore. Gli score risultanti variano da 0 (nessuna similarità) ad 1 (egualità) e rappresentano il grado di correlazione tematica poiché in uno spazio di feature definito da termini, la direzione di un vettore può essere interpretata come il tema del testo.

2.4.3. CAESAR

CAESAR (*CAndidatE Search And Rank*) [149] è un sistema computazionale per selezionare geni candidati per tratti complessi. CAESAR usa tecniche di text e data mining per classificare geni in accordo con il loro potenziale coinvolgimento in un tratto complesso. CAESAR sfrutta la conoscenza dei tratti complessi in letteratura usando le ontologie per mappare semanticamente le informazioni del tratto ad una varietà di fonti di informazioni basate sui geni e proteine, includendo espressione dei geni, domini di proteine, interazioni proteina-proteina, pathway metaboliche e i geni omologhi di fenotipi mutati. CAESAR è disponibile online <http://visionlab.bio.unc.edu/caesar/>. CAESAR è composta di tre step principali.

2.4.3.1. Text-Mining

CAESAR richiede un corpo testuale come input, che può essere fornito dall'utente oppure usare i record di OMIM. Poiché il corpus è scritto in linguaggio naturale esso deve essere convertito in una forma adatta per le macchine, questo viene realizzato in due passi. Per prima cosa vengono identificati i simboli dei geni, i geni sono pesati basandosi sulla frequenza di occorrenze nel corpus, f_g , dove il peso c_g del gene estratto g è calcolato come f_g diviso la somma di tutti gli f_g degli n geni estratti. Secondo, il corpus è usato per quantificare la rilevanza dei termini all'interno di differenti ontologie biomediche, le ontologie usate in CAESAR sono *gene ontology biological process* (GO bp), *molecular function* (GO mf) [150], *mammalian phenotype ontology* (MP) [100] e *eVOC anatomical ontology* [151]. La rilevanza è quantificata usando la seguente similarità. Per ogni ontologia, i termini sono divisi in due separati documenti contenenti il nome del termine e la sua descrizione se disponibile. Questi documenti insieme formano un database o uno spazio di ricerca contro cui il corpus deve essere interrogato. Il corpus ed ogni

documento sono convertiti in vettori $v_i = \langle w_{i1}, w_{i2}, \dots, w_{in} \rangle$ con dimensione uguale alla dimensione dello spazio della parola n , che è il numero totale di parole uniche nel database. Ogni elemento del vettore per il documento i è calcolato come $w_{ij} = e_{ij}$, dove e_{ij} è il numero di occorrenze della parola j nel documento. La similarità del corpus ad ogni documento è calcolato come il coseno dell'angolo tra vettori, un valore ampio indica che la similarità è grande.

2.4.3.2. Data-Mining

Otto fonti di informazioni basate sul gene sono usate per mappare i termini ontologici ai geni annotati con essi. Il risultato è costituito da otto liste di score di geni, una per ogni categoria funzionale. Lo score r_{ij} del gene i per la fonte j è poi calcolata o come il massimo, somma o media dei pesi dei k termini ontologici che fanno *match* o dei c_1, \dots, c_k geni estratti.

2.4.3.3. Data-Integration

Gli score dei geni dalle otto fonti sono integrate per un unico combinato score per ogni gene. L'integrazione è compiuta seguendo uno dei quattro metodi, ognuno dei quali rappresenta un differente approccio che un ricercatore potrebbe scegliere quando manualmente sceglie i geni candidati sulle basi delle prove da diverse fonti.

I primi tre metodi coinvolgono massimo, somma e media dello score r_{ij} z-trasformato per ogni gene. Il massimo favorisce geni con una forte prova da una fonte di dati, la somma favorisce i geni con prove in molte fonti di dati e la media favorisce geni solo con forte prova, penalizzando i geni con alcune prove deboli. Il massimo, media e somma sono riferiti come *int1*, *int2* e *int3* rispettivamente. Gli score trasformati sono calcolati come $z_{ij} = (r_{ij} - \bar{u}_j) / s_j$, dove \bar{u}_j è la media e s_j la deviazione standard degli score dalla fonte j . Lo score combinato $\varphi_{:,i}$ è poi ottenuto calcolando

il massimo

$$\varphi_{int1,i} = \max z_{ij}$$

La media

$$\varphi_{int2,i} = \sum_{j=0}^n z_{ij} / n$$

O la somma

$$\varphi_{int3,i} = \sum_{j=0}^n z_{ij}$$

Degli score trasformati per il gene i .

Il quarto metodo indicato come int4, differisce dagli altri tre perché considera il numero di geni ritornati per quella data fonte. Un score trasformato s_{ij} è ottenuto

$$s_{ij} = \frac{r_{ij}}{\sum_{i=0}^n r_{ij}}$$

Gli score dei geni trasformati sono poi sommati insieme per fornire un singolo score per ogni gene

$$\varphi_{int4,i} = \sum_{j=1}^J s_{ij} \frac{g_j}{G}$$

Dove g_j è il numero di geni ritornati dalla fonte j e $G = \sum_{j=1}^J g_j$.

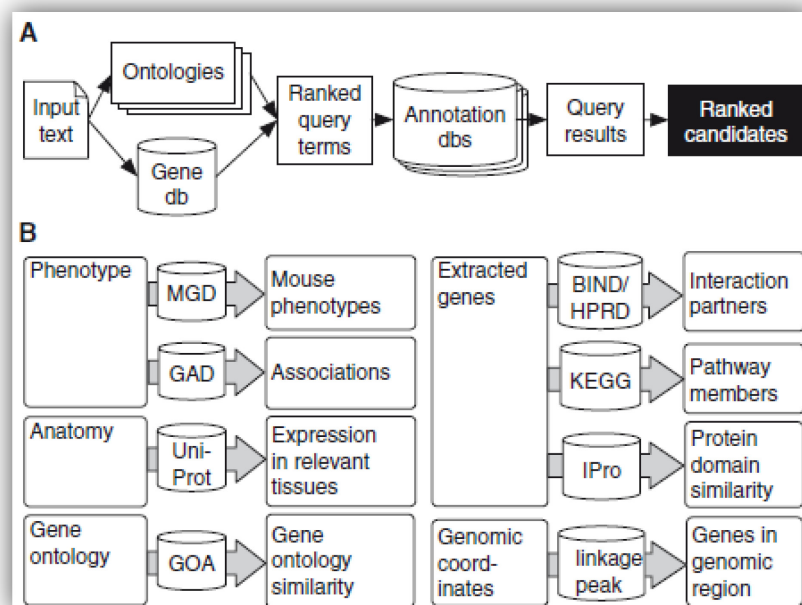


FIGURE 35: CAESAR: A) IL TEXT-MINING È USATO PER ESTRARRE I SIMBOLI DEI GENI E I TERMINI ONTOLOGICI DALL'INPUT. B) OTTO TIPI DI INFORMAZIONI FUNZIONALI SONO INTERROGATI USANDO I GENI ESTRATTI E ANATOMIA, FENOTIPI E I TERMINI DEI GENI ONTOLOGICI.

2.4.4. MimMiner

van Driel et al usano tecniche di text-mining per classificare oltre 5000 fenotipi umani contenuti in OMIM, dimostrando che la similarità tra fenotipi si riflette su moduli biologici sull'interazione di geni collegati funzionalmente. Il raggruppamento dei fenotipi dimostra la natura modulare della genetica delle malattie umane, quindi il *mapping* delle malattie può essere usato per predire geni candidati per malattie così come relazioni funzionali tra geni e proteine. MimMiner [152] è disponibile on line <http://www.cmbi.ru.nl/MimMiner>.

2.4.4.1. Metodi

OMIM è un database di informazioni in cui i record sono in forma testuale, un gene o un disordine genetico per record, inoltre contiene riferimenti della letteratura ad altri database. In MimMiner vengono usati due sezioni, fulltext (TX) e clinical synopsis (CS) di tutti i record che descrivono disordini genetici, una combinazione di questi viene intesa come 'record'.

Per estrarre informazioni da OMIM si usano le sezioni di anatomia (A) e disease (C) del vocabolario MeSH. Ogni termine MeSH e i loro plurali e componenti sono concetti. MeSH fornisce un modo standardizzato per recuperare informazioni quando si usano differenti terminologie per riferirsi agli stessi concetti. La sua dimensione e struttura gerarchica lo rendono un ricco dizionario che è necessario per fare il *match* con i testi di OMIM. I concetti di MeSH servono come feature fenotipiche caratterizzanti i record OMIM; ogni *entry* nel vettore delle feature rappresenta un concetto MeSH. Il numero di volte che i termini per un dato concetto sono trovati in un record OMIM riflette la rilevanza del concetto al fenotipo. Concetti non specifici come 'syndrome' o 'disease' sono esclusi.

I concetti MeSH possono essere molto ampi come 'Eye' o più specifici come 'Retina', ma al suo interno la struttura gerarchica descrive le relazioni tra 'Eye'-'Retina'-'Photoreceptors'. 'Eye' è iperonimo di 'Retina' che a sua volta è iperonimo di 'Photoreceptors'. Inversamente 'Retina' è iponimo di 'Eye'. Per assicurarsi che i concetti di 'Eye' e 'Retina' siano riconosciuti come simili si usa la gerarchia MeSH per codificare questa similarità nei vettori delle feature incrementando il valore di tutti gli iperonimi come descritto:

$$r_c = r_{c, \text{counted}} + \frac{\sum r_{\text{hypo}'s}}{n_{\text{hypo}'s}}$$

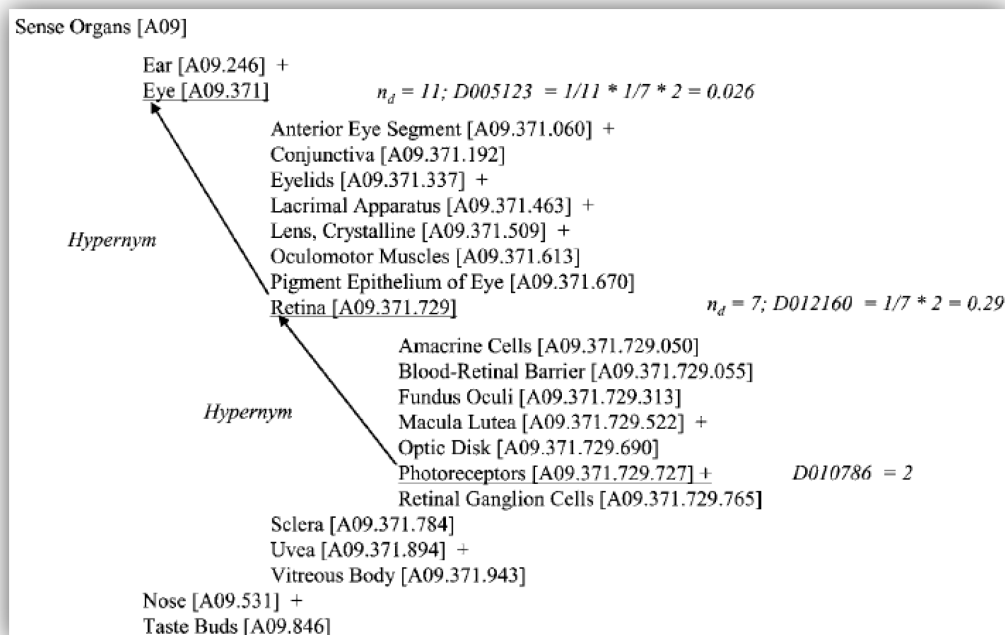


FIGURE 36: ESEMPIO DELL'ESPANSIONE DEL CONCETTO USANDO LA STRUTTURA MESH GERARCHICA. IL CONCETTO 'PHOTORECEPTORS' ESPANSO RESTITUISCE L'IPERONIMO 'RETINA' CHE A SUA VOLTA È IPERONIMO DI 'EYE'

Per qualsiasi concetto c , la sua rilevanza r_c diventa il numero attuale del concetto in un documento $r_{c, \text{counted}}$ più la sommatoria della rilevanza degli iponimi del concetto $r_{\text{hypo}'s}$. Questa somma è divisa per il numero di iponimi $n_{\text{hypo}'s}$. Questa equazione è applicata iterativamente dal livello più dettagliato nell'albero MeSH finché il livello dell'iperonimo più alto è raggiunto. Non tutti i concetti nei record OMIM sono equamente informativi, per esempio 'retina pigment epithelium' occorre raramente, e quindi fornisce informazioni più specifiche rispetto ai termini che occorrono più frequentemente come 'Brain'. Le differenze nell'importanza delle frequenze del concetto sono permesse usando la misura 'inverse document frequency':

$$gw_c = \log_2 \frac{N}{n_c}$$

La frequenza inversa del documento o peso globale del concetto c (gw_c) è il logaritmo del numero totale di record analizzati (N), diviso il numero di record che il concetto contiene c , n_c . Non tutti i record OMIM contengono descrizioni estensivamente uguali, queste differenze rendono difficili paragoni tra record a causa della diversità e la frequenza dei concetti nei record più grandi poiché eccederanno i record più piccoli. La seguente equazione viene usata per correggere parzialmente queste differenze:

$$r_c = 0.5 + 0.5 * \frac{r_c}{r_{mf}}$$

Il peso locale del concetto c in un record è una funzione della frequenza del concetto r_c diviso la frequenza del concetto MeSH che occorre maggiormente in quel record r_{mf} .

La similarità tra record OMIM può essere quantificata paragonando i vettori delle feature che sono espansi e corretti. La similarità tra vettori di feature è determinata dal coseno dei loro angoli:

$$s(x, y) = \frac{\sum_{i=1}^l x_i y_i}{\sqrt{\sum_{i=1}^l x_i^2} \sqrt{\sum_{i=1}^l y_i^2}}$$

La similarità tra i vettori di feature X e Y ($s(x,y)$) è una funzione delle rispettive frequenze dei loro concetti x_i e y_i . L'indice i va da 1 al numero di concetti MeSH l .

Questo stesso metodo è stato usato in Eshmun (Vedi paragrafo 3.6) per il calcolo della matrice di similarità fenotipica (SSMP similarità score matrix phenotype), che sarà alla base del tool per individuare la relazione tra geni in base ai fenotipi delle malattie in cui essi risultano coinvolti. Sebbene la matrice di similarità fenotipica generata da van Driel sia liberamente disponibile si è scelto di implementare nuovamente questo algoritmo poiché le descrizioni dei fenotipi così come il loro numero in OMIM è in crescente aumento, insieme con il vocabolario MeSH, questo ovviamente incide sugli score e la dimensione della stessa matrice, infatti la matrice originale possedeva 5.132 record mentre la versione aggiornata 7.112, questo si traduce in una maggiore possibilità di trovare relazione tra geni così come un risultato qualitativamente migliore.

3. Drug

3.1. Drug Discovery

Drug Discovery è il processo con cui avviene la scoperta e/o il design delle *drug*. La fase più importante nel processo è la selezione del target, una molecola chiave coinvolta in una specifica *pathway* nella condizione di una malattia e che possiede il potenziale per essere modificato da uno stimolo esterno. Target potenziali, per esempio, sono proteine i cui geni sono sovra espressi o che sono associati con proteine difettose ma essi non sono adatti per essere targettati da una piccola molecola (*small molecule*). Le *drug* (farmaci) devono legarsi a queste proteine target per curare una condizione di malattie. Anche se fondamentale per il processo, l'identificazione del target è ampiamente basata su prove circostanziali dove una proteina è selezionata basandosi sui suoi riferimenti ad una malattia. Negli studi biologici e nella salute umana, singole proteine e geni sono esaminati uno alla volta per capire la loro funzionalità ed il loro contributo in un aspetto funzionale specifico dell'organismo. Questo ignora la complessità degli organismi multicellulari, che porta ad una comprensione limitata del corpo umano e del suo funzionamento, limitando quindi la capacità di prevedere, prevenire e correggere potenziali problemi di salute.

3.1.1. System Biology in Drug Discovery

I sistemi biologici ci permettono di capire come i sistemi molto dinamici funzionano così che si possano trovare i modi per "targettare" malattie.

I sistemi biologici formano un ciclo con esperimenti, risultati di analisi e nuove conoscenze.

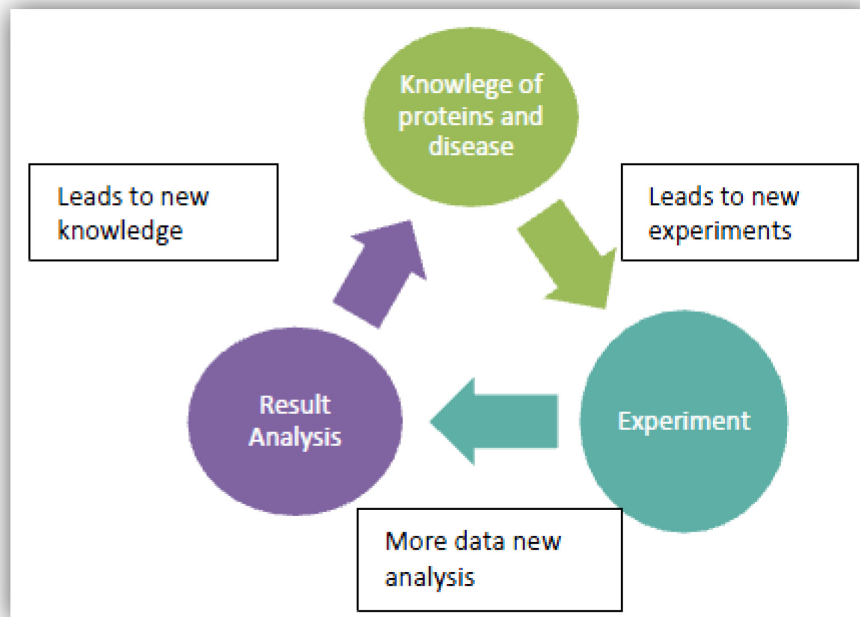


FIGURE 37: CICLO DEI SISTEMI BIOLOGICI PER GENERARE NUOVA CONOSCENZA

- Per comprendere la complessità dei sistemi biologici, è necessario eseguire esperimenti che di volta in volta generano una quantità di dati.
- I dati generati dagli esperimenti non avrebbero alcun valore senza analisi, quindi questo processo è fondamentale perché genera nuove conoscenze.
- La nuova conoscenza ottenuta come risultato di analisi di nuovi dati è poi usata per eseguire nuovi esperimenti, ricreando così il ciclo.

L'obiettivo dei moderni sistemi biologici è di comprendere la fisiologia e la malattia dal livello molecolare della *pathway*, *network* regolatorie, cellule, tessuti, organi e ultimamente l'intero organismo. Il termine "systems biology" ricopre molti differenti approcci e modelli per sondare e comprendere la complessità biologica, e studi di molti organismi dai batteri all'uomo. Molti studi accademici si concentrano sullo sviluppo di *tool* computazionali ed informatici per integrare una grande quantità di dati in modelli di *network* regolatorie e comportamenti cellulari.

I tre approcci principali sono:

1. Integrazione informatica dei dataset 'omics' (approccio bottom-up): gli approcci omics a sistemi biologici si focalizzano sulla costruzione di blocchi di sistemi complessi (geni, proteine e metaboliti). Questi approcci sono stati adottati dalle industrie farmaceutiche per completare gli approcci tradizionali per l'identificazione e la validazione del target, per generare ipotesi e per

l'analisi sperimentale in metodi tradizionali basati sulle ipotesi. La classificazione omics degli stati della malattia può portare ad un targeting più efficiente o anche personalizzazione delle terapie identificando *pathway* molecolari specifiche attive in particolari stati della malattia e in pazienti individuali.

2. Compute modeling di malattie o di sistemi organo (un approccio top-down per la selezione di target, indicazioni cliniche e il design di trial clinici): l'obiettivo è fornire un *framework* per generare ipotesi e predizioni basate sulla simulazione *in silico* della biologia della malattia umana attraverso distanze e mappe di scale multiple di un organismo.
3. L'uso di sistemi cellulari umani complessi per interpretare e predire le attività biologiche di drug e geni target (un approccio sperimentale diretto).

Hopkins et al [153] introduce il concetto di "*Druggability*", rivelando che delle quattro macromolecole, lipidi, acidi nucleici, polisaccaridi e proteine, sono principalmente queste ultime che possono essere usate come target. Ma non tutte le proteine sono "*druggable*", non possono tutte essere bersagli di drug, ma solo le proteine che hanno la capacità di legarsi e sovrapporsi ad una malattia possono essere targettati.

3.1.2. Approccio corrente nella Drug Discovery

Ad oggi la *drug discovery* si focalizza sull'approccio singolo-target: "*One target, one drug and one disease*" [154] con l'azione delle drug come chiave selettiva che si adatta nel blocco di uno specifico drug target. Nonostante il fatto che il paradigma basato sui target faciliti lo sviluppo di certe attività, ha anche alcuni inconvenienti:

- Il funzionamento non sempre è garantito. Potrebbe non sempre influenzare sistemi complessi nel modo desiderato, anche se cambia completamente il comportamento del suo obiettivo immediato. [155]
- Certe drug lavorano solo per una certa popolazione di pazienti. [156]
- Ignora l'enorme complessità delle cellule e dei tessuti. [155]
- Limita l'abilità dei ricercatori di identificare target innovativi e/o meccanismi di azione limitando lo spazio "*druggable*" per riconoscere i target. [157]
- Di solito non può combattere malattie multi-genetiche come il cancro, o malattie che influenzano tessuti multipli o tipi di cellule.

Yildirim et al [158] svilupparono una network di drug con i loro target. Attraverso la loro analisi della network essi rivelavano che anche se le drug al massimo hanno un solo obiettivo, ci sono drug che hanno obiettivi multipli e molte proteine target sono di volta in volta targettati da più di una drug.

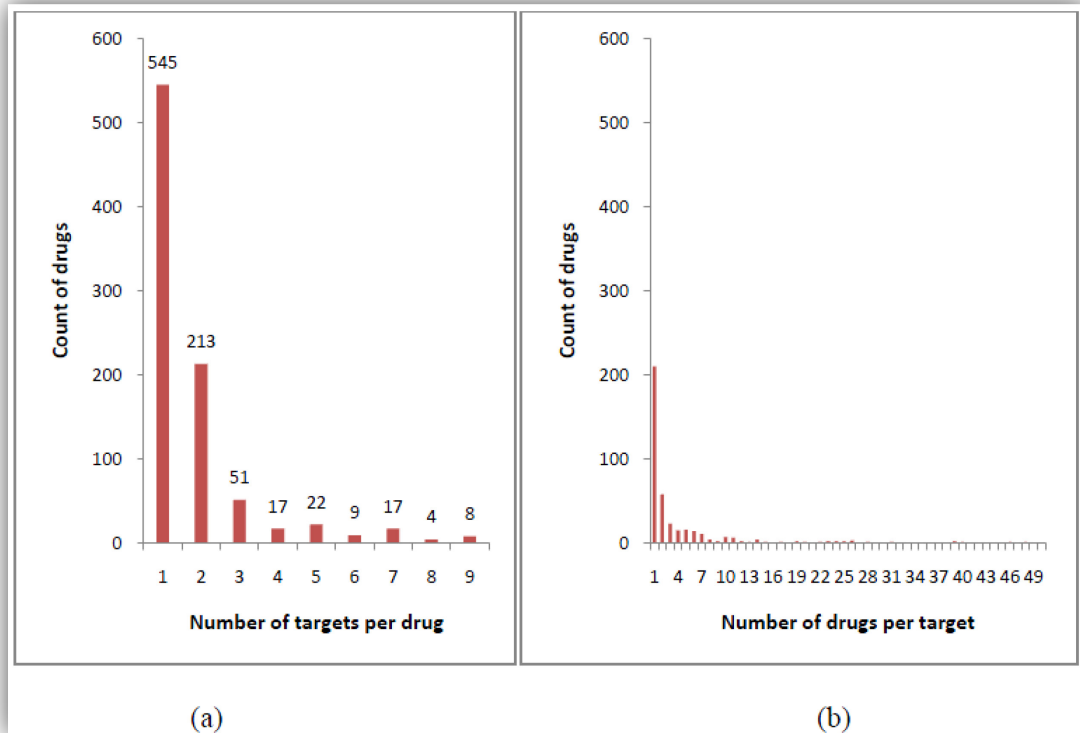


FIGURE 38: (a) Numero di target per drug (b) Numero di drug per target

Quando mappiamo la network drug-target in una network malattia-gene, si dimostra non solo che le drug agiscono su target multipli, ma i target sono inoltre coinvolti in malattie multiple [154]. Hopkin et al nel loro studio "Network Pharmacology" mappano la network delle drug sulla network delle interazioni proteina-proteina per rilevare tutte le proteine che sono targettate dalla stessa drug e la loro interconnessione. Essi hanno sottolineato il fatto che l'efficacia e la tossicità delle drug può essere ben compresa dalle azioni su specifici nodi e hub.

3.1.3. Drug Database

- DrugBank

DrugBank (<http://www.drugbank.ca/>) [159] è una database che combina informazioni dettagliate (chimica) sulle drug con informazioni sui target delle drug (proteine). Il database contiene 6.711 drug tra cui 1.447 sono *small molecule* approvate dal FDA, 131 sono drug biotech (proteine/peptidi) approvati dal FDA e 5.080 drug sperimentali. Più di 30 campi descrivono ogni drug, con dettagliate informazioni come il suo generico nome, indicazioni, tipo di drug, peso molecolare, tossicità, ect. I target delle drug sono descritti con il nome del target, nome del gene, Uniprot ID, ect.

DrugBank è un database web con una interfaccia user-friendly che permette la visualizzazione, l'ordinamento e l'estrazione di drug e/o target drug.

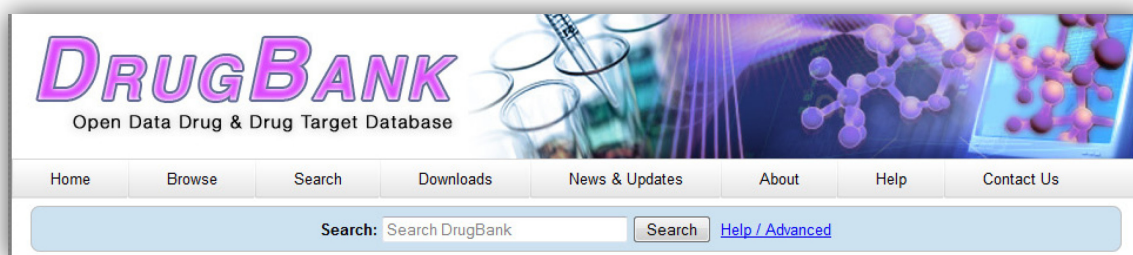


FIGURE 39: SITO DRUGBANK

- SuperTarget e Matador

SuperTarget fornisce informazioni circa le relazioni drug-target e integra tali relazioni da differenti fonti usando metodi eterogenei. Viene considerata una relazione drug-target come una specifica interazione di un piccolo composto chimico per trattare o diagnosticare una malattia. SuperTarget contiene oltre 332.828 interazioni drug-target una parte delle quali sono fornite separatamente da Matador (<http://matador.embl.de/>) (Manually Annotated Targets And Drugs Online Resource). Matador differisce da altre fonti come DrugBank poiché include interazioni dirette ed indirette.



FIGURE 40: SITO SUPERTARGET

Le relazioni descritte in SuperTarget sono ottenute in tre modi diversi. Iniziando con 2.400 drug e i loro sinonimi dal database SuperDrug, il tool di text-mining EbiMed viene usato per estrarre testo rilevante contenente relazioni drug-target da circa 15 milioni di abstract estratti da PubMed. Altre relazioni possono essere estratte da Medline cercando sinonimi di drug, proteine e MeSH descrittivi gruppi di proteine. Come ultimo step vengono prese le relazioni da altri database quali DrugBank, KEGG, PDB, SuperLigands e TTD.

3.2. Side Effect

Gli effetti collaterali sono risultati nocivi ed indesiderati di una drug o di altri effetti. Essi sono osservazioni del fenomeno complesse che sono state attribuite ad un numero di scenari molecolari, che includono l'interazione con i target primari o addizionali (off-target). Dopo che una drug entra in una cellula, può interagire o direttamente con un recettore o evocare risposte positive o negative da una up o down regolazione delle vie di segnalazione dei recettori.

Di tutte le cause, il legame degli off-target è il più importante. È stato trovato che il Viagra progettato per targettare PDE-5 e promuovere il rilassamento della muscolatura liscia si legava anche all'omologo PDE-6 nell'occhio che porta in pazienti ad un effetto collaterale definito "blue vision" [160]. Ciò evidenzia la necessità di studiare i target rilevanti nella network. L'approccio della network esamina gli effetti delle drug nel contesto di una network di interazioni rilevanti proteina-proteina [155]. Questo potrebbe aiutare a minimizzare gli effetti collaterali.

Gli effetti collaterali sono risposte al trattamento delle drug a livello fenotipico. Gli effetti collaterali possono fornire alcune informazioni importanti per i ricercatori in scienze farmaceutiche e tecnologiche. Per esempio, la similarità di effetti collaterali tra due drug potrebbe essere un'importante metrica per identificare nuovi target per le stesse [161]. Inoltre vicini di target di drug in una network biologica contribuiscono alla similarità degli effetti collaterali delle drug [162]. Poiché risulta importante per le industrie farmaceutiche razionalizzare il design delle drug, è importante investigare una distanza ottimale nella network tra i target delle drug e i geni delle malattie.

Wang et al [163] usando le network di segnalazioni umane, dimostrano che con una distanza di tre e quattro le drug mostrano una incidenza minima di effetti collaterali, invece drug con distanze tra il loro target ed i geni delle malattie più piccole di tre o maggiori di quattro mostrano una significativa incidenza di effetti collaterali.

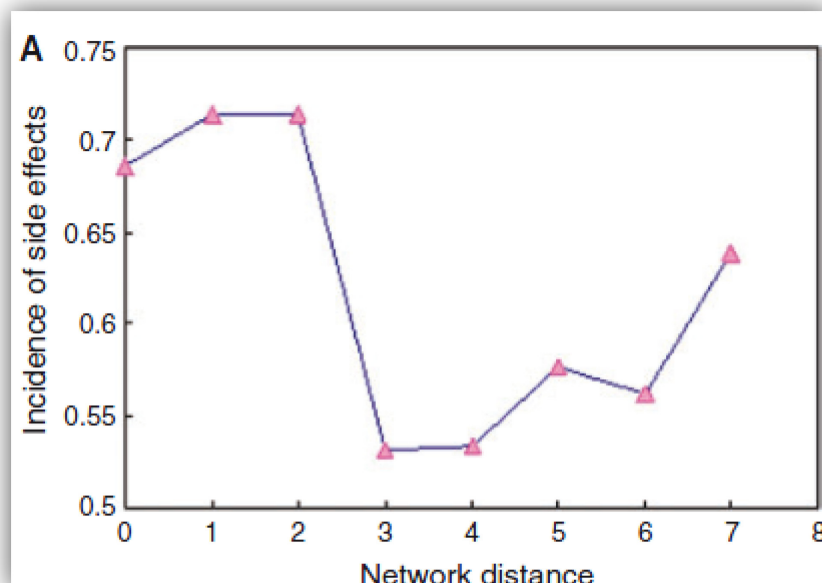


FIGURE 41: RELAZIONE TRA L'INCIDENZA DEGLI EFFETTI COLLATERALI DELLA DRUG E LA DISTANZA NETWORK DEI TARGET DELLA DRUG E DEI GENI DELLE MALATTIE.

Sebbene la ragione per cui una distanza minore di tre porti ad un aumento degli effetti collaterali rimane oscura, la disposizione dei geni della malattia nelle network biologiche possono fornire indizi. Infatti è noto che normalmente i geni della malattia hanno più interazioni con i geni non della malattia nelle network [164].

Sebbene le attività derivate dagli off-target siano inaspettate e nocive essi possono qualche volta avere dei benefici e portare a nuove indicazioni terapeutiche per le drug. Effetti collaterali simili di drug non collegati possono essere causati dai loro comuni off-target. Ad esempio, due diverse drug la cisapride e astemizole causano entrambi aritmia cardiaca poiché essi inibiscono il canale cardiaco hERG in aggiunta ai loro target primari (5). In generale, drug che *in vitro* hanno profili di legame proteico simili tendono a causare effetti collaterali simili (6,7), implicando una correlazione diretta tra target e la similarità degli effetti collaterali e perciò una possibilità di predire off-target. Campillos et al [161] hanno sviluppato una misura per la similarità degli effetti collaterali, analizzando la probabilità di condividere proteine target. Per classificare gli effetti collaterali, usano le ontologie UMLS (Unified Medical Language System) per sintomi medici ed estraggono termini rilevanti dai fogli illustrativi per le drug. Si usano le relazioni tra termini nelle ontologie per catturare le similarità tra drug annotate con termini distinti ma vicini. Non tutti gli effetti collaterali sono indipendenti da ogni altro, per esempio molte drug che causano nausea causano vomito.

3.2.1. *SIDER 2: Side Effect Resource*

La comprensione molecolare di fenotipi causati dalle drug negli umani è essenziale per chiarire meccanismi di azione e per sviluppare le medicine personalizzate. Gli effetti collaterali delle drug sono una fonte importante per l'informazione fenotipica umana. Gli effetti collaterali sono risposte fenotipiche dell'organismo umano al trattamento delle drug. Negli anni recenti, gli effetti collaterali sono diventati un soggetto importante dei ricercatori nell'industria farmaceutica, poiché possono inoltre essere usati per predire nuove interazioni drug-target. SIDER (<http://sideeffects.embl.de/>) [165] contiene 62.269 coppie drug-effetti collaterali e copre un totale di 888 drug e 1.450 distinti effetti collaterali. Il 70% delle drug hanno tra 10 e 100 differenti effetti collaterali.

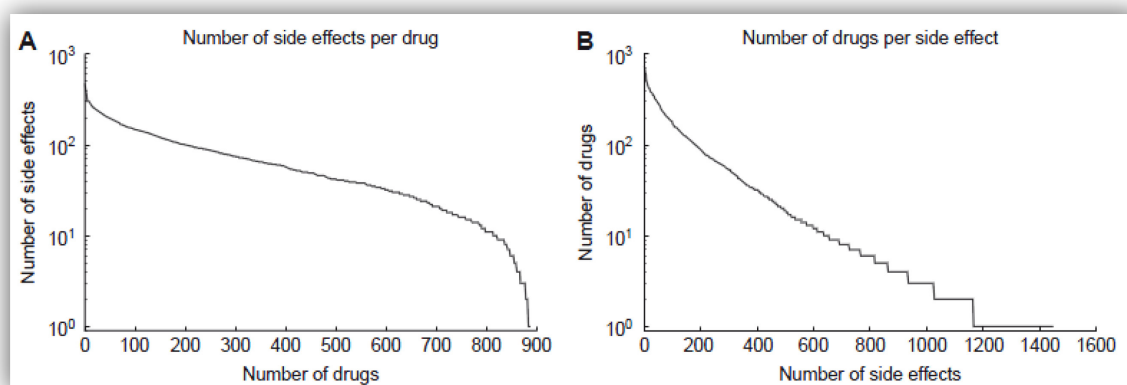


FIGURE 42: STATISTICHE DEL DATABASE. (A) IL NUMERO DI EFFETTI COLLATERALI PER OGNI DRUG (B) NUMERO DI DRUG PER EFFETTI COLLATERALI

In totale 798 di queste drug sono approvate dal FDA (Food and Drug Administration) le restanti 90 drug o sono state approvate precedentemente ma ritirate dal mercato o sono fuori dal mercato degli Stati Uniti.

Le drug in SIDER sono raggruppate seconda la loro classe, cioè, il gruppo anatomico principali della loro area di indicazione come indicato dal codice ATC (Anatomical Therapeutic Chemical Classification System) ed analizzate per quanto specifico l'effetto collaterale sia alla classe della drug. È chiaro che molti effetti collaterali possono occorrere per più che di una classe di drug.

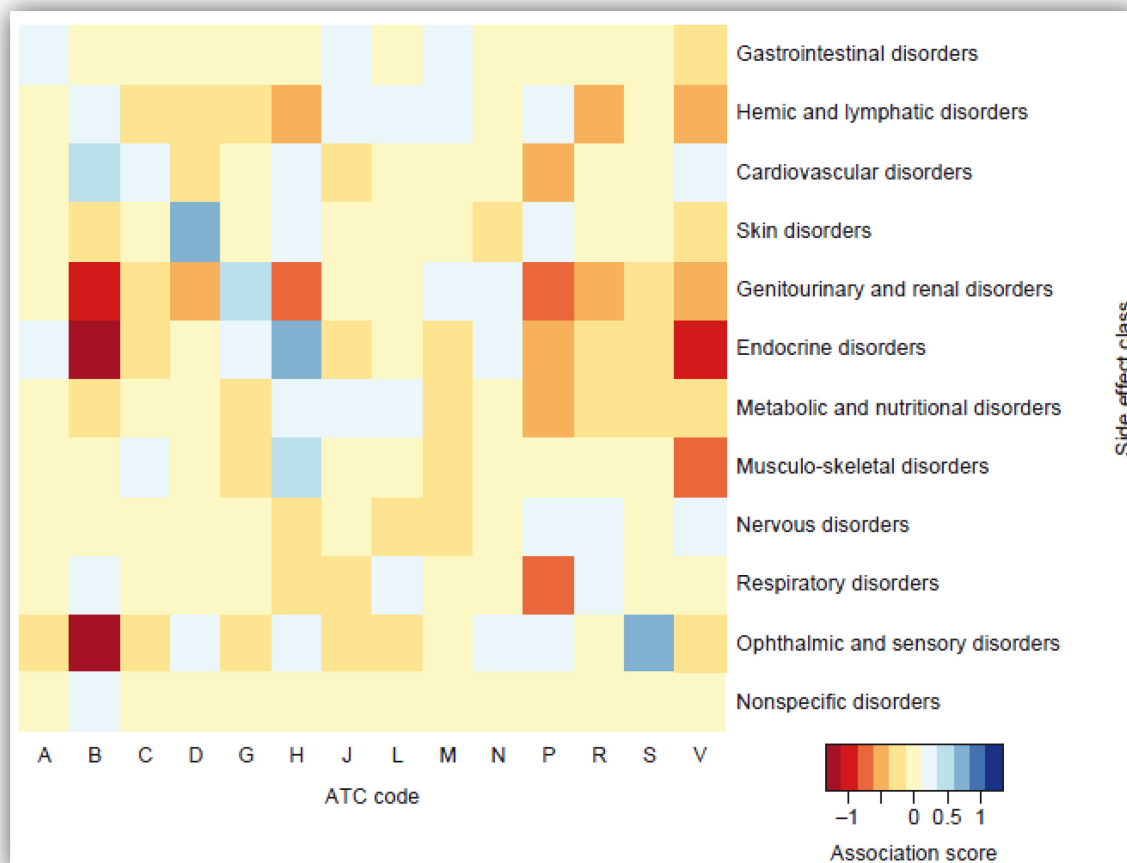


FIGURE 43: ASSOCIAZIONE TRA LE CLASSI DELLE DRUG E LE CLASSI ANATOMICHE DEGLI EFFETTI COLLATERALI.

3.3. Systems Pharmacology

La *system pharmacology* punta a capire le azioni e gli effetti collaterali delle drug considerando i target nel contesto delle network biologiche in cui essi esistono. La medicina genomica forma la base su cui la *system pharmacology* può svilupparsi. Approcci sperimentali e computazionali permettono alla *system pharmacology* di ottenere informazioni sulla network della malattia e sulla risposta delle drug, identificando anche nuovi target per le drug e specifiche combinazioni di esse. L'analisi network delle interazioni coinvolte nella patofisiologia e nella risposta delle drug tra varie scale di organizzazione, dal molecolare all'organismo, permetteranno l'integrazione della comprensione dei sistemi multi livello dell'azione delle drug con la medicina genomica. Molte malattie, come l'ipertensione, ulcera e diverse tipi di cancro, che due generazioni fa non potevano essere trattate, adesso possono essere gestite con successo, se non curate. La "*drugome*" (proteine e geni che sono targettati da drug approvate dall'FDA), coprono solo una piccola frazione della "*proteome*" o "*diseasome*" (geni che sono collegati con la malattia), e molte drug si focalizzano in poche aree [166] [158]. Questa

disparità mostra la relazione corrente tra scienza biologica e il suo uso per gli scopi terapeutici.

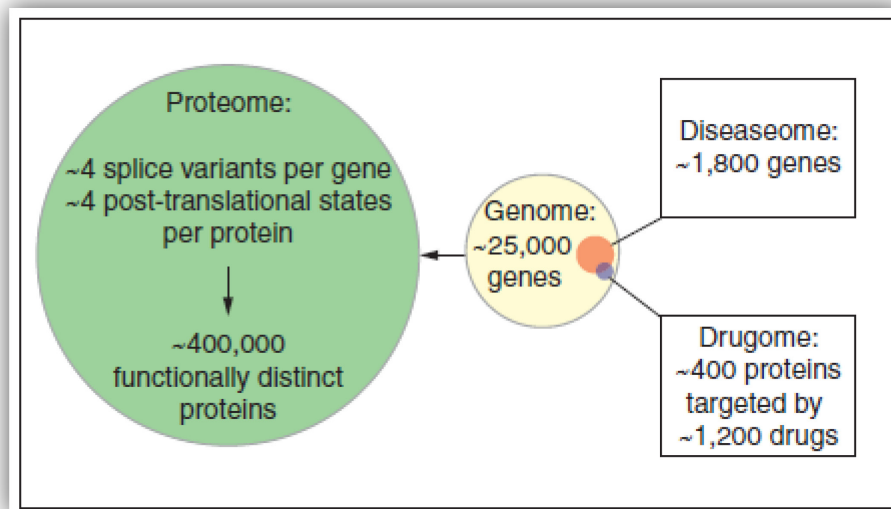


FIGURE 44: RELAZIONI TRA IL GENOME, PROTEOME, DISEASOME E DRUGOME

La *system pharmacology* cerca di sviluppare una comprensione globale delle interazioni tra patofisiologie e l'azione della drug. Per sviluppare una tale comprensione è necessario analizzare le interazioni tra le varie scale di organizzazione.

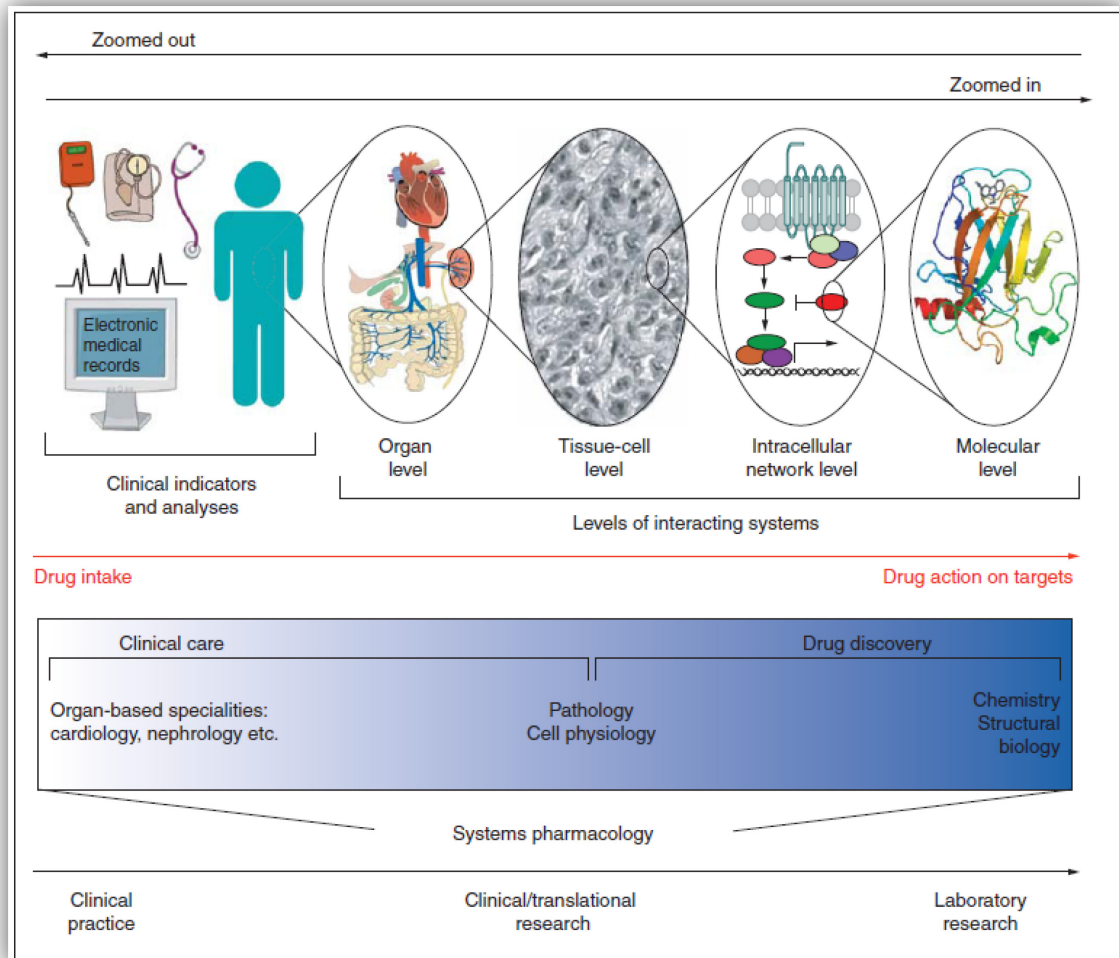


FIGURE 45: ANALISI MULTI-SCALA NELLA SYSTEM PHARMACOLOGY

L'analisi dei sistemi multi-scala richiede un'indagine *zoom in* o *zoom out* in relazione al tipo di analisi che si conduce. Le risposte fisiologiche ad una drug o malattia sono manifestazioni che possono essere studiate in una prospettiva *zoom-in*. L'intera network di eventi che sono coinvolti negli interventi terapeutici possono essere analizzati in studi proteomici o genomici, o, ad un livello più dettagliato, interazioni proteina-proteina o reazioni enzimatiche possono essere esaminate. La *system pharmacology* usa una serie di approcci che si estendono su scale multiple di organizzazione.

- **Approccio bottom-up**

Molti sforzi per comprendere sistematicamente i processi cellulari coinvolti nella malattia portano ad insiemi di proteine che sono mutate nello stato della malattia. Per ottenere una analisi multi-sfaccettata, tecniche come RNA interference (RNAi), gene-expression profiling, chromatin immunoprecipitation (ChIP) analisi o proteine microarrays sono

combinare. Queste analisi globali per identificare geni, proteine e altri componenti cellulari collegate all'origine o progressione di una specifica malattia producono un elenco di "seed nodes" che possono essere usati per costruire computazionalmente le network della malattia. L'identificazione di relazioni precedentemente non correlate tra pathway di malattie e funzioni cellulari è il primo step nell'identificazione dei target delle drug. Anche a livello di una singola proteina, una comprensione dei sistemi a livelli diventa utile. Una proteina mutata implicata in una malattia può avere cambi multipli nel suo comportamento nella network. Questi cambi devono essere inseriti nel contesto del "sistema" della malattia in modo da determinare quali proprietà hanno un impatto maggiore.

- **Approccio top-down**

Relazioni dirette tra malattie, drug e proteine permettono di inferire statistiche sulle loro relazioni meno evidenti. Un'analisi globale delle drug approvate dall'FDA ha trovato che i target delle drug tendono ad essere espressi nei tessuti specifici, tendono anche ad interagire con molte altre proteine nella network cellulare sebbene siano regolate indipendentemente [166] [158]. Metriche statistiche dalle analisi della network, come una misura di centralità, che quantifica la relativa importanza di una proteina nella comunicazione tra moduli differenti all'interno di una network, sono state suggerite per l'identificazione di nodi che hanno proprietà attrattive come potenziali target di drug [167]. Un lavoro recente ha generato una network che connette le drug sulle basi della loro similarità strutturali e la similarità dei profili degli effetti collaterali [161]. Questo metodo ha dimostrato l'efficacia nell'identificare gruppi di drug che condividono target comuni. Quando drug con target comuni condivisi non hanno di fatto target comuni, le predizioni forniscono un framework per testare queste drug per legami contro ogni altro target. Perciò, questo approccio porta all'identificazione di nuovi target per drug esistenti.

- **Integrare approcci bottom-up e top-down: network analysis**

Studi sulle pathway di segnalazione cellulare e network guidano la *system pharmacology* affinché possa portare a scoperte che permettano nuovi interventi terapeutici. Queste network sono spesso basate su informazioni genomiche. L'analisi network integra la medicina genomica e la *system pharmacology* come mostrato dalla sua disposizione centrale nella parte di sovrapposizione nel diagramma di Venn

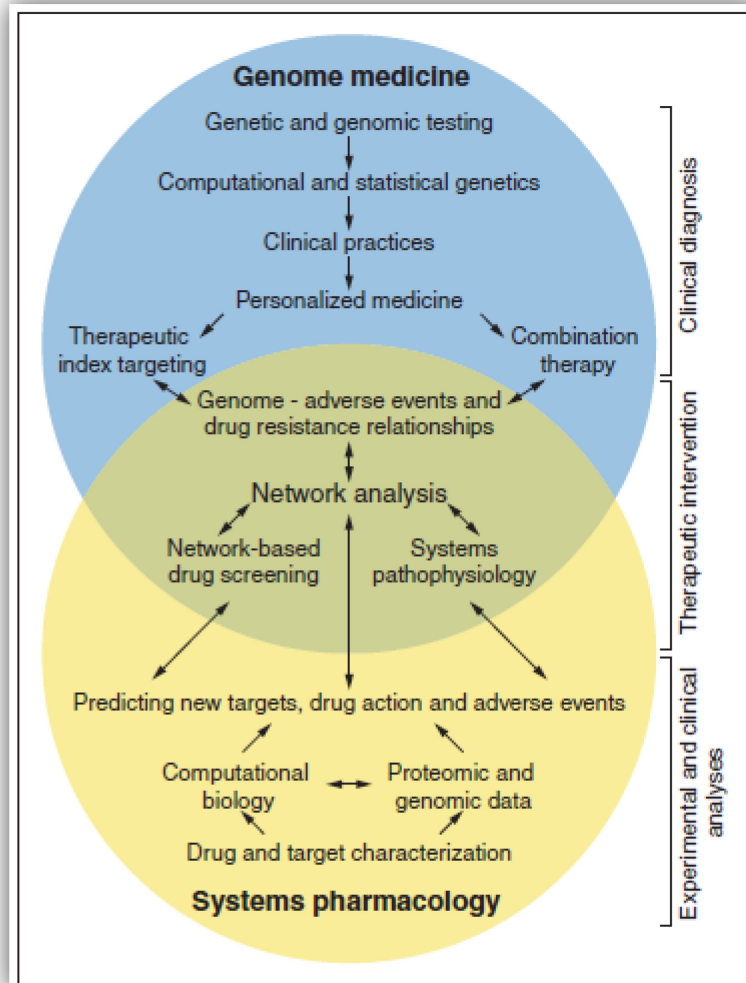


FIGURE 46: LE RELAZIONI TRA MEDICINA GENOMICA E SYSTEM PHARMACOLOGY. IL DIAGRAMMA RIASSUME I VARI ASPETTI DELLA MEDICINA GENOMICA (IN BLU) E LA SYSTEM PHARMACOLOGY (IN GIALLO), GLI ASPETTI IN OVERLAP DELL'ANALISI SONO IN VERDE.

L'analisi network ha facilitato lo sviluppo di algoritmi di predizione computazionale che predice tutte le possibile molecole influenzate da specifiche perturbazioni di target da drug ipotetiche [168]. Le analisi network suggeriscono che le analisi computazionali delle network di segnalazioni cellulari sono un aspetto importante della *system pharmacology* e della scoperta di future drug.

Gli approcci network in biologia hanno mostrato essere utili per l'organizzazione di dataset biologici molto grandi e per estrarre informazioni significative. Una network è un modo per la rappresentazione dei dataset enfatizzando le relazioni tra i nodi. Un grafo è costruito dove questi nodi, che rappresentano geni, proteine, *small molecule* o qualsiasi altra entità capace di interazione nel sistema sono connessi con archi, che rappresenta la natura dell'interazione. Nodi e archi possono avere vari attributi e annotazioni. Archi rappresentano interazioni tra i nodi e quando l'informazione è disponibile, gli archi

possono avere direzioni, pesi e altri attributi che forniscono informazioni circa la gerarchia degli effetti. Relazioni differenti tra nodi che possono essere usate per studiare drug sono rappresentate nella seguente figura.

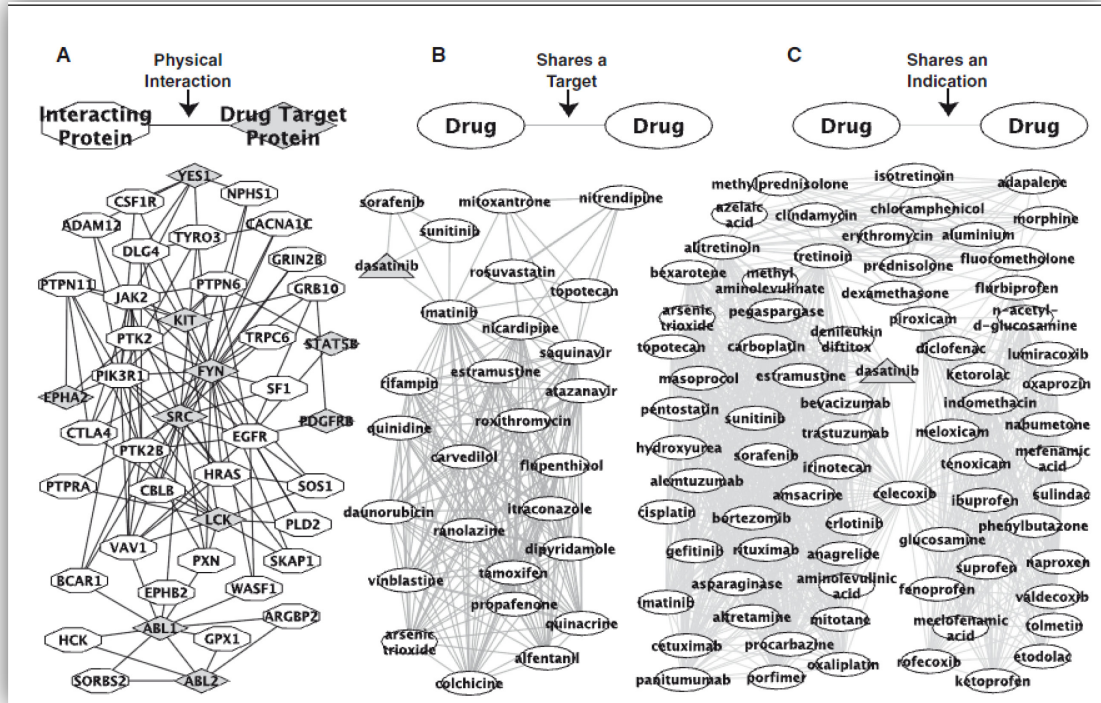


FIGURE 47: RELAZIONI DIFFERENTI TRA NODI NELLE NETWORK USATE NELLA SYSTEM PHARMACOLOGY. A) NODI SONO PROTEINE CONNESSE DA INTERAZIONI FISICHE. B) NODI OVALI SONO DRUG CONNESSE DA TARGET CONDIVISI ALL'INTERNO DEL DASATINIB, MOSTRATO IN UN TRIANGOLO. C) NODI OVALI SONO DRUG CONNESSE CONDIVIDENDO UNA INDICAZIONE TRAPEUTICA

Le network permettono l'integrazione di diverse fonti di dati sperimentali e conoscenze biologiche in un *framework* che fornisce nuovi indizi nei sistemi. Questi approcci possono combinare dataset genomici con informazioni specifici di geni e proteine. L'analisi network per studi farmacologici permette di identificare nuovi target delle drug per molte malattie, comprendere meglio cosa rende un target un buon target di drug, migliorare l'abilità di predire le combinazioni di drug efficienti e di effetti collaterali.

3.4. Network Analysis

Le malattie sono raramente la conseguenza dell'anormalità di un singolo gene, ma riflette le perturbazioni del complesso intracellulare e network intercellulare che collegano tessuti e sistemi organo. Le drug possono avere più che un solo target così che la loro efficacia è determinata dalle loro interazioni multiple portando ad effetti indesiderati. Le drug con

target singoli possono correggere alcuni aspetti disfunzionali del modulo ma possono anche alterare l'attività delle molecole che sono situate vicino nella network della malattia, portando ad effetti collaterali. Molte drug sono palliativi perché esse non targettano la proteina associata alla malattia ma le proteine vicini nella network. Ci si chiede quindi se è possibile identificare target multipli per drug che hanno un impatto ottimale sul fenotipo della malattia?

Le network che operano in sistemi biologici, tecnologici o sociali non sono random, ma caratterizzate da un set di principi organizzativi, i nodi sono proteine, metaboliti o malattie, e i legami sono interazioni proteina-proteine, reazioni metaboliche o geni condivisi. I geni delle malattie hanno caratteristiche uniche e quantificabili che li distinguono dagli altri geni? I geni delle malattie sono situati randomicamente nell'interattoma, o sono individuabili correlazioni tra la loro locazione e la topologia della loro network?

- **Localizzazione dei geni delle malattie all'interno della network:** Nelle network biologiche ci sono molti nodi che hanno pochi link e pochi nodi che sono altamente connessi, questi sono chiamati "hub". Le proteine rappresentate da questi hub hanno uno speciale ruolo biologico. Le proteine tendono ad essere codificate da geni essenziali, questa assunzione porta all'ipotesi che negli essere umani, gli hub sono associati con i geni della malattia.
- **Gene Essenziale:** Nella network di interazioni proteina-proteina la rimozione di proteine selezionate in maniera random non influenzano la topologia della network, ma al contrario quando le proteine più connesse sono eliminate il diametro della network aumenta rapidamente. Se questo invece è dovuto alla componente topologica le proteine meno connesse dovrebbero essere meno essenziali che quelle altamente connesse. Jeong et al [169] dimostrò che le proteine con cinque o meno collegamenti costituiscono circa il 93% del numero totale di proteine e solo il 21% sono essenziali. Di contro solo l'0,7% delle proteine del lievito con profili fenotipici conosciuti hanno più di 15 collegamenti, ma una singola delezione del 62% di questi aumenta la letalità. Questo implica che proteine altamente connesse con un ruolo centrale nell'architettura della network sono tre volte più essenziali che proteine con un piccolo numero di collegamenti con altre proteine.

Goth et al trovò che i geni essenziali che non sono associati con la malattia mostrano un forte tendenza ad essere associate con hub e sono espresse in tessuti multipli, essi tendono ad essere localizzati al centro funzionale dell'interattoma. Tuttavia, i geni della malattia non essenziali mostrano una tendenza a codificare hub, e tendono ad essere tessuto-specifiche e localizzate alla periferia dell'interattoma.

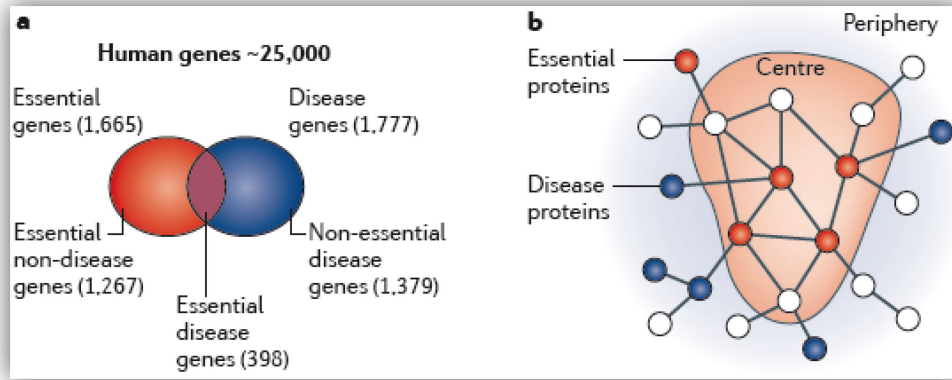


FIGURE 48: GENI ESSENZIALI E DISEASE NELL'INTERATTOMA. A) DEI 25.000 GENI UMANI, 2.418 SONO ASSOCIATI A SPECIFICHE MALATTIE, OVERLAP TRA 1.777 GENI ASSOCIATI ALLA DISEASE B) DIAGRAMMA SCHEMATICO DELLE DIFFERENZE TRA GENI DISEASE ESSENZIALI E NON ESSENZIALI

- **Clustering locale dei geni della malattia:**
 - **Ipotesi Locale:** Le proteine coinvolte nella stessa malattia hanno una tendenza crescente ad interagire con ogni altro.
 - **Corollario dell'ipotesi locale:** Mutazioni in proteine interagenti spesso portano a fenotipi simili nella malattia.
 - **Ipotesi "Disease Module":** Componenti cellulari associate con un fenotipo di una specifica malattia mostrano una tendenza a creare cluster nel vicinato della stessa network.

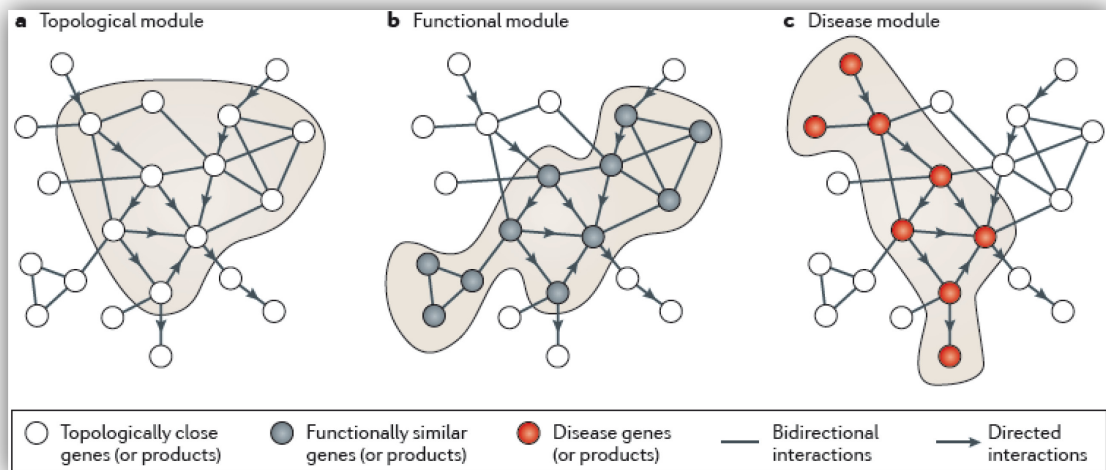


FIGURE 49: DIAGRAMMA SCHEMATICO DEI TRE CONCETTI DI MODULARITÀ, TOPOLOGICAL, FUNCTIONAL E DISEASE MODULE

- I moduli topologici corrispondono ad un vicinato localmente denso dell'interattoma, tali che i nodi del modulo mostrano una tendenza maggiore di interagire con ogni altro rispetto ai nodi fuori dal modulo.

- I moduli funzionali corrispondono alle network dei vicini in cui c'è una segregazione statistica significativa di nodi di funzioni correlati.
- Il modulo della malattia rappresenta un gruppo di nodi le cui perturbazioni (mutazioni, delezioni, ect) possono essere collegati a fenotipi di particolari malattie.

In letteratura c'è una tacita assunzione che questi tre moduli sono collegati e sovrapposti: i componenti cellulari che formano un modulo topologico hanno funzioni correlate vicine, una malattia è il risultato di un guasto di un particolare modulo funzionale, intimando che un modulo funzionale sia anche un modulo della malattia. Tuttavia un modulo di una malattia ha caratteristiche uniche:

- Un modulo della malattia potrebbe non essere identica, ma simile e con sovrapposizioni, ai modulo funzionale e/o topologico
- Un modulo della malattia è definito in relazione ad una particolare malattia e ogni malattia ha il proprio modulo unico
- Un gene, una proteina, un metabolita possono essere implicate in diversi moduli malattia, questo significa che diversi moduli malattia possono sovrapporsi.
- **Small-world phenoma:** In molte network complesse ci sono cammini relativamente brevi tra qualsiasi coppia di nodi. Questo significa che molte proteine (o metabolita) sono solo a poche interazioni (o reazioni) da qualsiasi altra proteina (o metabolite). Perciò perturbare lo stato di un dato nodo può influenzare l'attività di molti nodi nelle loro vicinanze così come il comportamento della network stessa.
- **Motivi:** Alcuni sottografi (o gruppi di nodi che si collegano ad ogni altro formando una sub-network all'interno della network) sono molto frequenti nelle network biologiche. Tali sottografi sono chiamati "*motifs*", ed essi sono associati con alcune funzioni biologiche ottimizzate.
- **Betweenness centrality:** Nodi con un alto "betweenness centrality" (una misura del numero di percorsi più brevi che attraversano ogni nodo) sono spesso dei colli di bottiglia.
- **Network parsimony principle:** Pathway molecolari casuali spesso coincidono con i cammini molecolari più brevi tra componenti noti associati alla malattie.
- **Shared components hypothesis:** La malattia che condivide componenti cellulari associate ad altre malattie (geni, proteine, metaboliti o microRNA) mostrano similarità e comorbidità fenotipica.

3.5. **Problemi e Ricerca**

Studi esistenti hanno indicato che è necessario cambiare la strategia corrente della selezione del target nella drug discovery che non ha successo nella lotta contro tutte le malattie. Un calo continuo nell'approvazione delle drug indica un bisogno immediato di rivedere le caratteristiche dei target delle drug candidati. L'importanza delle network biologiche nella drug discovery è stata riconosciuta dai alcuni ricercatori come non completamente implicita. Gli hub sono indicati come i target migliori ma la definizione di hub non è del tutto chiara. Vi è la necessità di entrare nei dettagli specifici su ciò che rende la proteine target di drug distinguibili dalla folla. L'essenzialità è un'altra importante caratteristica della network di una proteina ma se sia realmente considerata durante la selezione del target o quali effetti ha sugli effetti collaterali delle drug è ancora sconosciuto. Alcune questioni chiave sono:

- Gli hub/non hub nelle network sono usati preferibilmente come target delle drug?
- Quale range di connettività nella network di una proteina potrebbe essere meglio per candidare target?
- La selezione di una proteina come target la qualifica essere un buon target?
- L'analisi delle network aiuta nel trovare buoni target?
- L'essenzialità gioca un ruolo importante nella selezione dei target delle drug?
- C'è una relazione tra l'essenzialità dei target e gli effetti collaterali delle drug?
- È possibile trovare target cattivi analizzando le drug e i loro effetti collaterali?

In questo studio presento un tool web, ESHMUN (vedi paragrafo 3.6), basato su un approccio delle network biologiche per analizzare le malattie, fenotipi, geni, drug, effetti collaterali e transcription factor. L'approccio network permette di organizzare rapidamente le conoscenze integrando diversi tipi di ampi database. Si è scelto di usare l'analisi delle network poiché permette di comprendere malattie complesse che non possono essere comprese in termini di un singolo target o una singola pathway. Inoltre permette la predizione e la spiegazione di effetti inaspettati delle drug e suggerisce fattori che influenzano l'efficacia e la sicurezza delle drug. Eshmun fornisce 5 network, il core è sempre costituito dai nodi (geni) che sono collegati da archi che rappresentano la similarità fenotipica delle malattie in cui sono coinvolti, una comprensione globale di molte malattie infatti richiede una mappa delle interazioni genica, la cui disfunzione è responsabile per l'emergenza di un particolare fenotipo della malattia. A seconda della network in esame oltre ai geni ci saranno nodi che rappresentano drug, questo approccio nella farmacogenomica permette di evidenziare l'azione delle drug, identificare facilmente nuove strategie terapeutiche e potenzialmente sviluppare e prescrivere delle drug in maniera più sicura ed efficiente. Da queste analisi è possibile iniziare a formulare

un set di criteri network che definiscono un buon target per le drug e permette la selezione di nuovi target dalla network. Per esempio, le misure di centralità di una network possono collegarsi all'importanza di un nodo e alla sua capacità di distruggere un processo biologico di interesse. In ESHMUN usiamo il concetto di network biologica per integrare dati da diversi database come DrugBank, OMIM, STRING, TargetScan ed altri, i risultati mostrano che i target noti rimangono i target preponderanti di nuove drug, con una recente lenta diversificazione delle proteine target. Le distanze metriche mostrano che molte delle drug siano palliativi e non perturbano direttamente le proteine corrispondenti la sottostante causa della malattia.

3.6. ESHMUN

3.6.1. Introduzione

Le interazioni tra i geni, microRNA e proteine, insieme con il loro livello di espressione determinano i fenotipi negli organismi, tali fenotipi possono essere modulate dalle drug. Ottenere un quadro comprensivo capace di evidenziare insieme tutti gli elementi responsabili di un fenotipo aiuterebbe nello studio delle malattie o fenomeni biologici. Tuttavia, guardando alla letteratura, sono stati fatti relativamente pochi sforzi per usare i fenotipi per andare oltre le singole relazioni genotipo-fenotipo. L'origine genetica di una malattia è spesso scoperta una volta che i suoi fenotipi sono stati chiaramente definiti, inoltre i tratti complessi di una malattia spesso comprendono una varietà di fenotipi e meccanismi biologici, rendendo difficile l'identificazione dei geni da studiare. Per molto tempo, i fenotipi sono stati visti come indicatori per cambi nei genotipi o nelle malattie. L'abilità dei fenotipi per interferire con componenti genetici in una maniera sistematica ha aumentato la loro importanza come mezzo per capire i processi biologici a livello molecolare. Invece, l'esistenza di specifici gruppi di fenotipi per una malattia suggerisce che la fenomica è possibile. Alcuni importanti studi sui fenotipi sono stati condotti da van Driel et al, che propose una tecnica di text mining per classificare 5.000 fenotipi umani classificati in OMIM, dimostrando come la similarità tra fenotipi dipende da relazione funzionali tra geni. Gaulton et al svilupparono un sistema per classificare tutti i geni umani annotati come candidati per fenotipi complessi, usando le ontologie per mappare semanticamente le descrizioni in linguaggio naturale dei fenotipi. Recentemente, Groth et al proposero un nuovo approccio per predire l'annotazione di geni, usando un 'phenoclustere' per esaminare la correlazione dei geni usando le

interazioni proteina-proteina (PPI) di proteine associate e annotazioni funzionali dalla Gene Ontology (GO).

Sebbene i fenotipi hanno geni come attori centrali, molte altre molecole e fattori determinano il loro comportamento. Per esempio, i fattori di trascrizione (Transcription Factors TF) sono responsabili per la trascrizione del gene, miRNA, drug e proteine sono coinvolte nella loro regolazione. I microRNA (miRNA) sono piccoli RNA non codificanti coinvolti nella regolazione dell'espressione genica post-trascrizionale. La loro importanza biologica, inizialmente dimostrata nel cancro, si è rivelata centrale in molte altre patologie come Alzheimer, Parkinson, infezioni virali, diabete e miopatie. Inoltre, i polimorfismi nei miRNA stanno emergendo come potente tool per studiare la biologia di una malattia ed hanno il potenziale per essere usati nella prognosi e diagnosi di diverse malattie. Inoltre i polimorfismi influenzano la funzione dei microRNA. Quindi quando target microRNA traducono per proteine target di drug questo potrebbe portare a resistenza ai drug o ad una loro efficacia. Dall'altro lato particolari fenotipi potrebbero essere i risultati degli effetti collaterali di particolari drug. Gli off-target delle drug dovrebbero essere identificate sistematicamente usando la misura di similarità delle drug. Inoltre, i Fattori di Trascrizione sono elementi critici nella corretta regolazione dei miRNA, infatti prove crescenti mostrano che le variazioni fenotipiche nelle malattie sono causate da una regolazione anomala dei miRNA da parte dei TF. È quindi importante studiare in maniera comprensiva i fenotipi causati dai geni nel livello della network considerando tutti questi elementi. ESHMUN (side effectS pHenotypes Mirnas drUgs geNes) è un sistema integrato con una interfaccia web che permette agli utenti di annotare funzionalmente un set di geni. L'annotazione dei geni in ESHMUN consiste nella creazione di alcune network ibride: 1) una network di similarità fenotipica, in accordo con la definizione di van Driel, in cui i geni sono connessi quando essi condividono gli stessi aspetti fenotipici di una malattia conservata in OMIM, 2) due network di geni-miRNA usando target predetti (TargetScan 6.0) e validati (miRecords, miRTarbase, Tarbase), 3) una network geni-drug prelevati da DrugBank insieme con gli effetti collaterali di SIDER, 4) una network in cui i geni sono connessi sulla base delle interazioni tra le proteine che essi traducono ottenute da STRING. Inoltre, ESHMUN fornisce un set di network predefinite create dalle pathway dei geni di KEGG. Quindi ESHMUN fornisce una fonte molto ricca di conoscenza per i ricercatori che conducono studi basati sulle network.

3.6.2. ESHMUN BackEnd

ESHMUN è un tool web-based che integra informazioni da database biologici specializzati. Il cuore del sistema è sviluppato in linguaggi di programmazione quali Ruby ed R, ed usiamo script in PHP per generare dinamicamente pagine web. Lo scopo principale di ESHMUN è di fornire un sistema aggiornato periodicamente in cui l'utente può analizzare una pathway di geni secondo la loro similarità fenotipica e facilmente identificare, grazie ad una visualizzazione basate sulle network, miRNA predetti e validati, drug e interazioni proteina-proteina coinvolti.

ESHMUN può essere diviso in classi che usano diversi database, come mostrato in figura, che rappresentano vari livelli di informazioni che saranno disponibili.

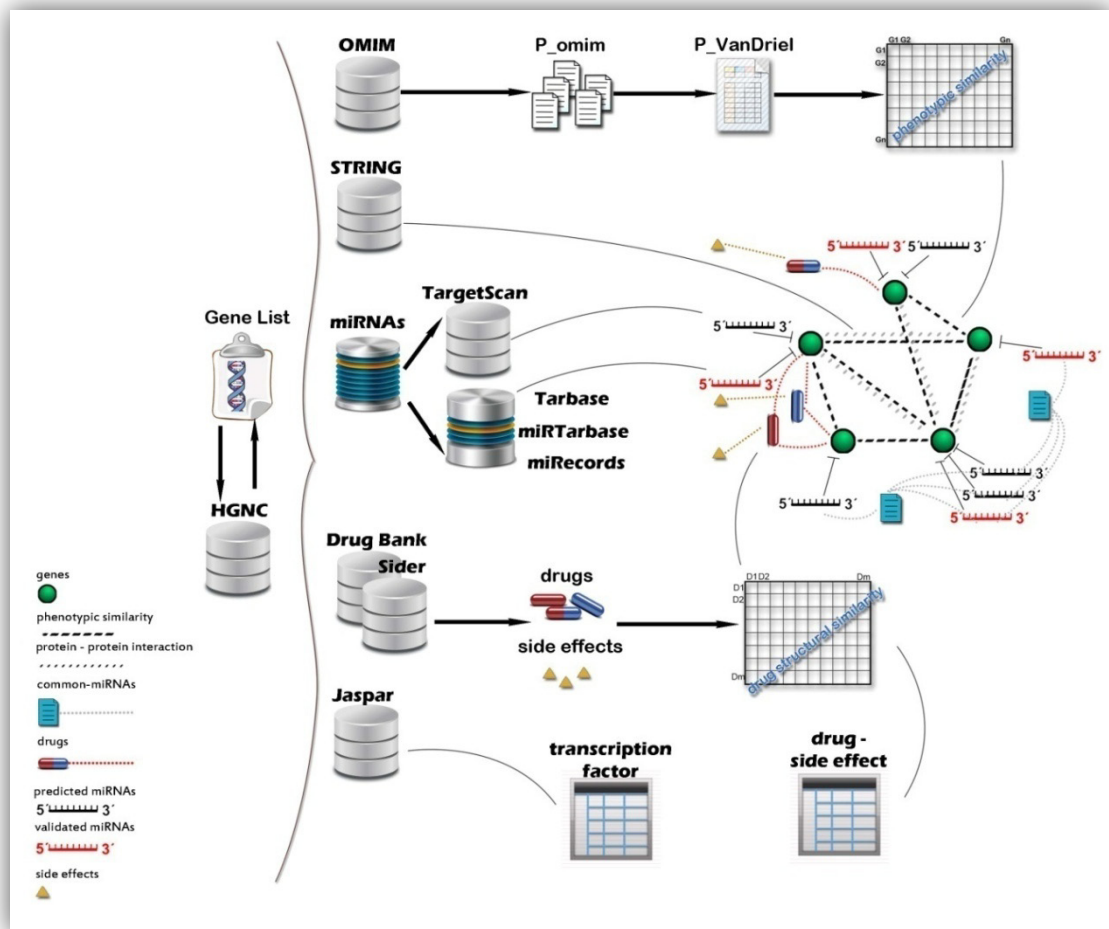


FIGURE 50: SCHEMA DEL FUNZIONAMENTO DI ESHMUN

La prima classe, l'**Input**, gestisce l'input dell'utente: il sistema richiede un minimo di due geni altrimenti il sistema ritorna un errore. Per ogni pathway genica

sottomessa è generato un job con un identificatore unico basato sulla data e l'orario di quando inviato. Ciò permette ad ESHMUN di creare un url unico per job che resterà attivo per 24 ore permettendo all'utente di recuperare i dati dopo l'esecuzione. Per ogni richiesta, ESHMUN interroga un database interno, creato in MySQL, per convertire l'input di differenti formati in un formato unisco usando la nomenclatura HGNC, infatti per ogni gene viene cercato il suo codice HGNC, e se trovato allora viene anche processato.

#	Nome	Tipo
1	hgnc	int(11)
2	approved_name	text
3	gene_name	text
4	synonyms	text
5	accession_number	text
6	entrez	text
7	ensembl	text
8	pubmed	text
9	refseq	text
10	omim	text
11	uniprot	text
12	drugbank	int(11)

FIGURE 51: TABELLA MYSQL PER LA CONVERSIONE INPUT

#	Nome	Tipo
1	hgnc	int(11)
2	synonyms	text

FIGURE 52: TABELLA MYSQL DEI CODICI HGNC E SINOMINI

La classe **Phenotype** interroga i dati prelevati dal database OMIM e la matrice di similarità tra fenotipi (similarità score matrix phenotypes SSMP) ottenuta da van Driel et al. In quest'ultimo caso noi creiamo una nuova SSMP attraverso i

database MeSH e OMIM, passando da una matrice di 5.132 x 5.132 ad una di 7.112 x 7.112 record. Ad ogni gene sono associati i fenotipi se c'è una documentazione correlata nella letteratura. In modo da calcolare la similarità fenotipica tra i geni (phenotypic similarity genes PSG) è usato uno script in R. Per ogni coppia di geni, A e B, sono presi i fenotipi da OMIM in cui sono correlati (definiamo questo set F) e per ognuno di essi sono presi i primi 100 fenotipi figli (indicati con f) più simili in accordo con la SSMP. Per entrambi i geni, A e B, unificiamo i due set di fenotipi e otteniamo $P_A = F_A \cup f_A$ e $P_B = F_B \cup f_B$, calcolando poi l'intersezione tra P_A e P_B , chiamata $IP_{A,B}$, similmente calcoliamo l'unione di P_A e P_B chiamata $UP_{A,B}$. La similarità fenotipica tra A e B è calcolato come segue:

$$S(A, B) = \frac{IP_{A,B}}{UP_{A,B}}$$

Quando tutti i PGS sono calcolati allora la network dei geni è creata (definiamo tale network geni-fenotipo, GPN), dove i nodi sono i geni mentre lo spessore di ogni arco rappresenta il PSG per una generica coppia di geni. In particolare, se non esiste un arco tra due geni allora il PSG di questi due geni è zero, in altre parole i geni non sono fenotipicamente simili. In questa sezione è inoltre calcolata la heatmap di associazione fenotipica. Infine ESHMUN calcola la significatività statistica per le componenti connesse di geni presenti nella network (connected component of genes CCG). Specificatamente, per ogni CCG noi aggiungiamo ogni PSG_i ($SCCG_i$), poi costruiamo un set di fenotipi collegati ad ogni gene presente in CCG, chiamato FCCG. Dopo di che, ESHMUN esegue una test di permutazione random in modo da calcolare un p-value per ogni CCG.

La classe **miRNA** permette di associare i miRNA predetti e validati per ogni gene in input. In particolare, attraverso queste informazioni ESHMUN genera due differenti network: 1) una GPN con miRNA predetti comuni, 2) una GPN con miRNA validati comuni.

La classe **Drug** per ogni gene prende i target delle drug e gli effetti collaterali associati ad ogni drug. Per ogni drug, se ha una smile, il database recupera le prime 10 drug più simili in accordo alla formula di Tanimoto.

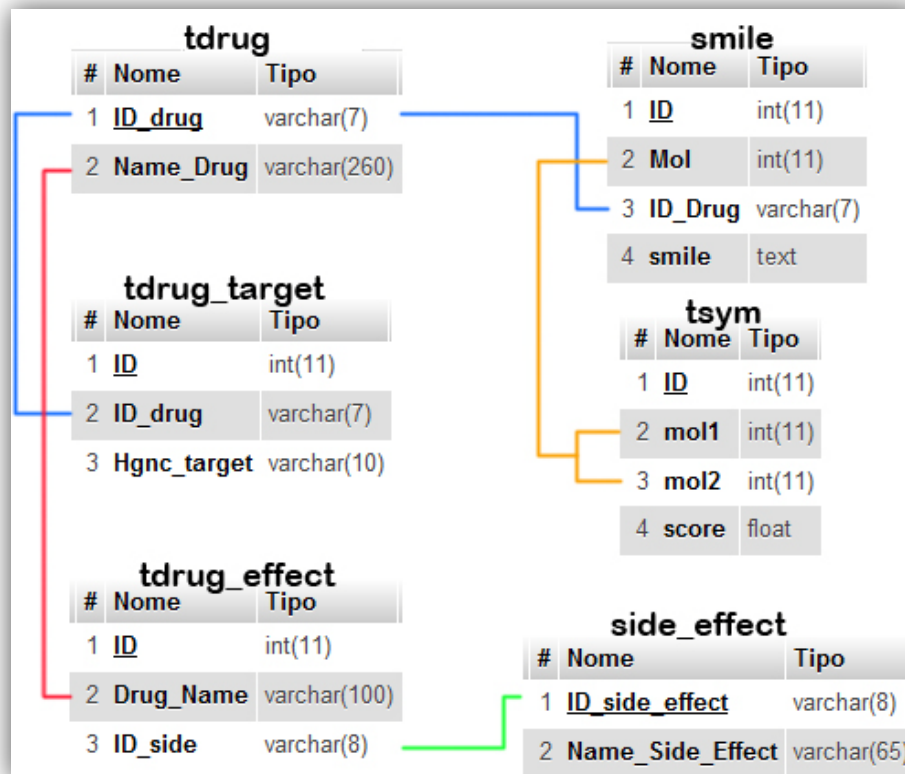


FIGURE 53: CONNESSIONE DELLE TABELLE MYSQL DI ESHMUN

La classe **PPI** proietta sugli stessi geni le interazioni proteina-proteina prelevate da STRING, in cui lo spessore degli archi rappresenta lo score dell'intersezione.

La classe **Transcription Factor** mostra tutti i fattori di trascrizione collegati ai geni delle pathway con informazioni sulle famiglie della classe.

Per rendere i risultati generati da ESHMUN più comprensibili essi sono visualizzati come network create tramite il tool Gephi, scritto in Java.

CAPITOLO 4:

ESHMUN - TSC

4. ESHMUN – TSC

4.1. TSC1/TSC2

Il complesso della sclerosi tuberosa (Tuberous Sclerosis Complex, TSC) è una condizione genetica a trasmissione autosomica dominante che interessa vari tessuti e organi. Il termine TSC è stato coniato per la prima volta dal patologo Molten nel 1942, ponendo l'accento sul coinvolgimento multi-sistemico e sull'estrema variabilità delle manifestazioni cliniche. I progressi clinici e genetici compiuti hanno iniziato a far luce su questioni centrali per la diagnosi, prevenzione e cura della sclerosi tuberosa, prima fra tutte il perché dell'enorme variabilità clinica. Dopo la scoperta dei due geni responsabili, TSC1 e TSC2 [170] [171], localizzati rispettivamente nel cromosoma 9 e 16, e delle funzioni principali delle proteine codificate da questi geni, amartina (gene TSC1) e tuberina (gene TSC2), sono stati introdotti nuovi criteri diagnostici ed è mutata profondamente la conoscenza della storia naturale e della biologia di quest'affezione. I segni clinici principali sono nella cute ed alcuni elementi connessi (unghie e denti), del sistema nervoso centrale, quindi cuore, reni, occhio, fegato e polmone [172], ma si possono avere manifestazioni della malattia in quasi tutti gli organi ed apparati (quali tessuto muscolare scheletrico e sistema nervoso periferico).

La TSC ad oggi è considerata una condizione rara e quasi inevitabilmente associata a gravi complicazioni neurologiche (come epilessia), cognitive (ritardo mentale) e disturbi del comportamento oltre alle tipiche manifestazioni cutanee.

L'incidenza della sclerosi tuberosa è circa di 1 su 6.000 individui nella popolazione generale, tuttavia la reale incidenza però non è perfettamente conosciuta a causa dell'elevato numero di persone affette con segni clinici molto lievi o quasi del tutto asintomatiche che, non richiedendo l'intervento del medico, sfuggono alle diagnosi. La prevalenza è quindi 1 su 13.000 – 30.000 individui nella popolazione generale e di 1 su 6.800 nella popolazione di bambini d'età compresa tra 11 e 15 anni.

Il gene TSC1 (Tuberous Sclerosis Complex gene 1): è localizzato sul cromosoma 9 (regione q34) ed è composto di 23 esoni distribuiti su 55 kb di DNA. I primi due esoni non sono codificanti; la regione codificante, cioè contenente l'informazione per la traduzione

dell'amartina inizia a metà del terzo esone. L'RNA messaggero ha una lunga coda non codificante, quindi più della metà dell'RNA messaggero maturo di TSC1 è costituito da sequenze non codificanti, in testa e in coda al gene. Le regioni trascritte e non tradotte sono spesso implicate nella regolazione dell'espressione genica; al momento non è ancora chiaro per i geni TSC quali siano le sequenze all'interno delle regioni trascritte e non tradotte coinvolte in questa regolazione.

Il gene TSC2 (Tuberous Sclerosis Complex gene 2): è localizzata sul cromosoma 16 (regione p13.3), è composto di 41 esoni codificati ed uno non codificante, ma trascritto. Nell'insieme i 42 esoni abbracciano circa 40 kb di DNA.

Fino ad ora nessun individuo con ST è risultato portatore di una mutazione sia del gene TSC1 che del gene TSC2, quindi il difetto di uno solo di essi è sufficiente a causare la ST. Inoltre non esistono prove a favore dell'esistenza di altri geni TSC, nel senso che non sono state osservate ad oggi famiglie in cui la malattia non segregasse né con TSC1 né con TSC2.

Le persone con ST in genere presentano in tutte le loro cellule un difetto di un allele (paterno o materno) in uno dei due geni TSC, e quindi potranno trasmettere alla prole o l'allele difettoso o l'allele normale, con la probabilità del 50%. Poiché né TSC1 né TSC2 sono localizzati sui cromosomi sessuali, la malattia si trasmette con egual probabilità sia ai maschi che alle femmine. Questo tipo di trasmissione viene detta autosomica dominante: autosomica, perché non è dovuta ad un gene sul cromosoma X o Y; dominante, perché è sufficiente avere un solo allele difettoso, perché la malattia prima o poi si manifesta. Si è infatti visto che soggetti eterozigoti, che hanno ereditato un allele difettoso da un genitore, ed un allele sano dell'altro genitore, hanno una probabilità vicina al 100% di manifestare nell'arco della loro vita un qualche segno della malattia. Sebbene i portatori di mutazione abbiano un'altissima probabilità di manifestare qualche segno di ST, questi segni possono variare di molto, causano forme molto lievi o forme molto gravi di malattia, a seconda del numero di lesioni e varietà di organi coinvolti. Quindi nella ST, l'alta penetrazione si associa ad una grande variabilità di espressione clinica. Questa variabilità d'espressione si verifica sia tra famiglie diverse, sia all'interno della stessa famiglia.

4.2. Funzioni principali delle proteina amartina e tuberina

La proteina prodotta del gene TSC1 (amartina) è costituita da 1.164 aminoacidi e non ha alcuna omologia con la proteina prodotta dal gene TSC2. La proteina prodotta dal gene TSC2 (tuberina) è costituita da 1807 aminoacidi. La tuberina ha un dominio GAP (GTPase Activating Protein) grazie al quale accelera il rilascio spontaneo di GTP da parte di RHEB

(Ras Homologue Enriched in Brain, una piccola G-binding protein della famiglia Ras, espressa ad alti livelli nel cervello), facilitando la sua transizione allo stato inattivo.

Poiché RHEB legata al GTP è un attivatore di mTOR (mammalian Target of Rapamycin) del complesso TORC1 (TOR Complex 1), la tuberina assume un ruolo importante in quanto regolatore della trasmissione di svariati segnali mediati da TORC1 e correlati alla crescita cellulare, come la sintesi proteica, la risposta all'ipossia e la rilevazione dei bassi livelli d'energia (ATP).

Le due proteine, tuberina ed amartina [173], sono presenti nelle cellule di tutti i tessuti ed interagiscono con un'elevata affinità formando eterodimeri (il legame dell'amartina con la tuberina stabilizza quest'ultima, impedendone la degradazione). Ciò spiega perché soggetti con mutazioni in TSC1 o mutazioni in TSC2 presentino manifestazioni cliniche. Non si può tuttavia escludere che una delle due proteine, o entrambe, svolgono anche attività autonome. Tra i ruoli aggiuntivi dell'amartina ci sono le capacità di legare proteine regolatrici del citoscheletro e di influire sul ciclo cellulare in quanto fosforilata da chinasi ciclina-dipendenti. L'eterodimero amartina/tuberina tra le sue tante funzioni interviene nella regolazione intracellulare di alcune vie di segnali: 1) Crescita cellulare e sintesi proteica (cosiddetta via di PI3K/AKT/mTOR), 2) adesione/migrazione/trasporto e traffico proteico cellulare (cosiddetta via della Glicogeno Sintasi Kinasi 3 (GSK3)/ β -catenina/chinasi dell'adesione focale (FAK)/Ras omologo (Rho)), 3) crescita e proliferazione cellulare (cosiddetta via della proteina chinasi mitogeno attivata (AMPK)). Il ruolo del complesso amartina/tuberina nella regolazione della crescita e della proliferazione cellulare ben si correla con la principale manifestazione della malattia, cioè la formazione di amartomi.

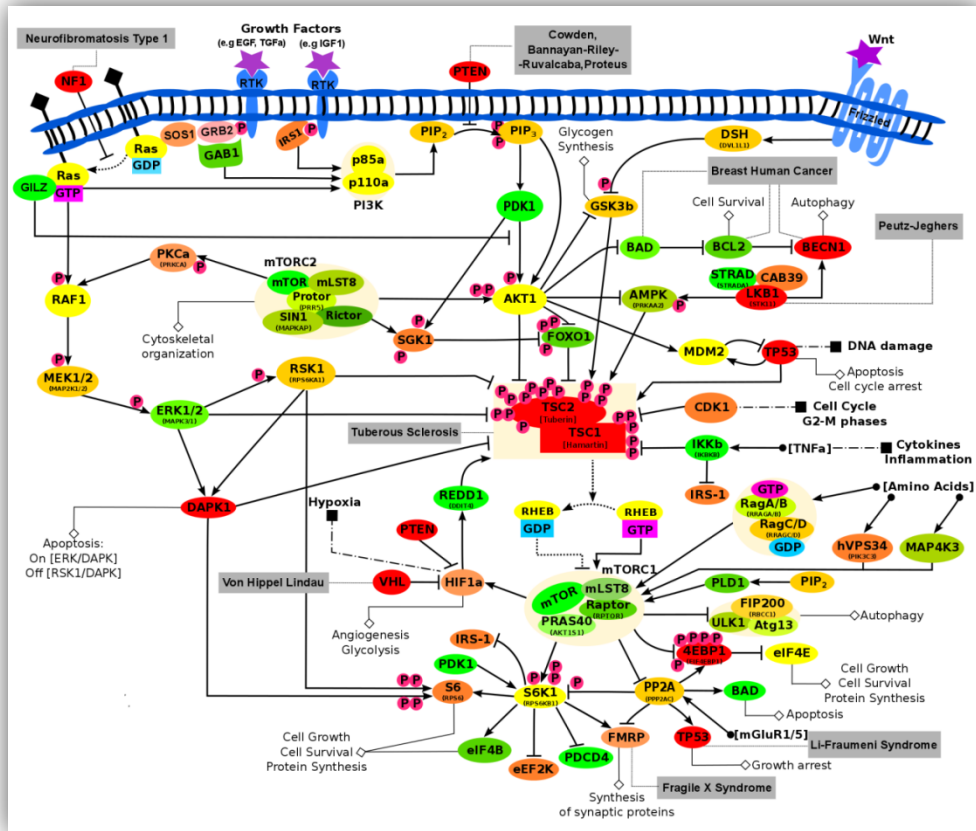


FIGURE 54: PATHWAY DELLA TSC [174]

4.3. Difetti genici, mutazioni e quadri clinici

Le analisi delle mutazioni nei soggetti con ST concordano nel segnalare una maggiore frequenza di mutazioni TSC2 (80%) rispetto a TSC1 (20%). Inoltre i difetti di TSC2 sono significativamente più rappresentati tra i casi sporadici, mentre i difetti di TSC1 si trovano più spesso in famiglie con più parenti affetti. Ciò suppone che le mutazioni TSC2 siano associate ad un quadro clinico più grave rispetto alle TSC e che quindi le persone con ST che hanno un difetto del gene TSC1 possano più frequentemente sfuggire alla diagnosi così da spiegare almeno in parte la loro apparente rarità. Questa spiegazione è avvalorata da una maggiore frequenza di spasmi infantili, ritardo di sviluppo psicomotorio e difficoltà di apprendimento tra i soggetti sporadici con una mutazione del gene TSC2 rispetto a soggetti sporadici con un difetto del gene TSC1.

Le ricerche e lo sviluppo della conoscenza su questa malattia hanno condotto a formare un set di criteri per la TSC, stabiliti dalla Tuberous Sclerosis Complex Consensus Conference:

Diagnosi definitiva: 2 segni maggiori o 1 segno maggiore + 2 segni minori

Diagnosi probabile: 1 segno maggiore + 1 segno minore

Diagnosi possibile: 1 segno maggiore o 2 o più segni minori

Segni maggiori:

- Angiofibromi facciali o placca fibrosa "frontale"
- Fibromi ungueali o periungueali non-traumatici
- Macchie ipomellanotiche (>3)
- Macchia zigrinata
- Amartomi retinici modulari multipli
- Tuberi corticali
- Astrocitoma subependimale e cellule giganti (ASCG)
- Rabdomiomi cardiaci, singoli o multipli
- Linfangiomatosi polmonare
- Angiomiolipomi renali

Segni minori:

- Lesioni multiple dello smalto dentario a distribuzione casuale
- Polipi rettali amartomatosi
- Cisti ossee
- Linee radiali di migrazione della sostanza bianca
- Fibromi gengivali
- Amartomi non-renali
- Chiazze acromiche retiniche
- Lesioni ipomelanotiche cutanee "a confetto"
- Cisti renali multiple

4.4. Fenotipo Sclerosi Tuberosa

I segni clinici della ST [175] possono essere a carico di molti organi ed apparati e possono manifestarsi durante periodi della vita. Vi possono essere persone con poche manifestazioni cliniche che tuttavia sono portatori di mutazioni in uno dei due geni della ST e famiglie all'interno delle quali l'espressione della malattia è assai variabile. La diagnosi e l'assistenza varia a seconda dell'età dell'individuo con sospetta ST o affetto da ST.

4.4.1. Cute

I segni cutanei della ST sono molteplici, abitualmente compaiono con un preciso ordine cronologico. Sebbene le manifestazioni cutanee della ST siano tra i segni più frequenti dopo la seconda decade di vita. L'esame dermatologico resta lo strumento più rapido, economico e di maggior efficacia diagnostica, se condotto da persone esperte in questa malattia.

- **Macchie ipomelanotiche:** Sono i segni cutanei più precoci. Sono presente alla nascita solo nel 30% circa dei casi, ma possono apparire o aumentare di numero nell'infanzia o in età puberale. Il numero di lesioni varia da <2 a diverse decine o centinaia, con forme e dimensioni variabili, rotondeggianti, lanceolate, molto piccole e multiple. Circa il 5% degli individui sani nella popolazione generale possono avere una o due macchie ipomelanotiche, è quindi il numero(>3) che conferisce a questo segno una rilevanza diagnostica.
- **Placca fibrosa:** E' una lesione di consistenza elastica e di colore (marrone-rossastro o giallo) un po' più scuro rispetto alla cute circostante posta generalmente nella regione frontale o sotto il cuoio capelluto.
- **Placca zigrinata:** Si intende un'area di cute dall'aspetto rugoso e di colore rosa-giallastro che compare di solito nella tarda infanzia. Inizialmente si manifesta con un'insieme di piccole papule leggermente rilevate, che confluiscono in un'area più estesa con il tipico aspetto zigrinato, quindi ruvido.
- **Angiofibroma facciale:** Istologicamente queste lesioni sono "amartomi" del tessuto connettivo e dei vasi del derma che si manifestano clinicamente come piccole papule sessili tondeggianti, rossastre o rosa, distribuite in guance, mento e punta del naso. Compaiono dai 3-5 anni di età sino alla pubertà quando poi tendono ad aumentare di numero e dimensione.
- **Fibromi cutanei:** Si possono avere dei piccoli fibromi penduli cutanei localizzati in prevalenza sul collo e sulle spalle.
- **Fibromi ungueali:** Appaiono in età adulta ma non in tutti i soggetti affetti da ST. Sono più comuni nelle dita dei piedi che delle mani. Sono di due varietà: 1) periunguali, cioè fibromi rotondi, ovali o piatti. 2) sub ungueali, che si annidano presso la matrice dell'unghia.

4.4.2. Occhio

Le lesioni oculari nella ST raramente sono sintomatiche e non sono progressive, nei pazienti possono essere presenti amartomi retinici che sono di tre varietà: 1) **piatti e lisci**, di forma rotonda o ovale e di colore rosa salmone. Sono la varietà più frequente ed appaiono come lesioni piatte a superficie liscia, traslucide e spesso poco circoscritte o rilevate. 2) **elevati, multinodulari e calcifici**, sono meno frequenti e sono composti da cluster di noduli opachi, cisti o calcificati, sono localizzati nella parte marginale della papilla ottica o al centro della retina, 3) **transizione o misto**, con caratteristiche intermedie tra le due varietà precedenti. Le lesioni retiniche sono più frequentemente rilevabili con l'aumento dell'età durante l'infanzia ma possono essere riscontrate a qualsiasi età nel neonato.

4.4.3. Lesioni Orali

- **Lesioni dentali:** Sono ipoplasie focali dello smalto dentale visibili sia in bambini che in adulti con ST. Sono asintomatiche e non interferiscono con le funzioni biologiche del dente. Possono presentarsi in forme e dimensioni diverse: piccole tacche o fossette rotonde, 2) lesioni microscopiche.
- **Lesioni della mucosa buccale:** In più del 40% dei soggetti con ST si hanno fibromi mucosi e/o gengivali, clinicamente hanno un aspetto di noduli di dimensioni variabili e di colore variabile dal rosso-giallastro al rosa, simile alle mucose circostanti. Si osservano principalmente nella gengiva superiore, labbra, guance e palato.

4.4.4. Cuore

Il 45-70% circa dei soggetti con ST presenta, eco-graficamente, rabdomiomi cardiaci. I rabdomiomi nella ST sono multipli, localizzati all'interno delle cavità o nelle parti cardiache. Istologicamente sono costituiti da cellule plurinucleate, molto più grandi delle fibrocellule muscolari cardiache e di colore giallastro o bianco per l'abbondante accumulo di glicogeno nel citoplasma. I rabdomiomi della ST sono abitualmente asintomatici e quindi non richiedono un trattamento specifico. In una piccola percentuale di casi (<15%), a causa della grande dimensione o localizzazione in sedi critiche, generano problemi in epoca fetale, neonatale o eccezionalmente in età successive. Si tratta, in ordine di frequenza, di 1) disturbi del ritmo cardiaco per danneggiamento delle vie di conduzione cardiache, 2) alterazione del flusso cardiaco per interferenza sulla cinesi

valvolare o parziale/totale ostruzione delle cavità così da condurre ad insufficienza cardiaca, 3) deterioramento delle fibrocellule muscolari per parziale sostituzione con tessuto patologico non contrattile e comparsa di segni/sintomi sovrapponibili a quelli di una miocardiopatia, 4) in epoca fetale, aritmia cardiaca con esito in idrope e morte fetale. Dal punto di vista diagnostico i rhabdomiomi rappresentano una delle lesioni più importanti perché, se presenti, sono rilevabili già alla nascita e talvolta anche in epoca prenatale.

4.4.5. Sistema Vascolare

Alcuni pazienti con ST possono avere aneurismi arteriosi, e sono spesso associati con una displasia fibromuscolare dell'arteria renale. Sembra che questo fenomeno sia causato da ampie delezioni che coinvolgo sia il gene TSC2 che il gene adiacente, PDK1. PDK1 è associato con la sindrome del rene policistico dell'adulto è quindi associato all'origine di alcuni sintomi della ST, tra cui aneurismi cerebrali, cisti epatiche multiple e rene policistico infantile.

4.4.6. Rene

Le tipiche lesioni renali associate con la TS sono gli angiomiolipomi (AML) multipli e bilaterali. Tali AML sono i responsabili delle complicazioni acute e croniche riguardanti la funzione renale e la prima causa di gravi complicazioni legate alla malattia nei soggetti adulti con ST.

- **Angiomiolipomi renali:** è un tumore benigno composto da tessuto adiposo, vasi anomali e tessuto muscolare liscio localizzato nella regione corticale del rene. Molto raramente si giunge ad insufficienza renale.
- **Cisti renali:** Le cisti sono di due tipi: cisti piccole, che si manifestano come cisti isolate o di piccoli gruppi, e cisti grandi che sono numerose e bilaterali.
- **Cisti renale linfagiomatosa:** è una cisti molto rara nella ST, circa <1%. Istologicamente ha componenti endoteliali, possono causare ipertensione e/o insufficienza renale.

4.4.7. Sistema Nervoso Centrale

Le lesioni del sistema nervoso centrale (SNC) nella ST sono principalmente delle malformazioni dello sviluppo corticale, ovvero sono disturbi della migrazione, proliferazione e differenziazione cellulare. Le tipiche lesioni del SNC sono:

- **Tuberi corticali:** appaiono rotondeggianti o di forma ovale, localizzati nella sostanza grigia corticale o sottocorticale
- **Noduli subependimali:** sono piccoli nodi rilevati presenti nelle pareti dei ventricoli cerebrali, di aspetto più compatto dei tuber corticali. Sono costituiti da cellule giganti e irregolari.
- **Astrocitoma subependimale a cellule giganti (ASCG):** presenta le stesse caratteristiche dei noduli subependimali. I sintomi sono comunemente causati dall'ostruzione della circolazione del liquor cerebrale e poiché al crescita dell'ASCG è molto lenta questo tipo di lesione può passare inosservata anche per anni.

4.4.7.1. Manifestazioni neurologiche e psichiatriche

- **Epilessia:** è la manifestazione più frequente nella ST. Circa l'80% dei soggetti affetti presenta crisi epilettiche che possono manifestarsi entro i primi anni di vita o in età adulta. L'epilessia nella ST ha le caratteristiche dell'epilessia di tipo parziale, con convulsioni parziali seguite da spasmi o da sincronia bilaterale secondaria. La storia naturale dell'epilessia nella ST è alquanto complessa. Spesso l'evoluzione è verso una spontanea remissione delle crisi avvicinandosi ed entrando nell'età adolescenziale. Una parte dei bambini, però, andrà incontro ad un aumento di frequenza e di gravità della crisi, a farmaco-resistenza, con evidente scadimento della qualità della vita.
- **Disturbi cognitivi:** nello sviluppo cognitivo nella ST può essere del tutto normale oppure si possono avere deficit cognitivi di frequenza e grado variabile associati o meno a disturbi dell'apprendimento e del comportamento. È ampiamente dimostrato che tutti i pazienti con handicap mentale hanno presentato o presentano epilessia, mentre il 30% circa di quelli con epilessia ha intelligenza normale.
- **Disturbi comportamentali:** sono molto comuni i deficit dell'attenzione con iperattività, comportamenti di tipo aggressivo con fenomeni di automutilazione o disturbi del sonno.
- **Disturbi del sonno:** sono abbastanza frequenti, specie nei soggetti con epilessia e disturbi cognitivi e sono caratterizzati da frequenti risvegli durante il sonno, risvegli

precoci, disturbi del sonno correlati alla presenza, frequenza e tipo di epilessia, eccesso di sonnolenza diurna, sonnambulismo.

4.4.8. Apparato Respiratorio

Il coinvolgimento polmonare nella ST è molto raro e da cinque a dieci volte più frequente nelle donne è rappresentato dalla lesione nota come linfagiomiomatosi (LAM) polmonare. La LAM è caratterizzata dalla presenza di lesioni cistiche al polmone, alterazioni dei vasi linfatici associati ad angiomiolipomi renali. Esistono due forme di LAM: i) la forma "isolata" ovvero non associata alla ST o ad altra affezione, si manifesta in donne in età fertile compresa tra i 20 e i 40 anni, ii) la forma "associata alla ST": una elevata percentuale di donne con ST può sviluppare un quadro polmonare di LAM.

4.5. Terapia

I principali farmaci usati per l'inibizione di mTOR sono la Rapamicina e il Vigabatrin, ma oltre a questi molti altri sono in fase di sperimentazione.

4.5.1. Rapamicina

La Rapamicina è un farmaco immunosoppressore usato per prevenire il rigetto nei trapianti d'organo. Recentemente è stato scoperto anche che i suoi effetti antiproliferativi potrebbero avere un ruolo importante nella cura del cancro. È presente in DrugBank sotto forma di tre composti; il Sirolimus, l'Everolimus e il Temsirolimus. Approfondendo lo studio sui target si nota che mTOR è un target di tutte e tre le drug. Utilizzando il database KEGG si può studiare la signaling Pathway dell'mTOR, essa è collegata a tre pathway riguardanti i processi cellulari, che sono la MAPK signaling Pathway, l'Apoptosi signaling Pathway e l'Insulin signaling Pathway. Inoltre è collegato ad una Pathway chiamata Regulation of autophagy, che descrive un processo molto importante nella crescita cellulare, volto a mantenere un equilibrio tra la sintesi, la degradazione e il successivo riciclaggio dei prodotti cellulari. L'ultima importante Pathway collegata ad mTOR è VEGF signaling Pathway, che descrive il comportamento del fattore di crescita

VEGF. L'importanza di quest'ultima Pathway è enorme in quanto la sovrapproduzione di VEGF è il primo passo delle formazioni di neoplasie e consente alle cellule tumorali un approvvigionamento costante di nutrienti necessari al soddisfacimento del loro metabolismo accelerato. Di seguito è riportato la Signaling Pathway dell'mTOR, dove si possono notare i vari collegamenti.

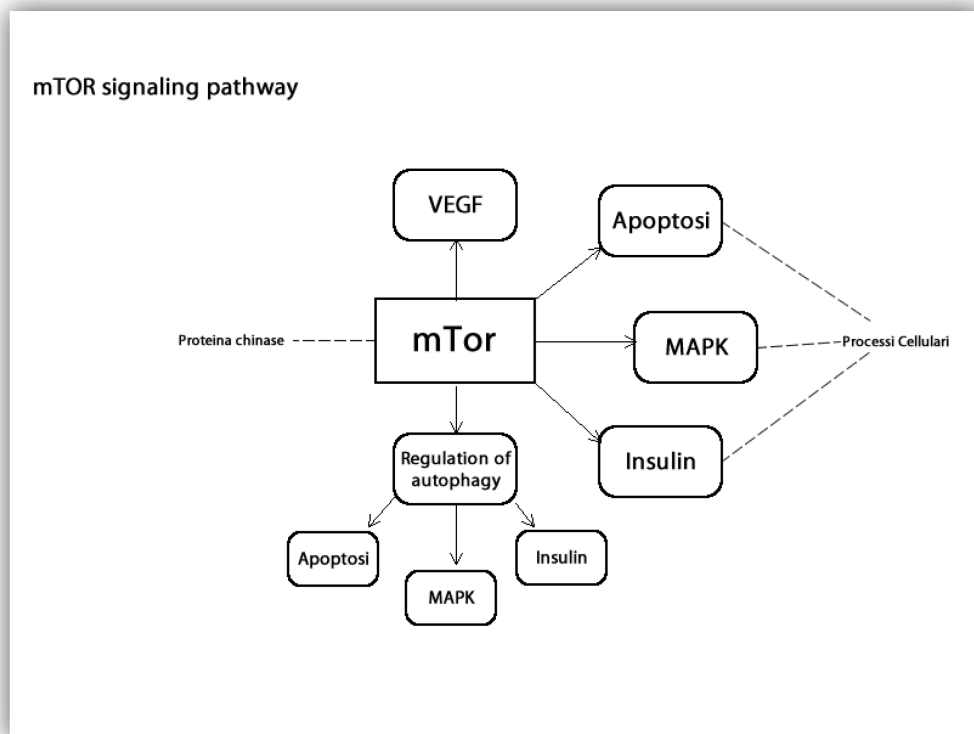


FIGURE 55: MTOR SIGNALING PATHWAY

Le Pathway riguardanti i processi cellulari, MAPK, l'Apoptosi e l'Insulin, a loro volta sono collegati ad altre Pathway. Studiando tale collegamenti si arriva ad altre Pathway che descrivono vari tumori, strettamente interconnesse con altre che descrivono invece processi cellulari. Tra questi troviamo la PI3K-AKT signaling Pathway, che attraverso l'attivazione di AKT può controllare i processi della sintesi delle proteine, del metabolismo, del ciclo cellulare è strettamente collegato all'Apoptosi, responsabile di una forma di morte cellulare programmata, che contribuisce al mantenimento del numero di cellule in un sistema.

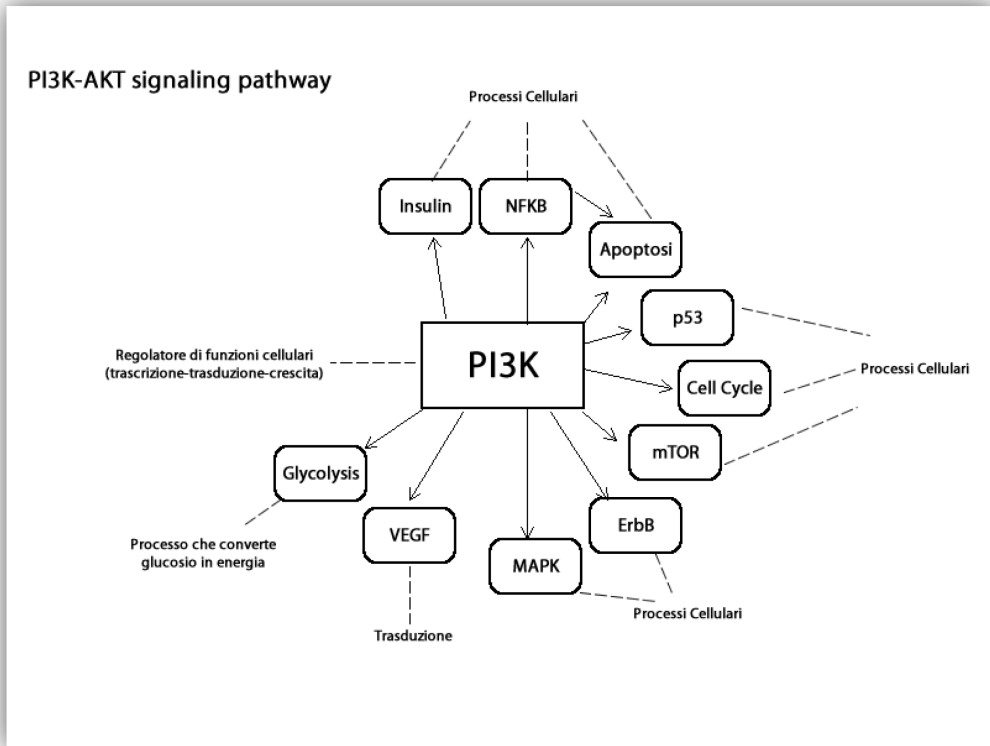


FIGURE 56: PI3K-AKT SIGNALING PATHWAY

In fine la p53 signaling Pathway, regolatore di geni e causa di ST, Atassia, malattie polmonari, emorragie, difetti cardiaci e tumori, ed anche quest'ultima collegata ad Apoptosi, al ciclo cellulare e all'Insulin signaling Pathway.

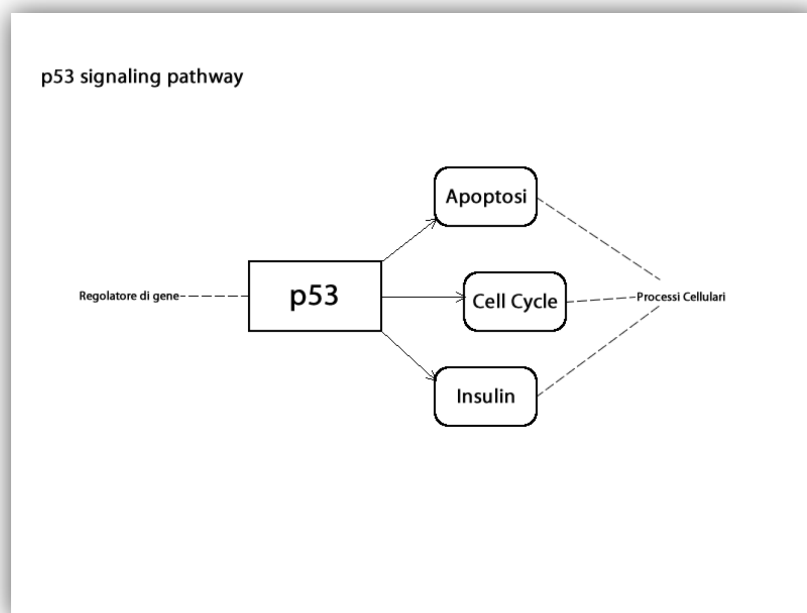


FIGURE 57: P53 SIGNALING PATHWAY

L'eterodimero 1 amartina/tuberina interviene nella regolazione intracellulare di alcune vie di segnale, come la crescita cellulare e sintesi proteica, la cosiddetta via di PI3K-AKT-mTOR e la crescita e proliferazione cellulare, la cosiddetta via della proteina chinasi mitogeno attivata MAPK. Quindi il ruolo del complesso amartina/tuberina nella regolazione della crescita e proliferazione cellulare e dei processi cellulari in generale, ben si correla con la principale manifestazione della malattia, cioè la formazione di amartomi². Nel Database DrugBank è inoltre presente un Tool che permette la ricerca di drug, già approvate o anche in fase di sperimentazione, che abbiano una struttura chimica simile alla drug oggetto di studi. Un confronto simile ha un'importanza notevole perché permette di studiare tutte le varianti della stessa drug, approfondendone i target e quindi i possibili effetti collaterali o eventuali effetti benefici inaspettati. Sono presenti altri due composti, simili ai primi tre, chiamati Pimecrolimus e Tracolimus. Il successivo approfondimento dei Targets e dei Pathway relativi ha evidenziato una somiglianza pressoché identica con il Sirolimus, il Temsirolimus e l'Everolimus.

4.5.2. Vigabatrin

Il Vigabatrin è un farmaco anticonvulsivante³ e, a differenza della Rapamicina, è presente in DrugBank sotto forma di un unico composto. Il target primario indicato da Drugbank è l'inibizione del catabolismo del GABA, principale neurotrasmettitore inibitorio, nei mammiferi, del sistema nervoso centrale, responsabile della regolazione dell'eccitabilità neuronale ed anche direttamente responsabile della regolazione del tono muscolare. Il Vigabatrin agisce in modo irreversibile sulla GABA-transaminasi, l'enzima responsabile della degradazione del GABA. Studiando questo aspetto tramite KEGG Pathway si nota che GABA si lega a tre classi di recettori, il GABA-A, il GABA-B e il GABA-C.

Nell'immagine sotto è riportato la Pathway di GABAERGIC SYNAPSE:

¹ Collegamento chimico differente, contrario di omodimero cioè collegamento di due molecole di natura chimica identica.

² Malformazione dovuta a un'anormale mescolanza di elementi costitutivi normali del tessuto.

³ Farmaco usato allo scopo di sedare le convulsioni.

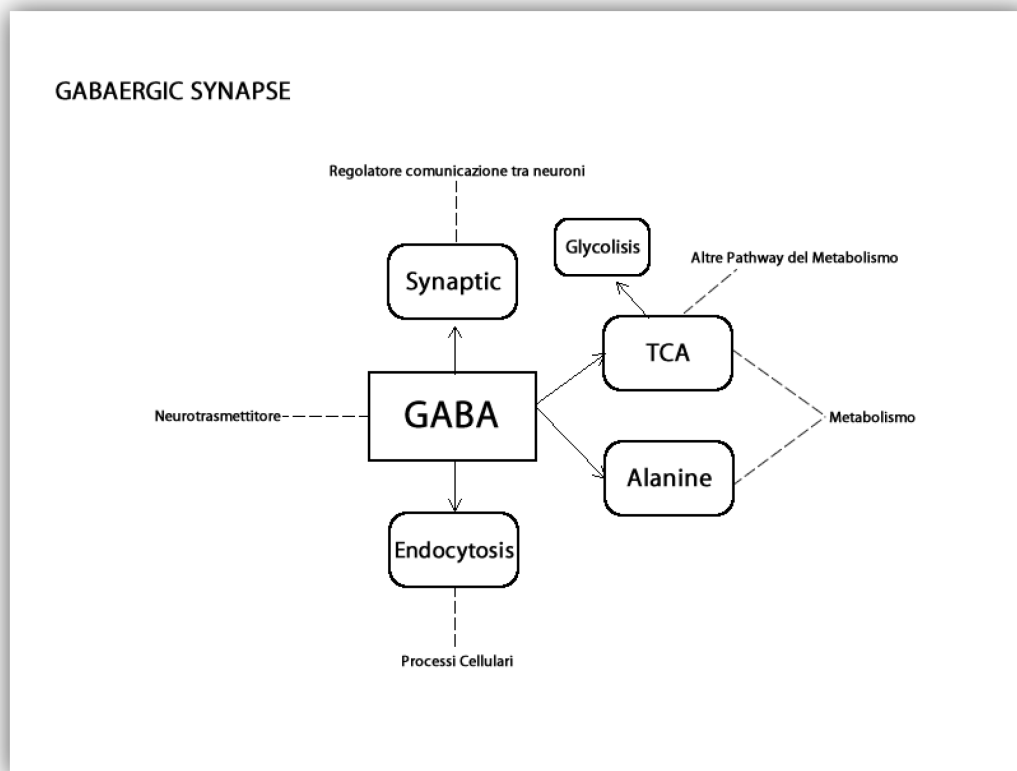


FIGURE 58: GABAERGIC SYNAPSE

Il GABA è collegato ad altre quattro Pathway: Cytrate Cycle (o TCA cycle), l'Alanine aspartate and glutamate metabolism, il Synaptic vesicle cycle e l'Endocytosis. Il TCA è una importante pathway per la parte finale dell'ossidazione dei carboidrati e degli acidi grassi. Come evidenziato dalla figura è a sua volta collegato alla Pathway Glycolysis e ad altre Pathway metaboliche. L'alanine aspartate and glutamate metabolism descrive un complesso metabolico di gruppi amminici prodotti dal muscolo, da qui il collegamento con GABA. Il Synaptic vesicle cycle è un regolatore per la comunicazione tra neuroni, mentre l'Endocytosis descrive un processo cellulare per il trasporto il catabolismo. Il secondo target del Vigabatrin indicato da Drugbank interagisce con molte Pathway riguardanti processi metabolici di amminoacidi e di carboidrati, quali: Valine, Beta-Alanine, Propanoate e Butanoate metabolism. Anche nel caso del Vigabatrin utilizzando il Tool di DrugBank per la ricerca di drug con strutture chimiche simili, emerge che quasi tutte sono in fase di sperimentazione e si riscontra in esse l'interazione con la Pathway Argine and proline metabolism, che descrive processi riguardanti il metabolismo degli amminoacidi. L'unica drug approvata è la L-Glutamic Acid, l'acido glutammico è un amminoacido polare, che l'organismo è in grado di sintetizzare, di importanza critica per il

funzionamento delle cellule che giungendo al cervello viene usato per la sintesi delle proteine. Oltre al suo ruolo di costituente delle proteine, è anche un neurotrasmettitore eccitatore del sistema nervoso centrale ed un precursore di GABA. Ciò che colpisce dell'L-Glutamic Acid è che, secondo DrugBank, ha 70 target. Un così alto numero di obiettivi presenti per questo farmaco porta inevitabilmente alla presenza di molteplici interazioni ed le Pathway chiamate in causa sono decine. In casi come questi quindi, uno strumento come KEGG facilita notevolmente ad approfondire ogni singolo target delle drug e vedere in quali Pathway esso va ad interagire, analizzando poi i singoli casi e mettendo in luce tutte le possibili interazioni.

4.6. **ESHMUN e caso studio**

ESHMUN è un web tool il cui frontend è stato realizzato in Html, PHP, CSS e JQuery. L'homepage è divisa in 3 sezioni:

- Upload di un file contenente una lista di geni, i formati accettati sono txt e csv

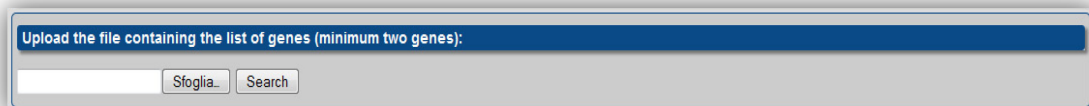


FIGURE 59: UPLOAD FILE DI LISTA GENICA

- Inserimento in una textarea di una lista di geni separate da virgola ','

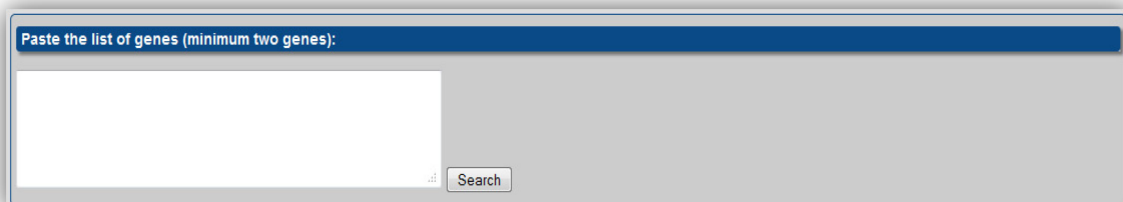


FIGURE 60: TEXTAREA PER L'INSERIMENTO DELLA LISTA DEI GENI

Il sistema riconosce i seguenti formati:

- o HGNC
- o RefSeq
- o Gene name

- o Approved name
 - o Accession number
 - o Id Ensemble
 - o Id PubMed
 - o Id Uniprot
- Selezionare una pathway prelevata da KEGG

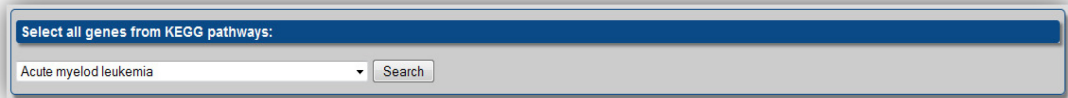


FIGURE 61: SELECT PATHWAY DI KEGG

Divise in 6 categoria, quali:

- 1) Cancer
- 2) Cardiovascular Disease
- 3) Endocrine and Metabolic Disease
- 4) Immune Disease
- 5) Infectus Disease
- 6) Neurodegenerative Disease

Consideriamo i seguenti geni relativi al complesso della sclerosi tuberosa estratti dalla letteratura, usiamo la seconda modalità di input inserendo nella textarea la colonna HGNC dei 68 geni.

Gene No.	Gene Name	HGNC	Gene Name HGNC	Approved Name
1	RAS	9871	RASA1	RAS p21 protein activator (GTPase activating protein) 1)
2	SOS1	11187		(son of sevenless homolog 1 (Drosophila))
3	RAF1	9829		(v-raf-1 murine leukemia viral oncogene homolog 1)
4	GILZ	3051		TSC22 domain family, member 3
5	MEK1	6840	MAP2K1	(mitogen-activated protein kinase kinase 1)
6	MEK2	6842	MAP2K2	(mitogen-activated protein kinase kinase 2)
7	ERK1	6877	MAPK3	(mitogen-activated protein kinase 3)
8	ERK2	6871	MAPK1	(mitogen-activated protein kinase 1)

9	DAPK	2674	DAPK1	(death-associated protein kinase 1)
10	PKCa	9393	PRKCA	(protein kinase C, alpha)
11	GRB2	4566		(growth factor receptor-bound protein 2)
12	NF1	7765		(neurofibromin 1)
13	GAB1	4066		(GRB2-associated binding protein 1)
14	IRS1	6125		(insulin receptor substrate 1)
15	PI3KCA	8975		(phosphoinositide-3-kinase, catalytic, alpha polypeptide)
16	PI3KR1	8979		(phosphoinositide-3-kinase, regulatory subunit 1 (alpha))
17	RSK1	10430	RPS6KA1	(ribosomal protein S6 kinase, 90kDa, polypeptide 1)
18	PIP	8993		(prolactin-induced protein)
19	SGK1	10810		(serum/glucocorticoid regulated kinase 1)
20	PDK1	8809		(pyruvate dehydrogenase kinase, isozyme 1)
21	AKT1	391		(v-akt murine thymoma viral oncogene homolog 1)
22	mTOR	3942		(mechanistic target of rapamycin (serine/threonine kinase))
23	MLST8	24825		(MTOR associated protein, LST8 homolog (S. cerevisiae))
24	Protor	31682	PRR5	(proline rich 5 (renal))
25	SIN1	18752	MAPKAP1	(mitogen-activated protein kinase associated protein 1)
26	RICTOR	28611		(RPTOR independent companion of MTOR, complex 2)
27	FOXO1	3819		(forkhead box O1)
28	TSC1	12362		(tuberous sclerosis 1)
29	TSC2	12363		(tuberous sclerosis 2)
30	RHEB	10011		(Ras homolog enriched in brain)
31	PRAS40	28426	AKT1S1	(AKT1 substrate 1 (proline-rich))
32	RAPTOR	30287	RPTOR	(regulatory associated protein of MTOR, complex 1)
33	VHL	12687		(von Hippel-Lindau tumor suppressor, E3 ubiquitin protein ligase)
34	HIF1A	4910		(hypoxia inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor))
35	S6K1	10436	RPS6KB1	(ribosomal protein S6 kinase, 70kDa, polypeptide 1)
36	S6	10429	RPS6	(ribosomal protein S6)
37	eIF4B	3285		(eukaryotic translation initiation factor 4B)
38	eEF2K	24615		(eukaryotic elongation factor-2 kinase)
39	PDCD4	8763		(programmed cell death 4 (neoplastic transformation inhibitor))
40	FMRP	3775		(fragile X mental retardation 1)
41	PP2A	9299	PPP2CA	(protein phosphatase 2, catalytic subunit, alpha isozyme)
42	P53	11998	TP53	(tumor protein p53)
43	BAD	936		(BCL2-associated agonist of cell death)
44	4EBP1	3288		(eukaryotic translation initiation factor 4E)
45	eIF4E	3287		(eukaryotic translation initiation factor 4E)

46	FIP200	15574	RB1CC1	(RB1-inducible coiled-coil 1)
47	ULK1	12558		(unc-51-like kinase 1)
48	ATG13	29091		(autophagy related 13)
49	PLD1	9067		(phospholipase D1, phosphatidylcholine-specific)
50	MAP4K3	6865		(mitogen-activated protein kinase kinase kinase 3)
51	hVPS34	8974	PIK3C3	(phosphoinositide-3-kinase, class 3)
52	RAGA	16963	RRAGA	(Ras-related GTP binding A)
53	RAGB	19901	RRAGB	(Ras-related GTP binding B)
54	RAGC	19902	RRAGC	(Ras-related GTP binding C)
55	RAGD	19903	RRAGD	(Ras-related GTP binding D)
56	IKKB	5960	IKBKB	(inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase beta)
57	CDK1	1722		(cyclin-dependent kinase 1)
58	MDM2	6973		(Mdm2, p53 E3 ubiquitin protein ligase homolog (mouse))
59	AMPK	9377	PRKAA2	(protein kinase, AMP-activated, alpha 2 catalytic subunit)
60	LKB1	11389	STK11	(serine/threonine kinase 11)
61	CAB39	20292		(calcium binding protein 39)
62	STRAD	30172	STRADA	(STE20-related kinase adaptor alpha)
63	BCL2	990		(B-cell CLL/lymphoma 2)
64	BECN1	1034		(beclin 1, autophagy related)
65	DSH	3085	DVL1L1	(dishevelled, dsh homolog 1 (Drosophila)-like 1)
66	GSK3B	4617		(glycogen synthase kinase 3 beta)
67	PTEN	9588		(phosphatase and tensin homolog)
68	REDD1	24944		(DNA-damage-inducible transcript 4)

La pagina dei risultati sarà valida per 24h e sarà sempre raggiungibile tramite un url unico composto con il timestamp di quando il job è stato sottomesso, nel caso della pathway della TSC l'url generato è il seguente:

http://atlas.dmi.unict.it/eshmun/execute_1.php?code=3c205d9f1cddb7cea42d1b6b7255c339

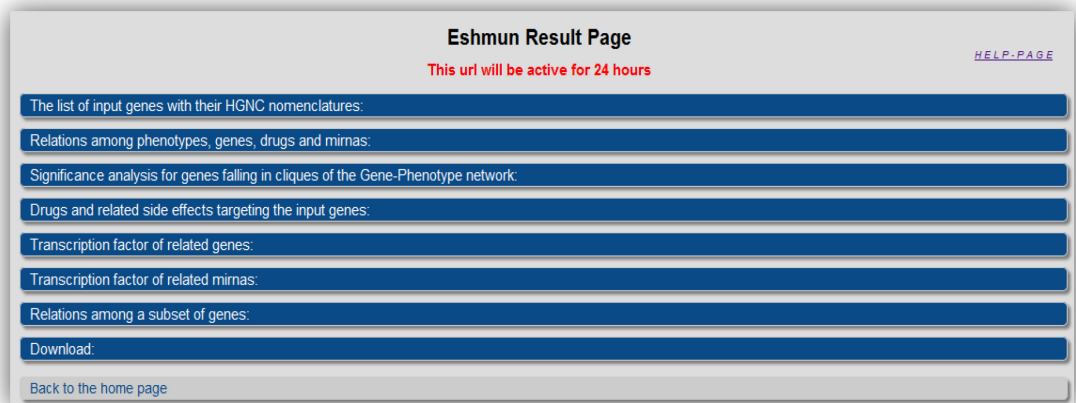


FIGURE 62: SEZIONE DI OUTPUT NEI RISULTATI DI ESHMUN

La pagina dei risultati è divisa in otto sezioni:

- **The list of input genes with their HGNC nomenclatures:** è composto da 3 possibili sottosezioni "Genes found", "Genes not found", "Genes discarded". La sezione dei geni mostra tutti i geni per cui è stato trovato un identificativo corrispondente in HGNC, i geni restanti sono mostrati nella sezione "genes discarded". Mentre la sezione "genes not found" mostra tutti i geni non contenuti nel database HGNC. Nel caso della TSC, tutti i 68 geni sono trovati nel database HGNC.

The list of input genes with their HGNC nomenclatures:

Genes found:

#	INPUT	GENE NAME	GENE HGNC
1	9871	RASA1	9871
2	11187	SOS1	11187
3	9829	RAF1	9829
4	3051	TSC22D3	3051
5	6840	MAP2K1	6840
6	6842	MAP2K2	6842
7	6877	MAPK3	6877
8	6871	MAPK1	6871
9	2674	DAPK1	2674
10	9393	PRKCA	9393
11	4566	GRB2	4566
12	7765	NF1	7765
13	4066	GAB1	4066
14	6125	IRS1	6125
15	8975	PIK3CA	8975
16	8979	PIK3R1	8979
17	10430	RPS6KA1	10430
18	8993	PIP	8993
19	10810	SGK1	10810
20	8809	PDK1	8809

FIGURE 63: ESEMPIO DEI GENI TROVATI RELATIVI ALLA TSC

- **Relations among phenotype, genes, drugs, and mirnas:** fornisce una heatmap che mostra la similarità fenotipica tra tutti i fenotipi correlati ai geni di input; una network fenotipica di geni dove i nodi sono i geni collegati da archi i cui pesi rappresentano la similarità fenotipica tra le malattie correlate a questi geni; due network geni-miRNA validati e geni-miRNA predetti in cui è possibile vedere la relazione tra i geni target e i loro miRNA; la network gene-drug che mostra le relazioni tra i geni target e le drug; ed infine una network dove i nodi sono i geni e gli archi rappresentano le relazioni proteina-proteina dei geni stessi. In ogni network, i nodi sono classificati con diversi colori: rosso per i geni, blu per gli oncogeni, giallo per gli oncosoppressori, verdi per i miRNA e verde scuro per le drug.

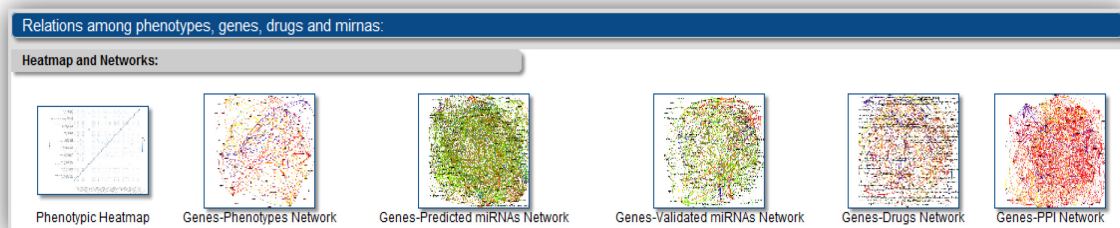


FIGURE 64. SEZIONE DELLE NETWORK GENERATE

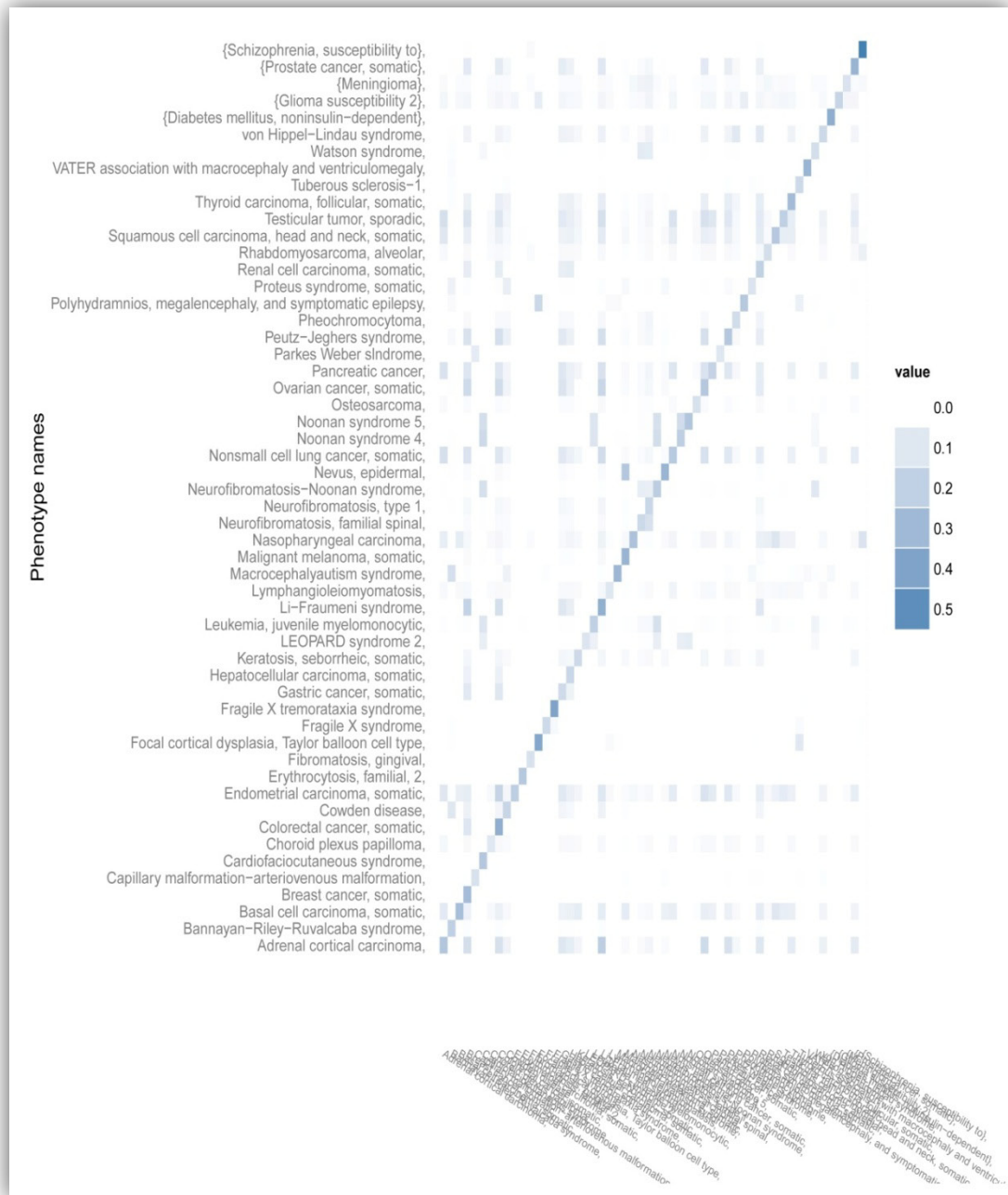


FIGURE 65: HEATMAP FENOTIPICA DELLA TSC

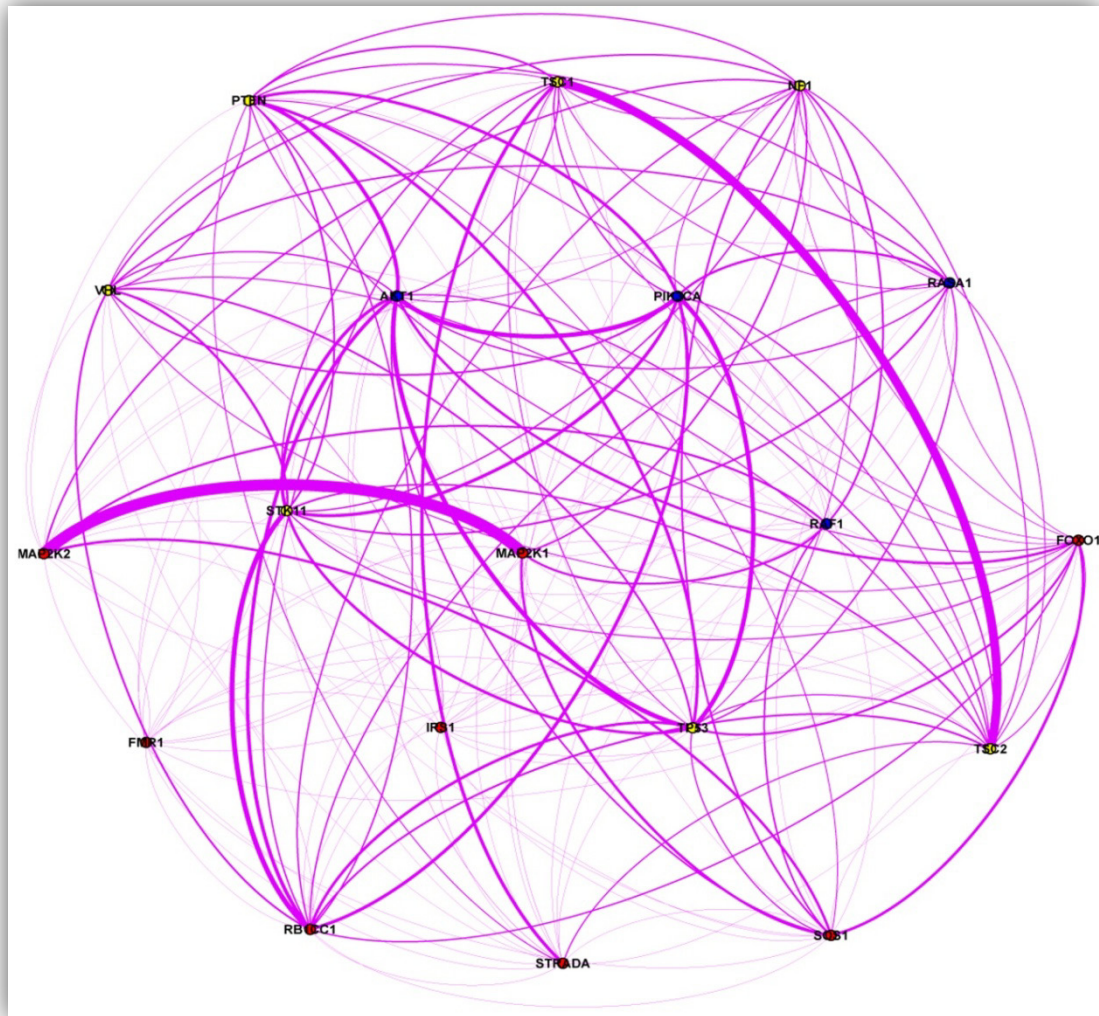


FIGURE 66: NETWORK GENE-GENE DELLA TSC. I NODI SONO I GENI GLI ARCHI LA CORRELAZIONE FENOTIPICA

La Figura mostra la network gene-to-gene in cui i nodi sono geni collegati dalla similarità fenotipica nel caso della TSC, dei 68 geni iniziali solo 19 di essi risultano formare un cluster strettamente connesso (PTEN, TSC1, NF1, VHL, AKT1, PIK3CA, RASA1, MAP2K2, STK11, MAP2K1, RAF1, FOXO, FMR1, IRS1, TP53, TSC2, RB1CC1, STRADA, SOS1). Dalla network emerge che alcuni geni condividono una elevata similarità fenotipica questo è evidente grazie allo spessore dell'arco, i casi rilevanti sono i seguenti:

- TSC1-TSC2, fenotipi padri prelevati da OMIM in cui sono coinvolti:

>605284#Hamartin (tuberous sclerosis 1 gene)#TSC1

191100#Tuberous sclerosis-1,

606690#Lymphangiomyomatosis,

607341#Focal cortical dysplasia, Taylor balloon cell type,

>191092#Tuberin (tuberous sclerosis 2 gene)#TSC2

191100#Tuberous sclerosis-2,

606690#Lymphangioliomyomatosis, somatic,

- MAP2K2-MAP2K1-SOS1-FOXO, fenotipi padri prelevati da OMIM in cui sono coinvolti:

>601263#Mitogen-activated protein kinase kinase 2#MAP2K2

115150#Cardiofaciocutaneous syndrome,

>176872#Mitogen-activated protein kinase kinase 1#MAP2K1

115150#Cardiofaciocutaneous syndrome,

>182530#Son of sevenless, Drosophila, homolog of, 1#SOS1

135300#Fibromatosis, gingival,

610733#Noonan syndrome 4,

>136533#Forkhead box O1A (forkhead in rhabdomyosarcoma)#FOXO1

268220#Rhabdomyosarcoma, alveolar,

- VHL-PIK3CA-PTEN-AKT1-STK11-RB1CC1-TP53, fenotipi padri prelevati da OMIM in cui sono coinvolti:

>608537#VHL gene#VHL

193300#von Hippel-Lindau syndrome,

144700#Renal cell carcinoma, somatic,

171300#Pheochromocytoma,

263400#Erythrocytosis, familial, 2,

>171834#Phosphatidylinositol 3-kinase, catalytic, alpha polypeptide#PIK3CA

167000#Ovarian cancer, somatic,

114480#Breast cancer, somatic,

114500#Colorectal cancer, somatic,

137215#Gastric cancer, somatic,

114550#Hepatocellular carcinoma, somatic,

211980#Nonsmall cell lung cancer, somatic,

182000#Keratosis, seborrheic, somatic,

162900#Nevus, epidermal,

>601728#Phosphatase and tensin homolog (mutated in multiple advanced cancers#PTEN

158350#Cowden disease,

153480#Bannayan-Riley-Ruvalcaba syndrome,

607174#{Meningioma},

613028#{Glioma susceptibility 2},

605309#Macrocephalyautism syndrome,

276950#VATER association with macrocephaly and ventriculomegaly,

176807#{Prostate cancer, somatic},

188470#Thyroid carcinoma, follicular, somatic,

155600#Malignant melanoma, somatic,

608089#Endometrial carcinoma, somatic,

275355#Squamous cell carcinoma, head and neck, somatic,

>602216#Serinethreonine protein kinase-11#STK11

175200#Peutz-Jeghers syndrome,

273300#Testicular tumor, sporadic,

>606837#RB1 -inducible coiled-coil 1#RB1CC1

114480#Breast cancer, somatic,

>191170#Tumor protein p53#TP53

114500#Colorectal cancer,

151623#Li-Fraumeni syndrome,

114550#Hepatocellular carcinoma,

259500#Osteosarcoma,

260500#Choroid plexus papilloma,

607107#Nasopharyngeal carcinoma,

260350#Pancreatic cancer,

202300#Adrenal cortical carcinoma,

114480#Breast cancer,

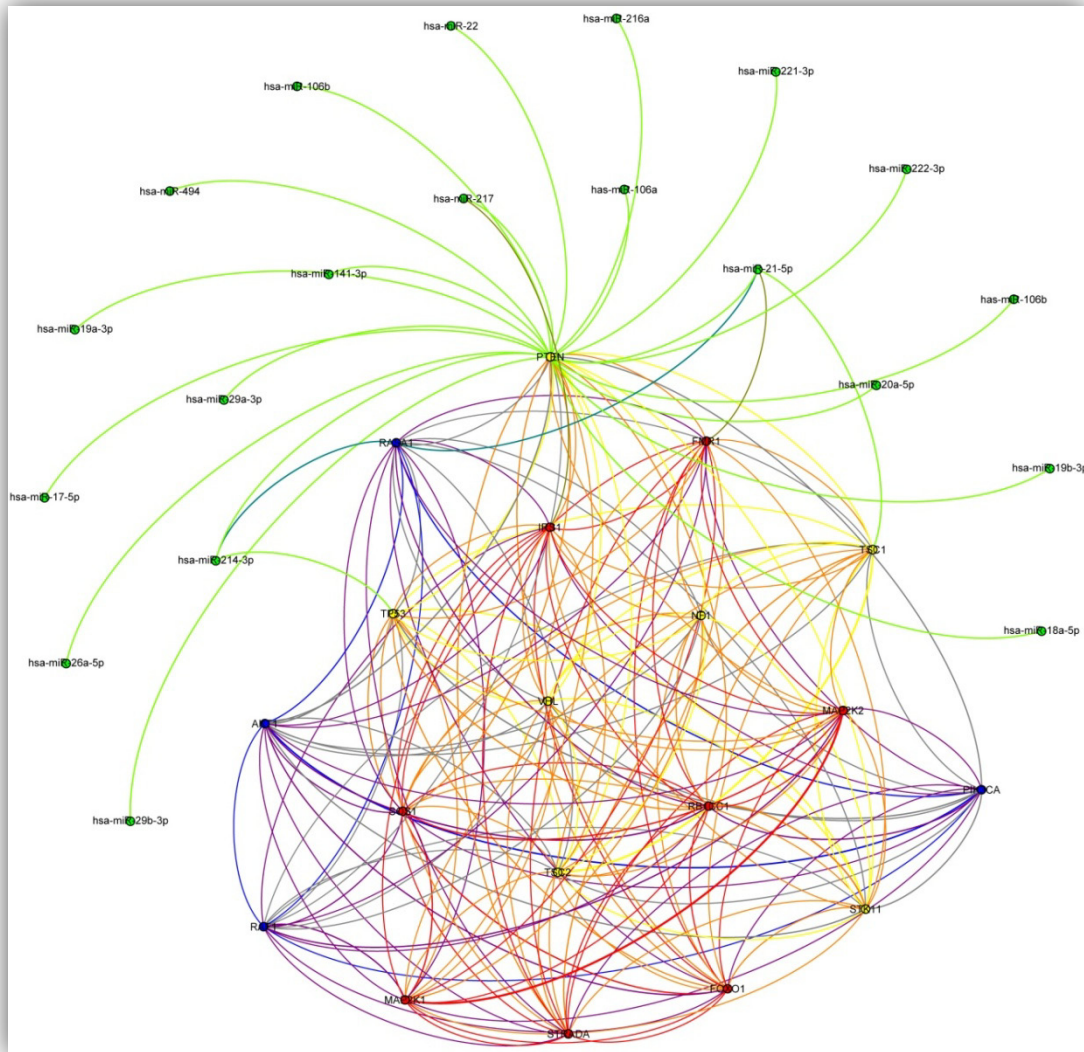


FIGURE 67: NETWORK GENI-MIRNA VALIDATI (TSC). DETTAGLIO DEL GENE PTEN E MIRNA VALIDATI

La figura mostra la network geni-miRNA validati con il dettaglio del gene PTEN, questo risulta essere insieme a TP53, il target del maggior numero di miRNA prelevati da TargetScan. PTEN è coinvolto in molti processi biologici come l'angiogenesi, lo sviluppo del muscolo cardiaco, sviluppo del sistema nervoso centrale, induzione dell'apoptosi, regolazione negativa del ciclo cellulare, regolare negativa della proliferazione cellulare, regolazione negativa della migrazione cellulare e regolazione della stabilità della proteina, mentre TP53 è coinvolto nella regolazione negativa dell'apoptosi, regolazione negativa del ciclo cellulare, regolazione negativa della replicazione del DNA, regolazione negativa della proliferazione dei fibroblasti, regolazione negativa della proliferazione dei neuroblastomi, risultando quindi molto collegati con i fenotipi descritti nel TSC come quelli relativi al sistema nervoso centrale, inoltre sono coinvolti in malattie come la sindrome Cowden, colorectal cancer e breast cancer che rafforzano i legami fenotipici espressi nella heatmap e nella network gene-gene. Entrambi i geni sono targettati da hsa-miR-214

che è coinvolto in molte malattie come l'ipertrofia cardiaca, attacco cardiaco e malattie vascolari, mentre hsa-miR-21-5p targetta PTEN sia il gene della malattia TSC1, recenti studi [176] dimostrano che tale miRNA governa l'attivazione di TORC1 nel cancro renale causando la proliferazione cellulare, come mostrato in figura maggior miR-21 attenua i livelli della proteina di PTEN, con conseguente attivazione di Akt, che inattiva la tuberina per incrementare l'attività di TORC1 portando alla proliferazione, migrazione ed invasione delle cellule renali tumorali, che risulta essere un organo interessato nel TSC.

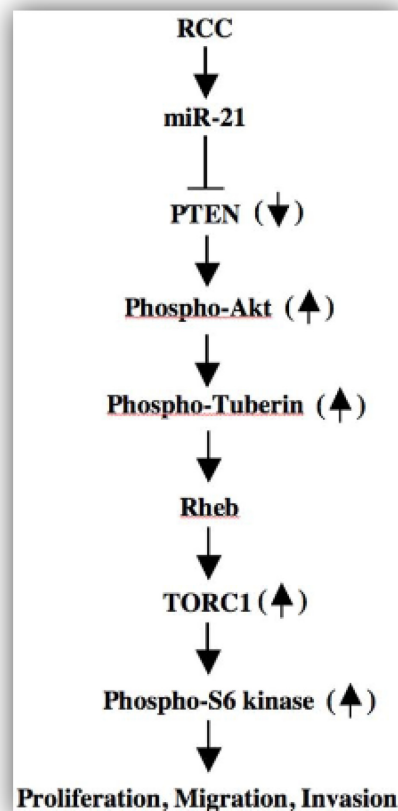


FIGURE 68: SCHEMA DELL'AZIONE DEL MIR-21

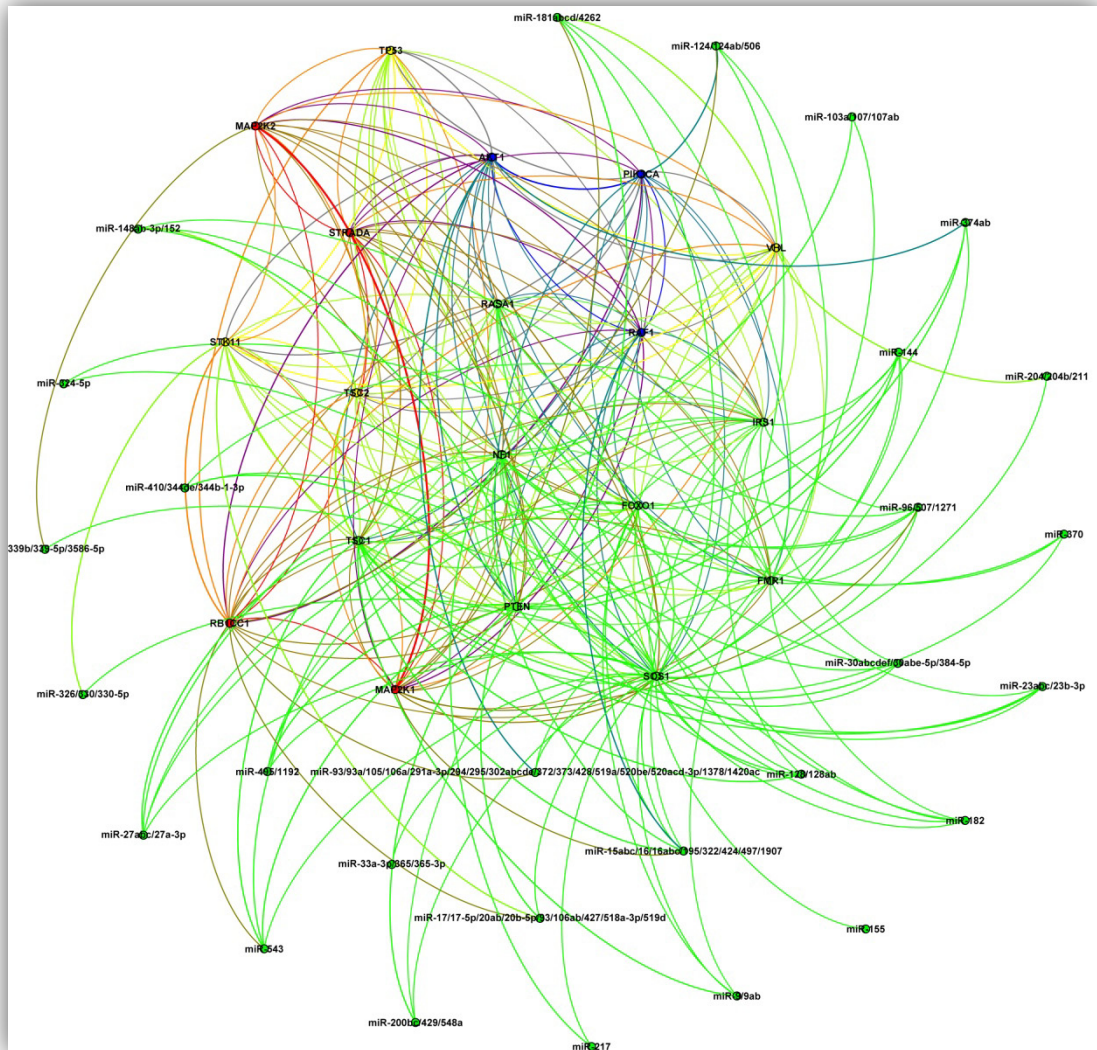


FIGURE 69: NETWORK GENI-MIRNA PREDETTI (TSC). DETAGLIO DEL GENE SOS1 E MIRNA PREDETTI

La figura mostra la network geni-miRNA predetti con dettaglio del gene SOS1 che insieme a FOXO1 risulta essere il gene con maggior numero di connessioni. La network mostra come questi miRNA targettino oltre il gene SOS1 altri geni appartenenti al cluster fondamentale dei geni collegati dal punto di vista fenotipico. SOS1 e TSC1 condividono 7 miRNA: miR-23abc/23b-3p, miR-534, miR-9/9ab, miR-27abc/27a-3p, miR-128/128ab, miR-30abcde/30abe-5p/384-5p, miR-15abc/16/16abc/195/322/424/497/1907. Il gene SOS1 fornisce istruzioni per creare una proteina coinvolta nelle pathway all'interno di cellule che controllano la crescita e lo sviluppo. Questa proteina aiuta a controllare (regolare) l'attivazione delle pathway di segnalazione RAS/MAPK, che aiuta nel controllo di diverse ed importanti funzioni cellulari. Specificatamente regola la crescita e la divisione delle cellule (proliferazione), il processo da cui cellule mature svolgono funzioni specifiche (differenziazione), movimento cellulare, e autodistruzione delle cellule (apoptosi). La proteina di SOS1 regola un'altra proteina, chiamata Ras, che stimola le cellule a crescere

e dividersi. Questa regolazione controlla strettamente la crescita di cellule e tessuti, ed è specialmente importante per il corretto sviluppo embrionale. SOS1 è inoltre coinvolto nella Noonan syndrome che è messa in relazione fenotipica con il TSC dallo stesso tool Eshmun.

Il miR-15abc/16/16abc/195/322/424/497/1907 risulta essere il miRNA che targetta il maggior numero di geni, tra questi uno dei più importanti è BCL2, infatti alcuni studi [177] dimostrano che i miRNA miR-15a e miR-16, che agiscono come soppressori tumorali putativi targettando l'oncogene BCL2, sono inoltre implicati nel ciclo cellulare, nell'apoptosi e la proliferazione cellulare.

Anche il miR-96/507/1271 targettano un elevato numero di geni, alcune ricerche [178] dimostrano che miR-96 e miR-1271 targettano lo stesso set di geni che sono coinvolti nella segnalazione neurale e nell'eccitazione cellulare, suggerendo che questi miRNA potrebbero avere un ruolo nella regolazione di aspetti dello sviluppo o funzionamento neurale umano.

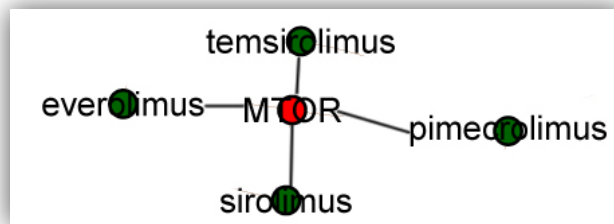


FIGURE 70: NETWORK GENI-DRUG (TSC). DETTAGLIO DEL GENE MTOR CON LE RELATIVE DRUG

Dettaglio della figura geni-drug, il particolare mostra il complesso mTOR che viene targettato da sirolimus, everolimus, temsirolimus e pimecrolimus. Il sirolimus (detto anche rapamicina) è usato come trattamento della TSC [179], recenti trial clinici hanno dimostrato una effettiva riduzione del volume degli angiomiolipomi renali, angiofibromi così come un miglioramento della funzione polmonaria, tuttavia quando la terapia viene fermata queste lesioni tendono ad incrementare ancora di volume. L'everolimus ed il temsirolimus sono dei derivati del sirolimus che agiscono come inibitori del complesso mTOR, nel dettaglio everolimus è usato come immunosoppressore per prevenire il rigetto dopo un trapianto di organi, mentre il temsirolimus è usato per il trattamento di carcinomi renali. L'altra drug, il pimecrolimus, che targetta mTOR è usata invece come trattamento per dermatite atopica (eczema).

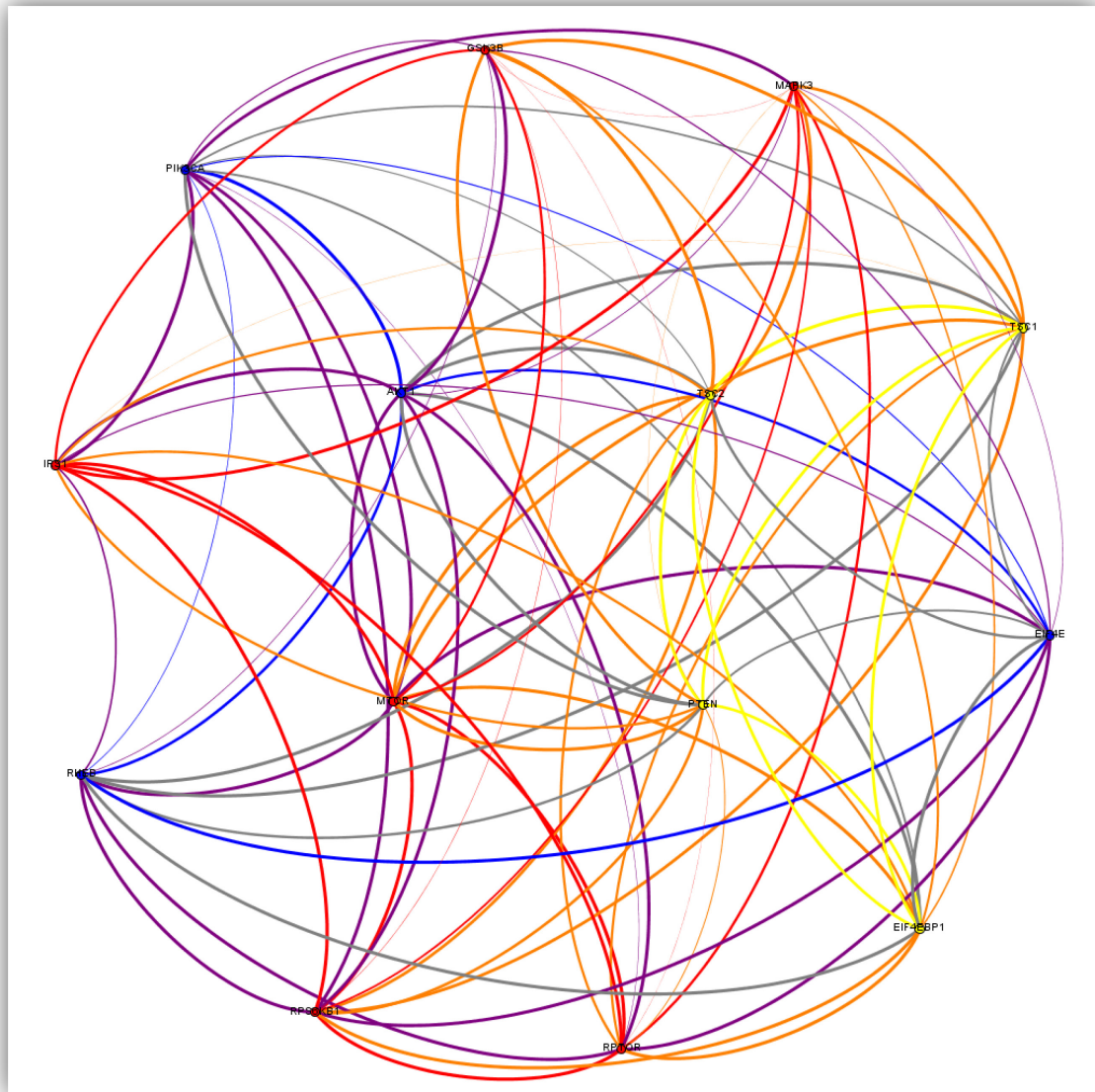


FIGURE 71: NETWORK GENE-GENE CON ARCHI PASSATI SULLE INTERAZIONE PROTEINA-PROTEINA (TSC)

La Figura mostra una network gene-gene in cui gli archi rappresentano le relazioni proteina-proteina dei geni stessi. La network mostra il cluster di geni maggiormente connesso con un min di 40 ad un massimo di 63 archi per gene, tale cluster è formato da 14 geni: GSK3B, PIK3CA, MAPK3, IRS1, AKT1, TSC2, TSC1, RHEB, MTOR, PTEN, EIF4E, RPS6KB1, RPTOR, EIF4EBP1 di cui PTEN, TSC1, AKT1, PIK3CA, IRS1, TSC2 sono parte del cluster fenotipico principale.

- **Subnetworks of connected input genes in Gene-Phenotype Network and its statistical significance:** mostra la significatività statistica per i cluster massimali di geni connessi.

Significance analysis for genes falling in cliques of the Gene-Phenotype network:	
Subnetworks of connected input genes	statistical significance
TSC2 - AKT1 - TSC1 - TP53 - PIK3CA - PTEN - NF1 - RASA1 - SOS1 - STRADA - VHL -	0.94
TSC2 - AKT1 - TSC1 - TP53 - PIK3CA - PTEN - NF1 - RASA1 - SOS1 - STRADA - FMR1 -	1
TSC2 - AKT1 - TSC1 - TP53 - PIK3CA - PTEN - NF1 - RASA1 - SOS1 - MAP2K1 - MAP2K2 - FMR1 -	0.97
TSC2 - AKT1 - TSC1 - TP53 - PIK3CA - PTEN - NF1 - RASA1 - SOS1 - MAP2K1 - MAP2K2 - STK11 - RAF1 - VHL - RB1CC1 - FOXO1 -	0.99
TSC2 - AKT1 - TSC1 - TP53 - PIK3CA - PTEN - IRS1 - RAF1 - VHL - RB1CC1 - FOXO1 -	0.85

FIGURE 72: SEZIONE DELLE CLIQUE MASSIMALI GENERATE RELATIVE ALLA TSC

I clique massimali nel caso del complesso della TSC vanno da 11 a 16 geni, con clique massimali si intendono clique che non possono essere estese oltre, la significatività statistica è alta proprio perché sono clique massimali questo implica che alcuni geni potrebbero essere sostituiti con altri geni che hanno un numero maggiore di fenotipi OMIM che porterebbero una relazione maggiore con altri geni e che renderebbe la clique più specifica.

- **Drugs and related side effects targeting the input genes:** contiene una tabella con tutte le drug che targettano i geni di input. Per ogni drug sono mostrate, il codice DrugBank, il nome ufficiale della drug, per le drug che hanno una simile il link alle prime 10 drug con simile simile alla drug data secondo la formula di Tanimoto, in asterisco vengono indicate le drug presenti nella pagina dei risultati, gli effetti collaterali in cui sono coinvolti con relativo link a SIDER, i fenotipi in cui tali effetti collaterali con collegamento ad OMIM e i target di queste drug. Per il caso della TSC vengono riportate 86 drug.

First 10 selected-by-Tanimoto smiles similar to chosen drug

ID	SMILE
DB00877	<chem>[H][C@@]1(C[C@H](C)[C@]2([H])CC(=O)[C@H](C)C=C(C)[C@@H](O)[C@@H](OC)C(=O)[C@H](C)C[C@H](C)C=CC=CC=C(C)[C@H](C)[C@]3([H])CC[C@@H](C)[C@@](O)(O3)C(=O)C(=O)N3CCCC[C@@]3([H])C(=O)O2)OC)CC[C@@H](O)[C@@H](C1)OC</chem>

ID	Score	SMILE	TARGET HGNC	OFF-TARGET
DB00864	0.81	<chem>CO[C@@H]1C[C@@H](CC[C@H]1O)C=C(C)[C@H]1OC(=O)[C@@H]2CCCCN2C(=O)C(=O)[C@]2(O)O[C@@H]([C@H]([C@H]2C)OC)[C@H]([C@@H](C)CC(C)=C[C@@H](CC=C)C(=O)C[C@H](O)[C@H]1C)OC</chem>	3711 -	
* DB00337	0.78	<chem>CC[C@@H]1C=C(C)C[C@H](C)C[C@H](OC)[C@H]2O[C@](O)([C@H](C)C[C@@H]2OC)C(=O)C(=O)N2CCCC[C@H]2C(=O)O[C@@H]([C@@H](C)[C@@H](O)CC1=O)C(C)=C[C@@H]1CC[C@H](C)[C@@H](C1)OC</chem>	3942 - 3711	3942 -
DB00175	0.58	<chem>CC[C@H](C)C(=O)O[C@H]1C[C@H](O)C=C2C=C[C@H](C)[C@H](CC[C@@H](O)C[C@@H](O)CC(O)=O)[C@@H]12</chem>	5006 -	
DB00681	0.58	<chem>C[C@H]1O[C@@H](O[C@@H]2[C@@H]3O[C@@](O)(C)[C@H](O)[C@H]3C(=O)C[C@H](O)C[C@@H](O)[C@H](O)CC[C@@H](O)C[C@H](O)CC(=O)O[C@@H](C)[C@H](C)[C@H](O)[C@@H](C)C=CC=CC=CC=CC=C2)[C@@H](O)[C@@H](N)[C@@H]1O</chem>	-	
DB00602	0.57	<chem>CO[C@H]1C[C@@H](O[C@@H](C)[C@@H]1O)O[C@H]1[C@H](C)O[C@H](C[C@@H]1OC)O[C@H]1[C@@H](C)C=CC=C2CO[C@@H]3[C@H](O)C(C)=C[C@@H](C=O)O[C@H]4C[C@@H](CC=C1/C)O[C@@]1(C)[C@H](C)[C@@H](C(C)O1)C4[C@]23O.CC[C@@H](C)[C@H]1O[C@@]2(CC[C@@H]1C)O[C@@H]1CC=C(C)[C@@H](O[C@@H]3O[C@@H](C)[C@H](O)[C@@H]4O[C@@H](C)[C@H](O)[C@@H](OC)C4)[C@@H](OC)C3)[C@@H](C)C=CC=C3/CO[C@@H]4[C@H](O)C(C)=C[C@@H](C=O)O[C@@H](C1)C2)[C@]34O</chem>	4328 - 4083	
DB00198	0.55	<chem>CCOC(=O)C1=C[C@@H](OC(CC)CC)[C@H](NC(C)=O)[C@@H](N)C1</chem>	1863 - 7758 7759	
DB00227	0.54	<chem>CC[C@H](C)C(=O)O[C@H]1C[C@@H](C)C=C2C=C[C@H](C)[C@H](CC[C@@H]3[C@H](O)CC(=O)O3)[C@@H]12</chem>	5006 - 6148 4853	
DB00207	0.53	<chem>CC[C@H]1OC(=O)[C@H](C)[C@@H](O[C@H]2C[C@@](C)OC)[C@@H](O)[C@H](C)O2)[C@H](C)[C@@H](O[C@@H]2O[C@H](C)C[C@@H]([C@H]2O)N(C)C)[C@](C)(O)C[C@@H](C)C[N](C)[C@H](C)[C@@H](O)[C@]1(C)O</chem>	-	
DB00284	0.53	<chem>C[C@H]1O[C@@H](O[C@@H]2[C@@H](CO)O[C@@H](O[C@@H]3[C@@H](CO)O[C@@H](O)[C@H](O)[C@@H]3O)[C@H](O)[C@H]2O)[C@H](O)[C@@H](O)[C@@H]1N[C@@H]1C=C(CO)[C@H](O)[C@H](O)[C@H]1O</chem>	7043 - 477 4065 10856	
DB00641	0.53	<chem>[H][C@]12[C@H](C[C@@H](C)C=C1C=C[C@H](C)[C@@H]2CC[C@@H]1C[C@@H](O)CC(=O)O1)OC(=O)C(C)CC</chem>	5006 - 6155	

FIGURE 73: DRUG SIMILARITY DB0087 (SIROLIMUS), MAGGIOR SIMILARITÀ CON DB00864 (TACROLIMUS) E DB00337 (PIMECROLIMUS)

- **Transcription factor of related genes:** una lista di tutti i fattori di trascrizione dei geni con i loro codici HGNC, la loro classe e famiglia, con il link al database Jaspard. Per l'esempio della TSC ci sono due fattori di trascrizione uno per il gene TP53 e l'altro per HIF1A

Transcription factor of related genes:				
#	Transcription Factor (Jaspard ID)	Hgnc Gene	Class	Family
1	MA0106.1	11998	Zinc-coordinating	Loop-Sheet-Helix
2	MA0259.1	4910::700	Zipper-Type	Helix-Loop-Helix

FIGURE 74: TF DEI GENI (TS)

- **Transcription factor of related mirnas:** Una lista di tutti i fattori di trascrizione legati ai miRNA, con il nome ed il codice HGNC del gene relativo, la funzione del miRNA e la patologia a cui è collegato, l'attività del miRNA e il codice dell'articolo NCBI da cui tale informazione è estratta.

Nel caso della TSC vengono individuati 53 fattori di trascrizione relativi ai miRNA

20	TP53	11998	mir-519d	activate its transcription	see HMDD (http://cmbi.bjmu.edu.cn/hmdd)	Activation	22262409
21	TP53	11998	mir-34a	MicroRNA-34a (miR-34a), a transcriptional target of p53	see HMDD (http://cmbi.bjmu.edu.cn/hmdd)	Activation	22198213
22	TP53	11998	mir-34a	miR-34 miRNAs are p53 targets	see HMDD (http://cmbi.bjmu.edu.cn/hmdd)	Activation	22020437
23	TP53	11998	mir-34b	miR-34 miRNAs are p53 targets	see HMDD (http://cmbi.bjmu.edu.cn/hmdd)	Activation	22020437
24	TP53	11998	mir-34c	miR-34 miRNAs are p53 targets	see HMDD (http://cmbi.bjmu.edu.cn/hmdd)	Activation	22020437
25	TP53	11998	mir-34a	MicroRNA-34a (miR-34a) is a transcriptional target of p53 and is down-regulated in pancreatic cancer.	see HMDD (http://cmbi.bjmu.edu.cn/hmdd)	Activation	21909380
26	TP53	11998	mir-29	transcriptional activation of miR-29 is triggered in response to DNA damage and occurs in a p53-dependent manner	see HMDD (http://cmbi.bjmu.edu.cn/hmdd)	Activation	21522133
27	TP53	11998	mir-200a	The miR-200 family members transactivated by p53 then repress ZEB1/2 expression.	see HMDD (http://cmbi.bjmu.edu.cn/hmdd)	Activation	21518799
28	TP53	11998	mir-200b	The miR-200 family members transactivated by p53 then repress ZEB1/2 expression.	see HMDD (http://cmbi.bjmu.edu.cn/hmdd)	Activation	21518799
29	TP53	11998	mir-200c	The miR-200 family members transactivated by p53 then repress ZEB1/2 expression.	see HMDD (http://cmbi.bjmu.edu.cn/hmdd)	Activation	21518799
30	TP53	11998	mir-192	p53-regulated miR-192 family members also repress ZEB2 expression.	see HMDD (http://cmbi.bjmu.edu.cn/hmdd)	Activation	21518799
31	VHL	12687	mir-210	involvement of the HIF signaling pathway Endothelial Cell Response to Hypoxia Receptor Tyrosine Kinase Ligand Ephrin-A3 inhibition	Heart Failure Lung Neoplasms Muscular Disorders, Atrophic Breast Neoplasms Carcinoma, Renal Cell Adenocarcinoma Lupus Vulgaris Pre-Eclampsia Adrenocortical Carcinoma	Activation	18316553
32	AKT1	391	let-7e	NA	Carcinoma, Hepatocellular Heart Failure Lung Neoplasms Neoplasms Digestive System Neoplasms Head and Neck Neoplasms Muscular Disorders, Atrophic Sarcoma, Kaposip Lupus Vulgaris Melanoma Ovarian Neoplasms Pancreatic Neoplasms Pituitary Neoplasms Lymphoma, Primary Effusion Neoplasms	activation	19699171
33	AKT1	391	mir-181c	NA	Lung Neoplasms Leukemia, Myeloid, Acute Leukemia, Lymphocytic, Chronic, B-Cell	activation	19699171

FIGURE 75: TF DEI MIRNA (TSC)

- **Relations among a subset of genes:** In questa sezione è possibile scegliere dei geni dalla lista iniziale e cercare gli elementi comuni tra fenotipi, miRNA predetti, miRNA validati, drug e fattori di trascrizione. La figura mostra un esempio di caso relativo alla pathway della TS, dei tre geni selezionati PIK3CA, TP53, AKT1 hanno in comune solo il fenotipo del breast cancer (114480), è possibile calcolare anche il p-value dei geni selezionati se essi condividono almeno un fenotipo, nel caso specifico la terna AKT1,TP53 e PIK3CA hanno un p-value di 0.16.

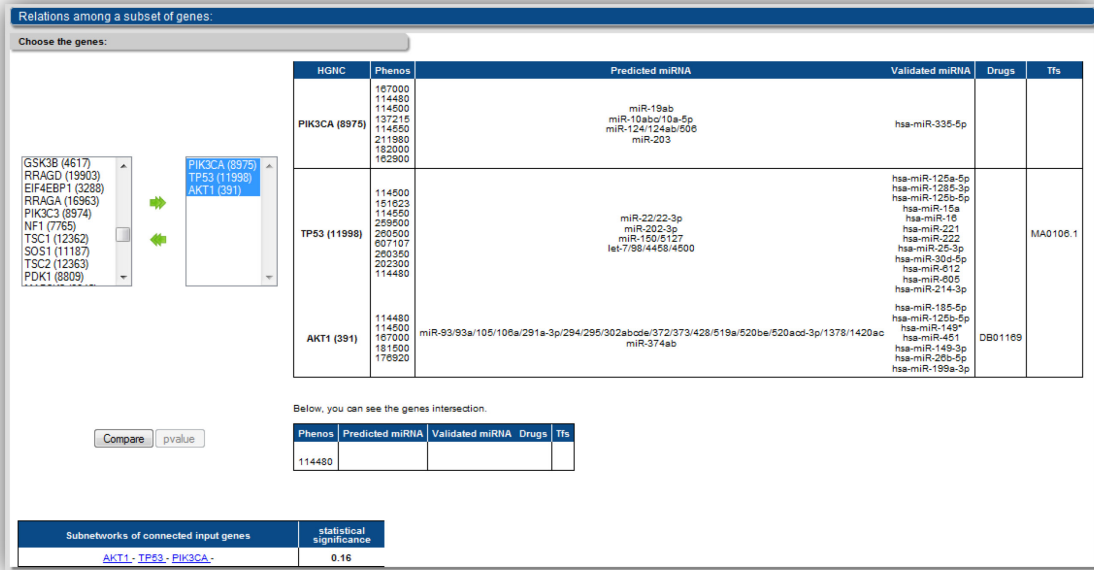


FIGURE 76: SEZIONE PER IL CALCOLO DEGLI ELEMENTI COMUNI TRA SET DI GENI E CALCOLO PVALUE

- **Download:** Contiene per ogni network il file pdf e il file gdf come codice sorgente che può essere modificato e visualizzato usando il tool Gephi.

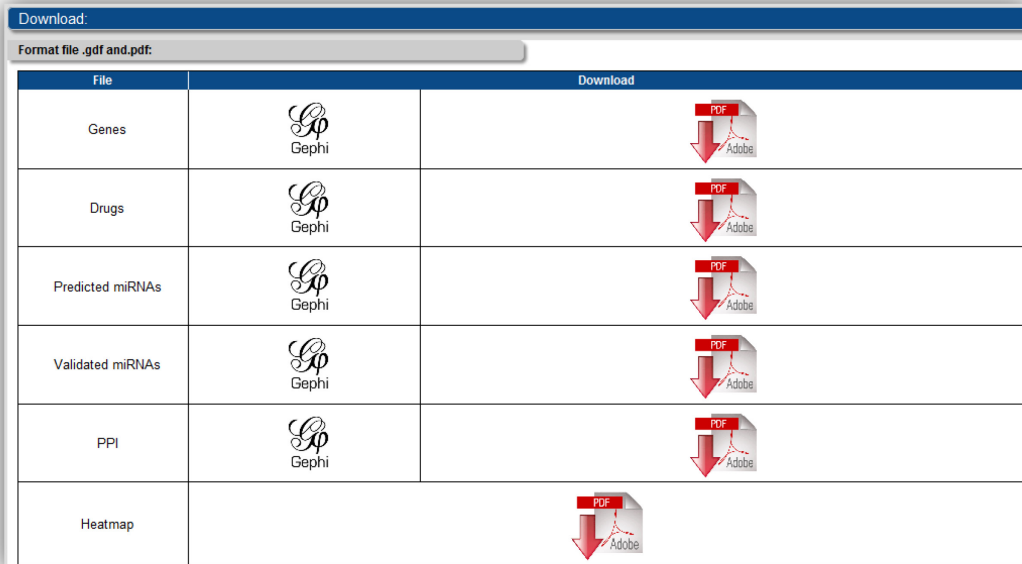


FIGURE 77: SEZIONE DOWNLOAD FILE

4.7. Conclusioni

Ci sono molte malattie ereditarie negli umani, ognuna delle quali ha una combinazione specifica di aspetti fenotipici, ma l'analisi computazionale dei dati fenotipici è stata ostacolata dalla mancanza di un vocabolario controllato per descriverli, infatti c'è un numero di problemi irrisolti circa la descrizione computazionale e l'analisi di fenotipi umani, tuttavia, alcuni lavori in progress forniscono uno standard uniforme.

L'analisi fenotipica ha una grande importanza nella patofisiologia delle network cellulari poiché può offrire indizi circa i gruppi di geni che insieme creano pathway e moduli, in cui le disfunzioni possono portare a conseguenze di fenotipi simili. Studi recenti mostrano come questo potenziale di correlazione fenotipica di aspetti si evidenzia su di una network genetica o cellulare su un'ampia scala genomica. La similarità fenotipica permette di usare informazioni circa una malattia le cui origini genetiche sono note per altre le cui origine genetiche sono sconosciute.

Eshmun è il primo web tool che permette di combinare informazioni eterogenee da database multipli fornendo una interfaccia user-friendly per l'analisi di tali dati. Lo scopo principale di Eshmun è di fornire un sistema periodicamente aggiornato, che permetta all'utente di analizzare una pathway di geni secondo le loro similarità fenotipiche ed identificare e visualizzare facilmente tramite network di miRNA prendetti e validati, le drug e le interazioni proteina-proteina collegate alle pathway scelte. In aggiunta ai file PDF, Eshmun fornisce i file sorgente in gdf per rigenerare e visualizzare queste network usando Gephi. Informazioni aggiuntive fornite dal tool sono gli effetti collaterali per ogni drug coinvolta ed i possibili fattori di trascrizione per queste geni. Per ogni pathway di input è generato un job con un identificativo unico che permette all'utente di recuperare i dati per un arco di 24 ore senza dover rieseguire il test. Per testare Eshmun, sono state analizzate le pathway collegate alle malattie umane contenute in KEGG, come il cancro alla prostata, malattie cardiovascolari, endocrine e metaboliche, immunitarie, infettive e neurodegenerative. I risultati possono essere visti cliccando su "Default Search" nella home-page. Ovviamente questo approccio può essere migliorato nelle seguenti direzioni. Il primo miglioramento è la standardizzazione delle descrizioni fenotipiche, la seconda permetterà agli utenti di visualizzare e diversificare i miRNA coinvolti secondo i loro livelli di espressione, la terza riguarda l'identificazione di geni off-target di una pathway.

BIBLIOGRAFIA

1. *MicroRNAs: small RNAs with a big role in gene regulation*. **Lin, He and J., Hannon Gregory**. Luglio 2004, *Nature Reviews, Genetics*, Vol. 5, pp. 522-532.
2. *Small RNAs: Classification, Biogenesis, and Function*. **Narry Kim, V.** 1, Febbraio 23, 2005, *Molecules and Cells*, Vol. 19, pp. 1-15.
3. *Short RNAs Repress Translation after Initiation in Mammalian Cells*. **Peterson, C. p., et al.** Febbraio 17, 2006, *Molecular Cell*, pp. 533-542.
4. *The functions of animal microRNAs*. **Ambros, Victor**. Settembre 16, 2004, *Nature*, Vol. 431, pp. 350-355.
5. *The Diverse Functions of MicroRNAs in Animal Development and Disease*. **Kloosterman, Wigard P. and Plasterk, Ronald H. A.** 2006, *Cell*, Vol. 11, pp. 441-450.
6. *Illuminating the silence: understanding the structure and function of small RNAs*. **Rana, Tariq M.** Gennaio 2007, *Molecular Cell Biology*, Vol. 8, pp. 23-36.
7. *MicroRNAs in Cardiovascular Biology and Heart Disease*. **Catalucci, Daniele, Gallo, Paolo and Condorelli, Gianluigi**. Giugno 19, 2012, *Circulation Cardiovascular Genetics*, pp. 402-408.
8. *Induction of specific micro RNA (miRNA) species by ROS-generating metal sulfates in primary human brain cells*. **Lukiw, Walter J. and Pogue, Aileen I.** 9, 2007, *PMC*, Vol. 101, pp. 1265-1268.
9. *MicroRNAs Induced During Adipogenesis that Accelerate Fat Cell Development Are Downregulated in Obesity*. **Xie, Huangming, Lim, Bing and Lodish, Harvey F.** Maggio 2009, *Diabetes*, Vol. 58, pp. 1050-1057.
10. *Oncomirs-microRNAs with a role in cancer*. **Esquela-Kerscher, Aurora and Slack, J. Frank**. Aprile 2006, *Nature Reviews, Cancer*, Vol. 6, pp. 259-269.
11. **Narry Kim, V. and Nam, Jin-Wu**. *Genomics of microRNA*. *TRENDS in Genetics*. 2006, Vol. 22, 3.
12. *MicroRNAs: Genomics, Biogenesis, Mechanism, and Function*. **Bartel, David P.** 2004, *Cell*, Vol. 116, pp. 281-297.
13. *Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight?* **Filipowicz, Witold, Bhattacharyya, Suwendra N. and Sonenberg, Nahum**. Febbraio 2008, *Nature*, Vol. 9, pp. 102-114.
14. **Ketting, Renè F., et al.** Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *GENES & DEVELOPMENT*. 2001, pp. 15:2654-2659.
15. *Asymmetry in the Assembly of the RNAi Enzyme Complex*. **Schwarz, Dianne S., et al.** Ottobre 17, 2003, *Cell*, Vol. 115, pp. 199-208.
16. *A microRNA Mediates EGF Receptor Signaling and Promotes Photoreceptor Differentiation in the Drosophila Eye*. **Xin, Li and W., Carthew Richard**. Dicembre 29, 2005, *Cell*, Vol. 123, pp. 1267-1277.
17. *Two Genetic Circuits Repress the Caenorhabditis elegans Heterochronic Gene lin-28 after Translation initiation*. **Seggerson, Kathy, Tang, Lingjuan and Moss, Eric G.** 2002, *Developmental Biology*, pp. 215-225.

18. *Determinants of targeting by endogenous and exogenous microRNAs and siRNAs.* **Nielsen, Cydney B., Shomron, Noam and Sandberg, Rickard.** 2007, RNA, Vol. 13, pp. 1-17.
19. *Isolation of microRNA targets by miRNP immunopurification.* **Easow, George, Teleman, Aurelio A. and Cohen, Stephen M.** Giugno 25, 2007, RNA, Vol. 13, pp. 1198-1204.
20. *Regulation by let-7 and lin-4 miRNAs Results in Target mRNA Degradation.* **Bagga, Shveta, et al.** Agosto 26, 2005, Cell, Vol. 122, pp. 553-563.
21. *mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes.* **Behm-Ansmant, Isabelle, et al.** 2006, Genes & Development, Vol. 20, pp. 1885-1898.
22. *Zebrafish MiR-430 Promotes Deadenylation and Clearance of Maternal mRNAs.* **Giraldez, Antonio J., et al.** Aprile 7, 2006, Science, Vol. 312, pp. 75-79.
23. *MicroRNAs direct rapid deadenylation of mRNA.* **Wu, Ligang, Fan, Jihua and Belasco, Joel G.** 11, Marzo 14, 2006, PNAS, Vol. 103, pp. 4034-4039.
24. *Comparison of siRNA-induced off-target RNA and protein effects.* **Alemàn, Lourdes M., Doench, John and Sharp, Philip A.** 2007, RNA, Vol. 13, pp. 385-295.
25. *c-Myc-regulated microRNAs modulate E2F1 expression.* **O'Donnell, Kathryn A., et al.** Giugno 9, 2005, Nature, Vol. 435, pp. 839-843.
26. *Reciprocal actions of REST and a microRNA promote neuronal identity.* **Conaco, Cecilia, Otto, Sefanie, Han, Jong-Jin and Mandel, Gail.** 7, Febbraio 14, 2006, PNAS, Vol. 103, pp. 2422-2427.
27. *The widespread regulation of microRNA biogenesis, function and decay.* **Krol, Jacek, Loedige, Inga and Filipowicz, Witold.** Settembre 2010, Nature Reviews, Genetics, Vol. 11, pp. 597-610.
28. *Phosphorylation of the Human MicroRNA-Generating Complex Mediates MAPK/Erk Signaling.* **Paroo, Zain, et al.** Ottobre 2, 2009, Cell, Vol. 139, pp. 112-122.
29. *Characterizing light-regulated retinal microRNAs reveals rapid turnover as a common property of neuronal microRNAs.* **Krol, Jacek, et al.** Maggio 14, 2012, Cell, Vol. 141, pp. 618-631.
30. *Role for a bidentate ribonuclease in the initiation step of RNA interference.* **Bernstein, E., et al.** 2001, Nature, Vol. 409, pp. 363-366.
31. *Specificity of microRNA target selection in translational repression.* **Doench, John G. and Sharp, Phillip A.** GENES & DEVELOPMENT : s.n., 2004, pp. 504-511.
32. *A combined computation-experimental approach predicts human microRNA targets.* **Kiriakidou, Marianthi, et al.** 2004, GENES & DEVELOPMENT, Vol. 18, pp. 1165-1178.
33. *Fast and effective prediction of microRNA/target duplexes.* **Rehmsmeier, Marc, et al.** 2004, RNA, Vol. 10, pp. 1507-1517.
34. *Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets.* **Lewis, Benjamin P., Gurge, Christopher B. and Bartel David.** 14, 2005, Cell, Vol. 120, pp. 15-20.
35. *Prediction of Mammalian MicroRNA Targets.* **Lewis, Benjamin P., et al.** 26, 2003, Cell, Vol. 115, pp. 787-798.

36. *MicroRNAs: A Role in Plant Development*. **Hake, Sarah**. 10 28, 2003, *Current Biology*, Vol. 13, pp. 851-852.
37. *MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies*. **Liu, Jidong, et al.** 7, 7 2005, *Nat Cell Biology*, Vol. 7, pp. 719-723.
38. *Circulating microRNAs as stable blood-based markers for cancer detection*. **Mitchell, Patrick S., et al.** 30, July 29, 2008, *PNAS*, Vol. 105, pp. 10513-10518.
39. *Circulating microRNA-21 as a novel biomarker for hepatocellular carcinoma*. **Tomimaru, Yoshito, et al.** 1, January 2012, *Journal of Hepatology*, Vol. 56, pp. 167-175.
40. *Circulating MicroRNAs as Biomarkers and Potential Paracrine Mediators of Cardiovascular Disease*. **Gupta, Shashi K., Bang, Claudia and Thum, Thomas**. 2010, *Circ Cardiovasc Genet*, Vol. 3, pp. 484-488.
41. *Circulating microRNA is a biomarker of pediatric Crohn disease*. **Zahm, A.M., et al.** 1, 2011, *J Pediatr Gastroenterol Nutr*, Vol. 53, pp. 26-33.
42. *Exosomes: Endosomal-derived vesicles shipping extracellular messages*. **Février, B. and Raposo, G.** *Curr Opin Cell Biol*, Vol. 16, pp. 415-421.
43. *Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma*. **Arroyo, J.D., et al.** 12, 2011, *Proc Natl Acad Sci USA*, Vol. 108, pp. 5003-5008.
44. *Characterization of extracellular circulating microRNA*. **Turchinovich, A., et al.** 11, 2011, *Nucleic Acids Res*, Vol. 36, pp. 7223-7233.
45. *The crystal structure of the Argonaute2 PAZ domain reveals ad RNA binding motif in RNAi effector complexes*. **Song, J.J., et al.** 2003, *Nat Struct Biol*, Vol. 10, pp. 1026-1032.
46. *Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain*. **Ma, J.B., Ye, K. and Patel, D.J.** 2004, *Nature*, Vol. 429, pp. 318-322.
47. *Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells*. **Valadi, H., et al.** 2007, *Nat Cell Biol*, Vol. 9, pp. 654-659.
48. *Secretory mechanism and intercellular transfer of microRNAs in living cells*. **Kosaka, N., et al.** 2012, *J. Biol. Chem.*, Vol. 285, pp. 17442-17452.
49. **Russo, F, et al.** *miRandola:FErro Lab. miRandola*. [Online] 2012. [Cited: Settembre 24, 2012.] <http://atlas.dmi.unict.it/mirandola/index.html>.
50. *miRò: a miRNA knowledge base*. **Laganà, A., et al.** 2009, *Database (Oxfors)*.
51. *MicroRNAs are transported in plasma and delivered to recipient cells by high-density*. **Vickers, K.C., et al.** 4, 2011, *Nat Cell Biol*, Vol. 13, pp. 423-433.
52. *Exosome released by melanoma cells prepare sentinel lymph nodes for tumor metastasis*. **Hood, J.L., San, R.S. and Wickline, S.A.** 2011, *Cancer Res*, Vol. 71, pp. 3792-3801.
53. *miR-15 and miR-16 induce apoptosis by targeting BCL2*. **Cimmino, Amelia and Calin, George Adrian**. 29, Settembre 27, 2005, *PNAS*, Vol. 102, pp. 13944-13949.

54. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. **Calin, George Adrian, et al.** 9, Marzo 2, 2004, PNAS, Vol. 101, pp. 2999-3004.
55. Characterization of MicroRNA Expression Levels and Their Biological Correlates in Human Cancer Cell Lines. **Gaur, Arti, et al.** Marzo 15, 2007, Cancer Research, pp. 2456-2468.
56. Causes and consequences of microRNA dysregulation in cancer. **Croce, Carlo M.** Ottobre 2009, Nature, pp. 704-714.
57. E2F1-Regulated MicroRNAs Impair TFGb-Dependent Cell-Cycle Arrest and Apoptosis in Gastric Cancer. **Petrocca, Fabio and Visone, Rosa.** Marzo 2008, Cancer Cell, Vol. 13, pp. 272-286.
58. The Expression of MicroRNA miR-107 Decreases Early in Alzheimer's Disease and May Accelerate Disease Progression through Regulation of B-Site Amyloid Precursor Protein-Cleaving Enzyme 1. **Wang, Wang-Xia, et al.** 5, January 30, 2008, J. Neurosci, Vol. 28, pp. 1213-1223.
59. Loss of microRNA cluster miR-29a/b-1 in sporadic Alzheimer's disease correlates with increased BACE1/b-secretase expression. **Hébert, Sébastien, et al.** 17, April 29, 2008, PNAS, Vol. 105, pp. 645-6420.
60. APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. **Rovelet-Lecrux, Anne, et al.** 1, January 2006, Nature Genetics, Vol. 38, pp. 24-26.
61. MicroRNA: Implications for Alzheimer disease and other human CNS disorders. **Maes, O.C., et al.** 2009, Curr Genomics, Vol. 10, pp. 154-168.
62. Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. **Wild, Sarah, et al.** 2004, Diabetes Care, Vol. 27, pp. 1047-1053.
63. Progress in the treatment of type 2 diabetes: new pharmacologic approaches to improve glycemic control. **Cohen, A. and Horton, E.S.** 2007, Curr Med Res Opin, Vol. 23, pp. 905-917.
64. Prediction and prevention of type 1 diabetes: progress, problems, and prospects. **J.S., Skyler.** 2007, Clin Pharmacol Ther, Vol. 81, pp. 768-771.
65. Epidemiology of diabetes. **Winer, N. and Sowers, J.R.** 2004, J Clin Pharmacol, Vol. 44, pp. 397-405.
66. A pancreatic islet-specific microRNA regulates insulin secretion. **Poy, M.N., et al.** 2004, Nature, Vol. 432, pp. 226-230.
67. Targeted inhibition of miRNA maturation with morpholinos reveals a role for miR-375 in pancreatic islet development. **Kloosterman, W.P., et al.** 2007, PLoS Biol, Vol. 5, p. e203.
68. MicroRNA-124a regulates Foxa2 expression and intracellular signaling in pancreatic beta-cell lines. **Baroukh, N., et al.** 2007, J Biol Chem, Vol. 282, pp. 19575-19588.
69. microRNAs and the regulation of glucose and lipid metabolism. **Poy, M.N., Spranger, M. and Stoffel, M.** 2007, Diabetes Obes Metab, Vol. 9, pp. 67-73.
70. MicroRNA miR-133 represses HERG K⁺ channel expression contributing to QT prolongation in diabetic hearts. **Xiao, J., et al.** 2007, J Biol Chem, Vol. 282, pp. 12363-12367.

71. *MicroRNAs in Diabetes: Tiny Players in Big Disease*. **Oandey, A.K., et al.** March 20, 2009, *Cellular Physiology and Biochemistry*, Vol. 23, pp. 221-232.
72. *Epidemiology of multiple sclerosis in US veterans: V. Ancestry and the risk of multiple sclerosis*. **Page, W.F., et al.** 1993, *Ann Neurol*, Vol. 33, pp. 632-639.
73. *The epidemiology of multiple sclerosis: a general overview*. **Poser, C.M.** 1994, *Ann Neurol*, Vol. 36, pp. S180-S193.
74. *Natural history of multiple sclerosis*. **Weinshenker, B.G.** 1994, *Ann Neurol*, Vol. 36, pp. S6-S11.
75. *The cellular immunology of multiple sclerosis*. **Al-Omaishi, J., Bashir, R. and Gendelman, H.E.** 1999, *J Leukoc Biol*, Vol. 65, pp. 444-452.
76. *Multiple Sclerosis*. **Compston, A. and Coles, A.** 2008, *Lancet*, Vol. 372, pp. 1502-1517.
77. *Period prevalence of multiple sclerosis in England: relationship to UV exposure*. **Ramagopalan, S.V., et al.** 2011, *Neurology*, Vol. 76, pp. 1410-1414.
78. *MicroRNA miR-326 regulates TH-17 differentiation and is associated with the pathogenesis of multiple sclerosis*. **Du, C., et al.** 2009, *Nat Immunol*, Vol. 10.
79. *Multiple sclerosis: microRNA expression profiles accurately differentiate patients with relapsing-remitting disease from healthy controls*. **Keller, A., et al.** 2009, *PLoS One*, Vol. 4, p. e7440.
80. *Circulating RNA as a tumor marker: detection of 574 mRNA in breast and lung cancer patient serum*. **Kopreski, M.S. and Benko, F., Gocke, D.C.** 2001, *Ann N Y Acad Sci*, Vol. 945, pp. 172-178.
81. *Circulating cell-free nucleic acids—a promising non-invasive tool for early detection of several human diseases*. **Swarup, V. and Rajeswari, R.M.** 2007, *FEBS Lett*, Vol. 581, pp. 795-799.
82. *Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases*. **Chen, X., et al.** 2008, *Cell Res*, Vol. 18, pp. 997-1006.
83. *MicroRNAs as therapeutic targets*. **Czech, M.P.** 2006, *The New Eng J Med*, Vol. 354, pp. 1194-1195.
84. *Exploiting and antagonizing microRNA regulation for therapeutic and experimental applications*. **Brown, B.D. and N., Naldini.** 2009, *Nat Rev Genet*, Vol. 10, pp. 578-585.
85. *Antagomir-17-5p abolishes the growth of therapy-resistant neuroblastoma through p21 and BIM*. **Fontana, L., et al.** 2008, *PLoS One*, Vol. 3, p. e2236.
86. *What is a gene, post-ENCODE? History and updated definition*. **Gerstein, M.B., et al.** 6, 2007, *Genome Res*, Vol. 17, pp. 669-681.
87. *Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid*. **Watson, P.J. and Crick, F.H.** 4356, 1953, *Nature*, Vol. 171, pp. 737-738.
88. *Guidelines for human genome nomenclature*. **Wain, H., et al.** 4, 2002, *Genomics*, Vol. 79, pp. 464-470.
89. **Campbell, N.A. and Reece, J.B.** *Biology, 8th ed.* s.l. : Pearson Education. 0-321-54424-2.
90. *Comparative genomics of the eukaryotase*. **Rubin, G.M., Yandell, M.D. and Wortman, J.R. et al.** 5461, 2000, *Science*, Vol. 287, pp. 2204-2215.

91. *Gene ontology: tool for the unification of biology.* **Ashburner, M., et al.** 1, 2000, *Nature Genetics*, Vol. 25, pp. 25-29.
92. *Automatic extraction of gene ontology annotation and its correlation with clusters in protein networks.* **Dareselia, N., et al.** 1, 2007, *BMC Bioinformatics*, Vol. 8, p. 243.
93. *Comparison of functional annotation schemes for genomes.* **Rison, S.C., Hodgman, T.C. and J.M., Thorton.** 1, 2000, *Functional & Integrative Genomics*, Vol. 1, pp. 56-69.
94. *Phenotype Data: A Neglected Resource in Biomedical Research?* **Groth, P. and Weiss, B.** 3, 2006, *Current Bioinformatics*, Vol. 1, pp. 347-358.
95. *Mpuse Phenome Project: understanding human biology through mouse genetics and genomics.* **Bogue, M.** 4, 2003, *J Appl Physiol*, Vol. 95, pp. 1335-1337.
96. *The Mouse Phenome Project.* **Bogue, M.A. and Grubb, S.C.** 1, 2004, *Genetica*, Vol. 122, pp. 71-74.
97. *EMPreSS: standardized phenotype screens for funzional annotations of the mouse genome.* **Brown, S.D., Chambon, P. and de Angelis, M.H.** 37, 2005, *Nat Genet*, Vol. 2005, p. 1155.
98. *Major European Project to create nre Knowledgebase of Gene-Disease Relantionships, The GEN2PHEN Consortium.* **Brookes, A.J.**
99. *The human phenom project.* **Freimer, N. and Sabatti, C.** 1, 2003, *Nat Genet*, Vol. 34, pp. 15-21.
100. *The Mammalian Phenotype Ontology as a tool for annotating, analyzing and comparing phenotypic information.* **Smith, C.L., Goldsmith, C.A. and Epping, J.T.** 1, 2005, *Genome Biol*, Vol. 6, p. R7.
101. *MGI-Statistics for the Mouse Genome Informatics database resource.* **MGI.** 2008, The Jackson Laboratory.
102. *The Human Phenotype Ontology: a tool for annotating and anlyzing human hereditary disease.* **Robinson, P.N., et al.** 5, 2008, *Am J Hum Genet*, Vol. 83, pp. 610-615.
103. *Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease.* **Botstein, D. and Risch, N.** 2003, *Nat Genet*, Vol. 33, pp. 228-237.
104. *Reduced sleep in Drosophila Shaker mutants.* **Cirelli, C., et al.** 7037, 2005, *Nature*, Vol. 434, pp. 1087-1092.
105. *Different clinical manifestations of hyperphenylalaninemia in three siblings with identical phenylalanine hydroxylase genes.* **DiSilvestre, D., Koch, R. and Groffen, J.** 5, 1991, *Am J Hum Genet*, Vol. 48, pp. 1014-1016.
106. *Monogenic traits are not simple: lessons from phenylketonuria.* **Scriver, C.R. and Waters, P.J.** 7, 1999, *Trends Genet*, Vol. 15, pp. 267-272.
107. *PAHdb 2003: what a locus-specific knowledge-base can do.* **Scriver, C.R., et al.** 4, 2003, *Hum Mutat*, Vol. 21, pp. 333-344.
108. *PAHdb: a locus-specific knowledgebase.* **Scriver, C.R., et al.** 1, 2000, *Hum Mutat*, Vol. 15, pp. 99-104.

109. *How PAH gene mutations cause hyper-phenylalaninemia and why mechanism matters: insights from in vitro expression.* **Waters, P.J.** 4, 2003, *Hum Mutat*, Vol. 21, pp. 357-369.
110. *After the genome-the phenome?* **Scriver, C.R.** 3, 2004, *Journal of Inherited Metabolic Disease*, Vol. 27, pp. 305-317.
111. *From genotype to phenotype: genetics and medical practice in the new millenium.* **Weatherall, D.** 1392, 1999, *Philos Trans R Soc Lond B Biol Sci*, Vol. 354, pp. 1995-2010.
112. *Large-scale mutagenesis of the mouse to understand the genetic bases of nervous system structure adn function.* **Goldowitz, D., et al.** 2, 2004, *Brain Res Mol Brain Res*, Vol. 132, pp. 105-115.
113. *Genome-wide RNAi as a route to gene function in Drosophila.* **Kuffenkeuler, D. and Boutros, M.** 2, 2004, *Brief Funct Genomics Proteomic*, Vol. 3, pp. 168-176.
114. *The art and design of genetic screens: Arabidopsis thaliana.* **Page, D.R. and Grossniklaus, U.** 2, 2002, *Nat Rev Genet*, Vol. 3, pp. 124-136.
115. *Forward genetics and map-based cloning approaches.* **Peters, J.L., Cnudde, F. and Gerats, T.** 10, 2003, *Trends Plant Sci*, Vol. 8, pp. 484-491.
116. *Systematic association of genes to phenotypes by genome and literature mining.* **Korbel, JO, Doerks, T. and Jensen, L.J. et al.** 2005, *PLoS Biol*, Vol. 3, p. e134.
117. *G2D: A Tool for Mining Genes Associated with Disease.* **Perez-Iratxeta, C., et al.** 2005, *BMC Genet*, Vol. 6, p. 45.
118. *The HUGO Gene Nomenclature Committee (HGNC).* **Povey, S., Lovering, R., et al.** 2001, *Hum Genet*, Vol. 109, pp. 678-680.
119. *Gene annotation from scientific literature using mappings between keyword systems.* **Perez, A.J., et al.** 2004, *Bioinformatics*, Vol. 20, pp. 2084-2091.
120. *Recognizing names in biomedical texts: a machine learning approach.* **Zhou, G., et al.** 2004, *Bioinformatics*, Vol. 20, pp. 1178-1190.
121. *Term identification in the biomedical literature.* **Krauthammer, M. and Nenadic, G.** 2004, *J Biomed Inform*, Vol. 37, pp. 512-526.
122. *An evaluation of GO annotation retrieval for BioCreAtIvE and GOA.* **Camon, E.B., Barrell, D.G. and Dimmer, E.C. et al.** 2005, *BMC Bioinformatics*, Vol. 6, p. S17.
123. *Evaluation of BioCreAtIvE assessment of task 2.* **Blaschke, C., et al.** 2005, *BMC Bioinformatics*, Vol. 6, p. S16.
124. *Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders.* **Hamosh, A., et al.** 2005, *Nucleic Acids Res*, Vol. 33, pp. D514-517.
125. *Database resources of the National Center for Biotechnology Information.* **Wheeler, D.L., Barrett, T. and Benson, D.A et al.** 2005, *Nucleic Acids Res*, Vol. 33, pp. D39-45.
126. *Genatlas database, genes and development defects.* **Frezal, J.** 1998, *C R Acad Sci III*, Vol. 321, pp. 805-817.

127. *Human genes involved in chromatin remodeling in transcription initiation, and associated disease: An overview using the GENATLAS database.* **Roux-Rouquie, M., et al.** 1999, *Mol Genet Metab*, Vol. 67, pp. 261-277.
128. *The human gene mutation database.* **Cooper, D.N., Ball, E.V. and Krawczak, M.** 1998, *Nucleic Acids Res*, Vol. 26, pp. 285-287.
129. *Human Gene Mutation Database (HGMD): 2003 update.* **Stenson, P.D., Ball, E.V. and Mort, M. et al.** 2003, *Hum Mutat*, Vol. 21, pp. 577-581.
130. *Human gene mutation database-a biomedical information and research resource.* **Krawczak, M., Ball, E.V. and Fenton, I. et al.** 2000, *Hum Mutat*, Vol. 15, pp. 45-51.
131. *PharmGKB: The Pharmacogenetics and Pharmacogenomics Knowledge Base.* **C.F., Thornm, Klein, T.E. and Altman, R.B.** 2005, *Methods Mol Biol*, Vol. 3111, pp. 179-192.
132. *PharmaGKB: the Pharmacogenetics Knowledge Base.* **Hewett, M., Oliver, D.E. and Rubin, D.L. et al.** 2002, *Nucleic Acids Res*, Vol. 30, pp. 163-165.
133. *Integrating genotype and phenotype information: an overview of the PharmGKB project. Pharmacogenetics Research Network and Knowledge Base.* **Klein, T.E., Chang, J.T. and Cho, M.K.** 2001, *Pharmacogenomics J*, Vol. 1, pp. 167-170.
134. *PharmGKB: the pharmacogenetics and pharmacogenomics knowledge base.* **Klein, T.E. and Altman, R.B.** 2004, *Pharmacogenomics J.*, Vol. 4, p. 1.
135. *PhenoGO: an integrated resource for the multiscale mining of clinical and biological data.* **Sam, L.T., et al.** 2008, *BMC Bioinformatics*, Vol. 10.
136. *PhenomicDB: a new cross-species genotype/phenotype resource.* **Gorth, P., et al.** 2006, *Nucleic Acids Research*, Vol. 35, pp. D696-699.
137. *The Mouse Genome Database (MGD): updates and enhancements.* **Blake, J.A., et al.** 2006, *Nucleic Acids Res*, Vol. 34, pp. D562-567.
138. *WormBase: better software, richer content.* **Schwarz, E.M., et al.** 2006, *Nucleic Acid Res*, Vol. 34, pp. D475-478.
139. *FlyBase: anatomical data, images and queries.* **Grumbling, G. and Strelets, V.** 2006, *Nucleic Acids Res*, Vol. 34, pp. D484-488.
140. *CYGD: the comprehensive yeast genome database.* **Guldener, U., et al.** 2005, *Nucleic Acids Res*, Vol. 33, pp. D364-368.
141. *The Zebrafish Information Network: the zebrafish model organism database.* **Sprague, J., et al.** 2006, *Nuclei Acids Res*, Vol. 34, pp. D581-585.
142. *MIPS Arabidopsis thaliana Database (MAiSB): an integrated biological knowledge resources for plant genomics.* **Schoof, H., et al.** 2004, *Nucleic Acids Res*, Vol. 32, pp. D373-376.
143. *Database resources of the National Center for Biotechnology Information.* **Wheeler, D.L., et al.** 2006, *Nucleic Acids Res*, Vol. 34, pp. D173-180.
144. *Phenoclustering: online mining of cross-species phenotypes.* **Groth, P., et al.** 15, 2010, *Bioinformatics*, Vol. 26, pp. 1924-1925.

145. *Mining phenotypes for gene function prediction*. **Groth, P. et al.** 2008, BMC Bioinformatics, Vol. 9, p. 136.
146. *Data clustering in life sciences*. **Zhao, Y. and Karypis, G.** 1, 2005, Molecular Biotechnology, Vol. 31, pp. 55-80.
147. **Toolkit, CLUTO: Clustering**. [Online] <http://glaros.dtc.umn.edu/gkhome/views/cluto>.
148. *A Comparison of Document Clustering Techniques*. **Steinback, M., Karypis, G. and Kumar, V.** KDD Workshop on Text Mining 2000.
149. *A computational system to select candidate genes for complex human traits*. **Gaulton, K.J., Mohlke, K.L. and Vision, T.J.** 9, 2007, Bioinformatics, Vol. 23, pp. 1132-1140.
150. *The Gene Ontology (GO) database and informatics resource*. **Harris, M, et al.** 2004, Nucleic Acids Res, Vol. 32, pp. D258-261.
151. *eVOC: a controlled vocabulary for unifying gene expression data*. **Kelso, J. et al.** 2003, Genome Res, Vol. 13, pp. 1222-1230.
152. *A text-mining analysis of the human phenome*. **van Driel, M., et al.** 2006, European Journal of Human Genetics, Vol. 14, pp. 535-542.
153. *The druggable genome*. **Hopkins, A.L. and Groom, C.R.** 9, 2002, Nat Rev Drug Discov, Vol. 1, pp. 727-730.
154. *Network pharmacology*. **Hopkins, A.L.** 10, 2007, Nat Biotechnol, Vol. 25, pp. 1110-1111.
155. *The efficiency of multi-target drugs: the network approach might help drug design*. **Csermely, P., Agoston, V. and Pongor, S.** 4, 2005, Trends Pharmacol Sci, Vol. 26, pp. 178-182.
156. *Drug discovery: playing dirty*. **Frantz, S.** 7061, 2005, Nature, Vol. 437, pp. 942-943.
157. *Systems biology in drug discovery*. **Kunkel, E.J.** 2006, Conf Proc IEEE Eng Med Biol Soc, Vol. 1, p. 37.
158. *Drug-target network*. **Yildirim, M.A., et al.** 2007, Nat Biotechnol, Vol. 25, pp. 1119-1126.
159. *DrugBank 3.0: a comprehensive resource for 'omics' research on drugs*. **Knox, C., et al.** 2011, Nucleic Acids Res, Vol. 39, pp. D1035-1041.
160. *Illuminating drug discovery with biological pathways*. **Apic, G.** 8, 2005, FEBS Lett, Vol. 579, pp. 1872-1877.
161. *Drug target identification using side-effect similarity*. **Campillos, M., et al.** 2008, Science, Vol. 321, pp. 263-266.
162. *Network neighbors of drug targets contribute to drug side-effect similarity*. **Brouwers, L., Iskar, M. and Zeller, G. et al.** 2011, PLoS One, Vol. 6, p. e22187.
163. *The relationship between rational drug design and drug side effects*. **Wang, J., et al.** 2011, Briefing in Bioinformatics.
164. *Network medicine: a network-based approach to human disease*. **Barabasi, A.L., Gulbahce, N. and Loscalzo, J.** 2011 : s.n., Nat Rev Genet, Vol. 12, pp. 56-68.

165. A side effect resource to capture phenotypic effects of drugs. **Kuhn, M., et al.** 6, 2010, *Molecular Systems Biology*, Vol. 343.
166. Network analysis of FDA approved drugs and thier targets. **Ma'ayan, A., et al.** 2007, *Mt Sinai J Med*, Vol. 74, pp. 27-32.
167. Identification of information flow-modulating drug targets: a novel bridging paradigm for drug discovery. **Hwang, W.C., Zhang, A. and Ramanathan, M.** 2008, *Clin Pharmacol Ther*, p. 129.
168. The signaling petrinet-based simulator: a non-parametric strategy for characterizing the dynamics of cell-specific signaling networks. **Ruths, D., et al.** 2008, *PLoS Comput Biolo*, Vol. 4, p. e1000005.
169. Lethality and centrality in protein networks. **Jeong, H., et al.** 2001, *Nature*, Vol. 411, p. 41.
170. Identification and characterization of the tuberous sclerosis gene on chromosome 16. **Consortium, The European Chromosome 16 Tuberous Sclerosis.** 1993, *Cell*, Vol. 75, pp. 1305-1315.
171. Identification of the tuberous sclerosis gene TSC1 on chromosome 9q34. **van Slechtenhorst, M., et al.** 1997, *Science*, Vol. 277, pp. 805-808.
172. Phenotypes of the tuberous sclerosis complex with a revision of diagnostic criteria. **Gomez, M.R.** 1991, *Ann. N. Y. Acad Sci*, Vol. 615, pp. 1-7.
173. Interaction between hamartin and tuberin, the TSC1 and TSC2 gene products. **van Slengtenhorst, M., et al.** 1998, *Hum Mol Genet*, Vol. 7, pp. 1053-1057.
174. **Macca, V., et al.** Analysis of TSC pathway by a new miRNA-drug-phenotype tool. Settembre 6-9, 2012. Oral Communication, International TSC Congress, Progetto Superare 2012.
175. Slerosi Tuberosa. **Ruggieri, M. and Migone, N.** 2, 2009, *Supplemento AESSETI NEWS*.
176. microRNA-21 Governs TORC1 Activation in Renal Cancer Cell Proliferation and Invasion. **Dey, N., et al.** 6, 2012, *PLoS One*, Vol. 7, p. e37366.
177. miR-15a and miR-16 affect the angiogenesi of multiple myeloma by targeting VEGF. **Sun, C.y., et al.** 2012, *Carciogenesis*.
178. Human miR-1271 is a miR-96 paralog with distinct non-conserved brain expression pattern. **Jensen, K.P. and Covault, J.** 2, 2010, *Nucleic Acide Research*, Vol. 39, pp. 701-711.
179. Recent advances in neurobiology of Tuberous Sclerosis Complex. **Napolioni, V., Moavero, R. and Curatolo, P.** 3009, *Brain & Development*, pp. 104-113.