



UNIVERSITÀ  
DEGLI STUDI  
DI TORINO



# Dipartimento di Scienze Mediche

*Il sequenziamento di nuova generazione e le tecnologie “omics”:  
dalla ricerca di base alla clinica*

**prof. Giuseppe MATULLO**

*DSM-UniTo and HuGeF*

*<http://www.dsm.unito.it>*



UNIVERSITÀ  
DEGLI STUDI  
DI TORINO



ISTITUTO NAZIONALE DI FISICA NUCLEARE



— HUMAN GENETICS FOUNDATION · TORINO —



# Dipartimento di Scienze Mediche

*Direttore: prof Franco Veglio*



## Le finalità del Dipartimento sono:

1. riunire in un unico Dipartimento gruppi di **ricerca clinica, pre-clinica e di base** in grado di condurre studi di eccellenza nel campo della diagnosi, della patogenesi, della fisiopatologia, delle basi ereditarie e della terapia di malattie acute e croniche, d'organo e sistemiche;
2. **sviluppare collaborazioni ed interazioni fra gruppi con competenze diverse**, in modo da facilitare la trasferibilità delle osservazioni derivate dalla ricerca di base alla clinica con la finalità ultima di offrire approcci diagnostici e terapeutici d'avanguardia;
3. promuovere un circuito di collaborazioni fra i diversi gruppi di ricerca clinica e di base al fine di sviluppare prodotti e tecnologie idonee allo sfruttamento brevettuale o applicato;

4. indurre sinergie fra gruppi con competenze complementari focalizzate alla **stesura di progetti di ricerca biomedica da proporre per il finanziamento a Istituzioni nazionali ed internazionali**, a Enti e Fondazioni, ed Organizzazioni non Lucrative di Utilità Sociale (ONLUS);
5. **facilitare la condivisione di tecnologie e strumentazioni complesse e la preparazione di protocolli di test farmacologici e diagnostici** di potenziale interesse per il mondo industriale, compresi i trials di intervento farmacologico;
6. generare un **collegamento diretto tra ricerca di base e ricerca clinica** in modo da fornire esperienze utili a realizzare un'offerta didattica per gli studenti del corso di Laurea in Medicina e Chirurgia, delle Lauree triennali e specialistiche in Biologia e Biotecnologie, delle Lauree Sanitarie, della Scuola Superiore di Medicina, delle Scuole di Specializzazione e delle Scuole di Dottorato di Ricerca. Tale collegamento rappresenta il contesto ideale per il percorso formativo di ricercatori e clinici con elevata professionalità in un ampio spettro di discipline delle Scienze Mediche.



# Dipartimento di Scienze Mediche

## FILONI DI RICERCA DSM



**Ricerca traslazionale e clinica:** studio dei meccanismi eziopatogenetici e fisiopatologici alla base di malattie umane frequenti e rare, marcatori diagnostici e approccio terapeutico:

aritmie cardiache

asma e broncopatie croniche

aterosclerosi

diabete e malattie del ricambio (retinopatia e nefropatia diabetica, rischio cardio-vascolare e modelli di self-management education)

ipertensione (essenziale, secondaria, eredo-familiare)

malattie dermatologiche: melanomi, linfomi primitivi della cute a cellule T

malattie endocrine e del metabolismo

malattie epatiche (epatopatie virali, metaboliche e steatoepatite, cirrosi)

malattie del tratto gastroenterico (gastrite, ulcera, malattie infiammatorie intestinali)

malattie **genetiche e genomiche** (patologie neurodegenerative e del neurosviluppo)

malattie immunitarie, autoimmuni, vasculiti e sindromi da attivazione mastocitaria

malattie infettive, farmacocinetica degli anti infettivi

malattie metaboliche dell'osso

medicina d'emergenza /urgenza

nefropatie

tumori e predisposizioni ereditarie allo sviluppo di tumori: endocrini, mammari, neurologici, della prostata, del tratto gastroenterico e del fegato, sindromi linfoproliferative e mieloma multiplo

benessere neuro-muscolare ed esercizio fisico





# Dipartimento di Scienze Mediche

## FILONI DI RICERCA DSM



### **Epidemiologia** e studio dei **fattori di rischio:**

tumori  
diabete  
epatiti virali  
malattie infiammatorie intestinali  
malattie multifattoriali (interazione geni/ambiente)  
epigenetica in ambito oncologico e cardiovascolare

### **Trapianti e medicina rigenerativa:**

istopatologia e genetica dei trapianti  
cellule staminali e angiogenesi  
ingegneria tissutale mediante utilizzo di onde d'urto ad alta energia e cellule staminali  
fattori di rischio per lo sviluppo di tumori cutanei non melanoma in pazienti immunosoppressi e trapiantati  
metodi fisici e medical devices per la medicina rigenerativa

### **Diagnostica ed imaging:**

medicina nucleare  
imaging digitale in istopatologia e sviluppo e validazione di tecniche di conservazione dei tessuti umani.  
sviluppo e validazione di biomarcatori e sviluppo di diagnostica molecolare





# Dipartimento di Scienze Mediche

## DSM BIG DATA



### SPAZIO DI STORAGE

- esomi
- genomi (WGS)
- epigenomica (DNA methylation arrays)
- proteomica
- metabolomica
- RNA sequencing
- miRNA sequencing

### TEMPO DI CALCOLO PER ANALISI E SIMULAZIONI

- analisi dei dati «omics»
- imputazione di genotipi da dati genomici
- modelli di simulazione dinamici
- simulazioni ad alta intensità computazionale
- modelli statistici ad alta intensità computazionale
- integrazione di dati di farmacogenetica



MiSeq

NextSeq500

HiScanSQ

**GENOMICS**  
**EPIGENOMICS**  
**TRANSCRIPTOMICS**  
**(RNAseq, miRNAseq)**



**Orbitrap Fusion**  
 Tribrid Mass Spectrometer

Unmatched Analytical  
 Performance with  
 Unprecedented Ease of Use



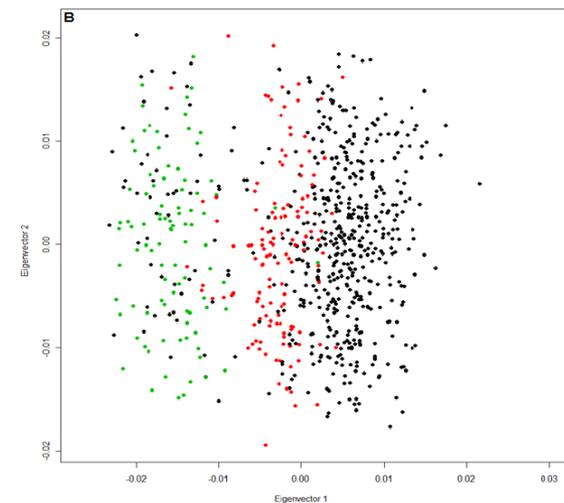
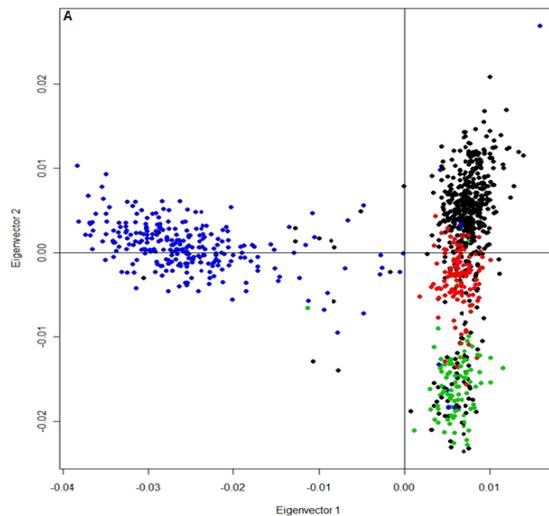
The Thermo Scientific™ Orbitrap Fusion™ Tribrid™ mass spectrometer combines the best of quadrupole, Orbitrap, and ion trap mass analysis in a revolutionary Tribrid architecture that delivers unprecedented depth of analysis. It enables life scientists analyzing even the most challenging low-abundance, high-complexity, or difficult samples to identify more compounds more quickly, quantify more accurately, and elucidate structures more thoroughly.

The Orbitrap Fusion system's next-generation hardware and software make setup easier, methods more powerful, and operation more robust. Users can spend more time thinking about their research and less time worrying about method development and experimental execution.

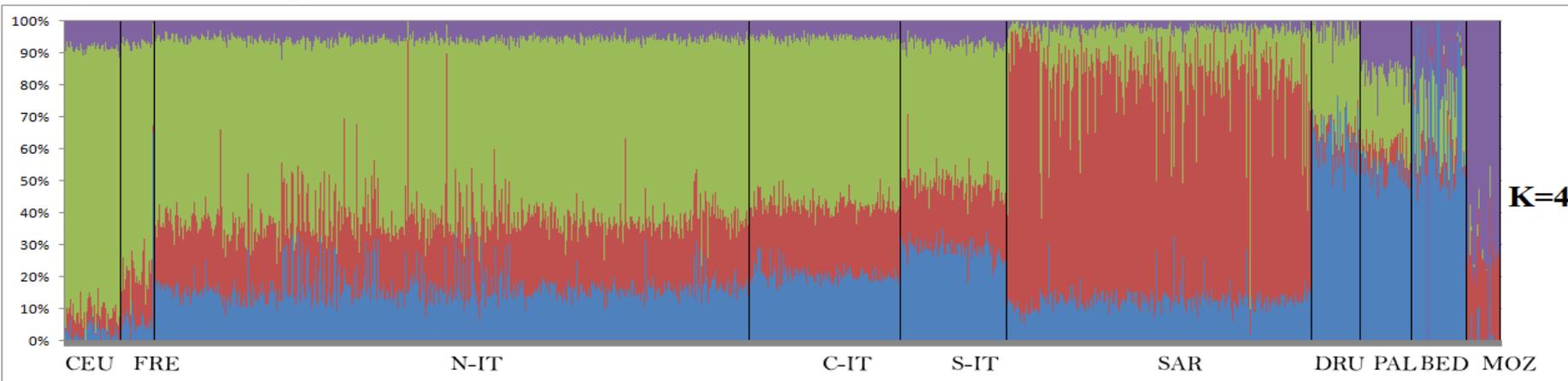
**PROTEOMICS**  
**METABOLOMICS**

## An Overview of the Genetic Structure within the Italian Population from Genome-Wide Data

1014 subjects 125.799 SNPs



Correlation between geographical distance and genetic distance (IBS)  $R^2 = 0.41$ ,  $p < 2.2 \cdot 10^{-6}$



## NETWORK FOR ITALIAN GENOMES (NIG)

Accordo di partenariato firmato dai rettori di diverse università italiane e, per delega del rettore UniTo, dal direttore del DSM

*Referente UniTo: prof. Matullo*

- **Giorgio Casari**, San Raffaele, Milano
- **Alessandro Desideri**, U. Tor Vergata, Roma
- **Paolo Gasparini**, Univ. Trieste
- **Giuseppe Matullo**, Univ. Torino
- **Vincenzo Nigro**, Univ. Napoli
- **Alessandra Renieri**, Univ. Siena
- **Marco Seri**, Univ. Bologna
- **Marco Tartaglia**, Osp. Bambin Gesù, Roma
- **Daniela Toniolo**, San Raffaele, Milano
- **Orsetta Zuffardi**, Univ. Pavia

- In Italy several large studies are currently performing **exome** or **whole genome sequencing** for the identification of pathological mutations and population-specific **rare variants** able to explain a larger part of the genetic variability of **traits and complex diseases**, useful also for **pharmacogenomics and personalized medicine** issues.
- The future goal is to create a true **Italian Network of Genomic data**, such as those existing in other European countries (e.g., GoNL and UK10K), maximizing the utility of the data that are produced for purposes of **public health** and opening up new possibilities for **collaborations and international funding**.
- The most important Italian sequencing centres have been already contacted and in the year 2015 a collaborative effort will be carried out to create a true Italian Network of Genomic data including partners as **Cineca and Telethon**.

ESOMI → 3000

WHOLE GENOMES → 1300

Search variants in GEUVADIS European Exome Variant Server

e.g. chr1:10000-1000000, ATM, ENSG00000187634

**Advanced Options**

Variant type:   Gene Definition:

Exonic Function:  Variant Function:

Subgroup:  Allele Frequency:

Search variants in GEUVADIS European Exome Variant Server

e.g. chr1:10000-1000000, ATM, ENSG00000187634

**Advanced Options**

Variant type:  Gene Definition:

Exonic Function:  Variant Function:

Subgroup:  Allele Frequency:

Search variants in GEUVADIS European Exome Variant Server

e.g. chr1:10000-1000000, ATM, ENSG00000187634

**Advanced Options**

Variant type:  Gene Definition:

Exonic Function:  Variant Function:

Subgroup:  Allele Frequency:

- All
- exonic
- splicing
- intronic
- UTR3
- UTR5
- downstream
- upstream
- intergenic
- ncRNA\_exonic
- ncRNA\_splicing
- ncRNA\_intronic
- ncRNA\_UTR3
- ncRNA\_UTR5

Search variants in GEUVADIS European Exome Variant Server

e.g. chr1:10000-1000000, ATM, ENSG00000187634

**Advanced Options**

Variant type:  Gene Definition:

Exonic Function:  Variant Function:

Subgroup:  Allele Frequency:

- All
- nonsynonymous SNV
- synonymous SNV
- stopgain SNV
- stoploss SNV
- frameshift insertion
- frameshift deletion
- frameshift substitution
- nonframeshift insertion
- nonframeshift deletion
- nonframeshift substitution

Gene:  Variant type:  Gene definition:  Exonic Function:  Variant Function:  Allele Frequency:

**Showing Results for brca1**

Show  entries

Chr	Position	Reference	Alteration	Allele Freq	Allele Count	Allele Number	DP	Gene	Transcript	AAC	Variant Function	Exonic Function	PhyloP (Pm)
17	41245471	C	T	0.06580386	361	5486	90	BRCA1	NM_007297	D646N	exonic	nonsynonymous SNV	0.125
17	41243509	T	C	0.00510204	28	5488	97	BRCA1	NM_007297	R1300G	exonic	nonsynonymous SNV	0.439
17	41245932	G	A	0.00018222	1	5488	65	BRCA1	NM_007297	T492H	exonic	nonsynonymous SNV	-0.412
17	41244429	C	T	0.06067784	333	5488	108	BRCA1	NM_007297	S993N	exonic	nonsynonymous SNV	0.796
17	41223094	T	C	0.26840380	1473	5488	113	BRCA1	NM_007297	S1566G	exonic	nonsynonymous SNV	0.128
17	41234445	G	T	0.05890266	378	5486	93	BRCA1	NM_007297	P1398T	exonic	nonsynonymous SNV	2.752
17	41244252	G	A	0.06905977	379	5488	98	BRCA1	NM_007297	P1052L	exonic	nonsynonymous SNV	1.402
17	41244936	G	A	0.28607872	1570	5488	90	BRCA1	NM_007297	P824L	exonic	nonsynonymous SNV	0.938



**Dipartimento di Scienze Mediche**



## **UNITA': VARIABILITA' GENOMICA DELLE POPOLAZIONI UMANE E MALATTIE COMPLESSE**

- 1800 SNP arrays ( $10^6$  SNPs)
- 1750 ESOMI ITALIANI
- 100 GENOMI COMPLETI
- 1400 DNA methylation arrays (450k CpGs)
- 400 gene expression arrays
- 100 miRNA seq NGS
- EPIC-Torino biobanca: 50,000 volontari





# Epidemiologia dei Tumori CPO, Laboratorio di Epidemiologia Molecolare, Epidemiologia Clinica

- Analisi su dati –omics in studi di coorte di grandi gruppi di popolazione (es: EPIC, Piccolipiù, NINFEA)
- Analisi statistiche ad elevata complessità computazionale su coorti di grandi dimensioni e su archivi di dati sanitari di grandi dimensioni (es: analisi di mediazione, modelli di crescita, analisi multilivello)
- Studi di simulazione per la valutazione di modelli in epidemiologia
- Modelli dinamici per la per la valutazione dell’impatto di vaccinazione e screening per la prevenzione del tumore della cervice uterina

# INTEGROMICS



**RICERCA DI  
BASE E  
TRASLAZIONALE**



- esomi
- genomi (WGS)
- epigenomica
- proteomica
- metabolomica
- RNA sequencing
- miRNA sequencing
- altri biomarcatori
- fattori di rischio
- stile di vita
- parametri clinici
- fenotipi



- **PREVENZIONE**
- **CLINICA**
- **DIAGNOSI PRECOCE**
- **PROGNOSI**
- **SOPRAVVIVENZA**
- **TERAPIA  
PERSONALIZZATA**



# Dipartimento di Scienze Mediche

## RICHIESTE C<sup>3</sup>S



### STORAGE DATI NGS E TEMPI DI CALCOLO

#### ESOMA

Dati di esoma 100x  $\approx$  **10 GB**

Spazio richiesto per l'analisi è **2-3 volte quello dei dati grezzi**

Per **1750 esomi a 30x**, estrapolando si ottiene una dimensione totale per i **dati grezzi di 6-7TB** e per **l'analisi di 12-20TB**. Solo le varianti richiedono  **$\sim$ 8TB di spazio**.

Per campioni tumorali la copertura media necessaria per trovare varianti somatiche cresce aumentando il coverage

#### GENOMA

Dati di genomi **5x  $\approx$  10GB** per i dati grezzi

Spazio richiesto per l'analisi: **3-4 volte** tanto quello dei dati grezzi

Per 100 campioni ci vorrà circa **1TB di stoccaggio e 4TB** per l'analisi base.

#### TEMPI DI CALCOLO

Allineare un genoma **5x richiede circa 24 ore**. Essendo l'allineamento altamente parallelizzabile, utilizzando il cluster OCCAM a piena capacità (1100 cores), possiamo allineare 100 genomi 5x in meno di tre ore. La chiamata delle varianti è più difficilmente parallelizzabile. La chiamata single sample un genoma a 5x richiede circa 18 ore.



# LAST BUT NOT LEAST !

- **AGREEMENT PER CONDIVISIONE DATI**
- **INSERIMENTO C<sup>3</sup>S IN RICHIESTE DI FONDI NAZIONALI E INTERNAZIONALI**