

I QUADERNI DI SAGEN 2

# SALUTE E AMBIENTE

ALLA LUCE DELLE INNOVAZIONI TECNOLOGICHE  
E SCIENTIFICHE

*Sintesi del workshop  
Roma, 8 maggio 2019*

FRANCISCI EDITORE



ASSOCIAZIONE SALUTE, AMBIENTE, GENOMA - SAGEN  
in collaborazione con il  
CENTRO DOCUMENTAZIONE E STUDI  
DEI COMUNI ITALIANI ANCI-IFEL

# SALUTE E AMBIENTE

*Alla luce delle innovazioni tecnologiche e scientifiche*

*Sintesi del workshop  
Roma, 8 maggio 2019*

ALDO FRANCISCI EDITORE

## SOMMARIO

- 3 **PREFAZIONE - Cesare Pillon**
- 5 **Giuseppe Novelli**  
*Genetica e impatto ambientale nella salute umana*
- 16 **Giorgio Pranterà**  
*Le Scienze Omiche nella Ricerca Epidemiologica Ambientale - I*
- 22 **Sara Rinalducci**  
*Le Scienze Omiche nella Ricerca Epidemiologica Ambientale - II*
- 37 **Luca Proietti De Santis**  
*Le variazioni dell'espressione dei geni della riparazione del DNA e dei geni di risposta allo stress ossidativo come biomarcatori di esposizione ai metalli pesanti: progetto di ricerca presso il Termovalorizzatore di San Zeno (Arezzo)*
- 47 **Ivan Arisi**  
*Statistica e Machine Learning per dati ambientali*
- 62 **Gerry Melino, Giuseppe Novelli**  
*Il rapporto tra ambiente e geni: dal rischio professionale alla medicina di precisione*
- 71 Tavola Rotonda  
**L'innovazione scientifica nella riduzione del rischio ambientale e nella valutazione dello stato di salute.**  
Interventi di: *Corrado Clini, Filippo Brandolini, Francesco Di Maria, Marzio Lasagni.*

### «I QUADERNI di SAGEN»

*Periodico dell'Associazione Salute, Ambiente, Genoma*

**Direzione Scientifica:** Claudio Clini  
**Direzione editoriale:** Ilaria Di Tommaso  
**Hanno collaborato a questo numero:**  
Nadia Bel Haddad, Amal Douhab

© Copyright 2019 by Aldo Francisci Editore - Abano Terme  
Tutti i diritti riservati - Printed in Italy

## PREFAZIONE

**Cesare Pillon**

*Vice Presidente Esecutivo SAGEN*

*Sta maturando sempre di più l'idea, nella consapevolezza generale, che esista una stretta correlazione tra ambiente e salute e, più in particolare, tra inquinamento e malattie.*

*Per non incorrere in risposte emozionali, ascientifiche, che potrebbero tramutarsi in azioni a volte irrazionali che non mutano in minima parte la situazione in essere, occorre che il dibattito sulla situazione ambientale si concentri, oltre che sul grave tema dei cambiamenti climatici, anche sui fenomeni legati all'influsso sulla salute di vari fattori inquinanti, i quali, sia per la pericolosità che per la quantità di soggetti coinvolti, sono altrettanto rilevanti dal punto di vista economico e sociale per la comunità umana e per l'ambiente in generale.*

*Molto spesso le ricette che vengono proposte hanno il difetto di essere inadeguate ed inattuabili nel breve e nel medio periodo. Un esempio per tutti può essere il problema dello smaltimento dei rifiuti. Puntare all'obiettivo "rifiuti zero" è apprezzabile dal punto di vista prospettico, ma necessita di tempi, programmazioni e strumentazioni adeguate. Mentre per i rifiuti urbani nell'Occidente industrializzato, attraverso il ciclo dell'economia circolare, la differenziazione e il riciclo possono ritenersi già in fase avanzata, così non si può dire delle realtà del mondo emergente o delle parti dei paesi poveri del terzo mondo. Resta comunque aperto il tema dei rifiuti pericolosi, il cui trattamento e smaltimento sono quasi alla fase iniziale, con qualche rara eccezione in alcuni Paesi europei e qualche Paese a forte industrializzazione. Una pianificazione razionale, basata sulle best practice mondiali, che si ponga il problema di una rete impiantistica moderna, in grado di affrontare tutte le problematiche relative ai temi dello smaltimento a valle di una differenziazione spinta, per quanto riguarda i rifiuti urbani, i residui delle lavorazioni e i rifiuti industriali, non può che orientarsi verso la ricerca e la promozione delle nuove tecnologie per impianti di trattamento che nel medio e lungo periodo saranno comunque necessari. Il risultato è sotto gli occhi di tutti: siti industriali inquinati da stoccaggi di rifiuti tossici e nocivi e montagne di rifiuti urbani lasciati per le strade, vere bombe ecologiche senza controlli.*

*Vista la molteplicità dei fattori che influenzano l'ambiente, non tutti riconducibili alle attività antropiche, analizzare gli effetti che i vari fenomeni hanno sulla salute umana è oggi possibile utilizzando le risultanze dei più recenti studi scientifici.*

*Il Workshop organizzato da Associazione Salute Ambiente Genoma - SAGEN in collaborazione con Il Centro Documentazione e Studi dei Comuni Italiani - ANCI IFEL, ha posto le basi di un approfondimento scientifico che appare collocato sui livelli significativi delle ricerche sulle mutazioni ed espressioni genomiche indotte da inquinanti ambientali. Sia la lectio magistralis del Prof. Giuseppe Novelli che*

*dei proff.ri Giorgio Prantera, Sara Rinalducci, Luca Proietti De Santis e Gerry Melino hanno sottolineato come le nuove scoperte scientifiche e le innovazioni tecnologiche ci diano la possibilità, attraverso indagini genomiche condotte con metodi scientifici all'avanguardia, di individuare e correggere i comportamenti delle persone o addirittura le situazioni ambientali causa di alcune patologie specifiche e influenti sulla salute in generale.*

*Sono risultati veramente interessanti, inoltre, gli interventi della tavola rotonda in cui si sono confrontate le opinioni di esperti e studiosi che, a vario titolo, operano nel campo dell'ambiente e della salute.*

*Il Prof. Corrado Clini e il Dott. Filippo Brandolini hanno offerto una visione panoramica delle carenze e delle prospettive dell'applicazione delle nuove tecnologie e delle innovazioni nella gestione dei processi industriali che hanno ed avranno impatto in ambito ambientale.*

*Il Prof. Francesco Di Maria ha analizzato le possibilità delle nuove tecnologie impiantistiche nella riduzione dei fenomeni e delle best practice che hanno effetto sull'impatto ambientale. Il Dott. Ivan Arisi ha illustrato il funzionamento delle metodiche del Machine Learning e le validazioni dei modelli applicati. Infine, l'Ing. Marzio Lasagni ha motivato la decisione di aderire al progetto di ricerca su eventuali effetti delle attività dell'impianto di recupero integrale dei rifiuti di San Zeno (Arezzo), indagine condotta dall'Associazione SAGEN, sotto la direzione scientifica del Prof. Claudio Clini.*

*Sicuramente la strada da percorrere è ancora lunga e tortuosa, ma la scienza a base delle innovazioni può contribuire a cambiare le metodologie sia sull'analisi dei rischi che sulla definizione di soluzioni praticabili per il miglioramento complessivo della salute.*

*L'Associazione SAGEN, associazione no profit, si pone l'obiettivo di promuovere e favorire la conoscenza, la pratica, lo sviluppo e la diffusione di attività scientifiche nell'ambito della salute nel suo più ampio significato. In tale direzione promuove iniziative attraverso pubblicazioni scientifiche e divulgative, indagini epidemiologiche, analisi genomiche predittive, studio e sviluppo del percorso di analisi del rischio individuale e programmi di educazione sanitaria.*

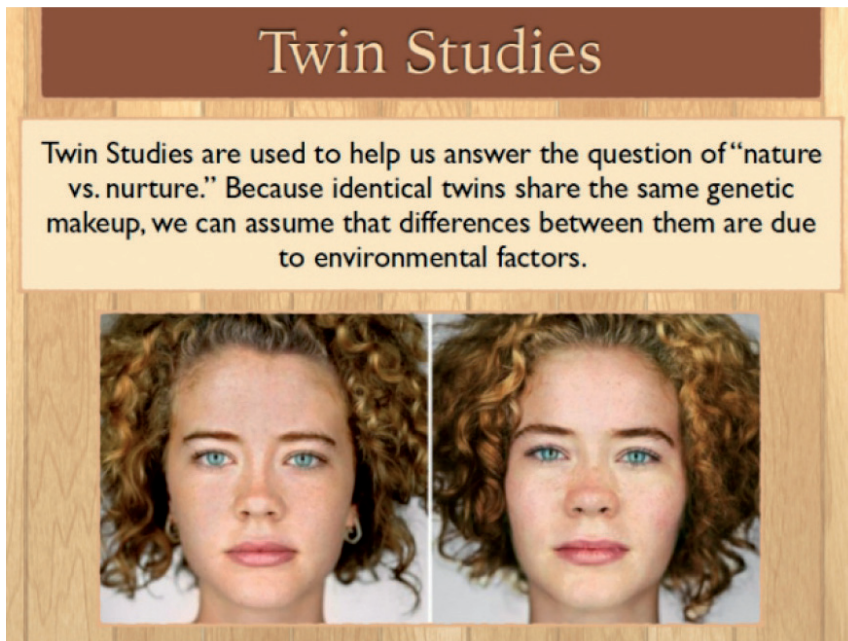
## Giuseppe Novelli

*Rettore e Direttore U.O.C. Laboratorio di Genetica Medica, Università degli studi di Roma "Tor Vergata"*

### Genetica e Impatto Ambientale nella salute umana

Il mio collega Richard Dawkins scrisse in un libro famosissimo questa frase che mi ha colpito molto: "se potessimo leggere il linguaggio nel DNA del tonno e della stella marina avrebbero scritto la parola "mare" nel testo". Immaginate, nel DNA c'è scritto "mare", l'ambiente dove vive questa specie. In effetti il rapporto tra specie biologica e ambiente è da tempo oggetto di ricerche per scoprirne le relazioni, cioè di come un ambiente possa influenzare l'espressione di una specie, il suo comportamento e le caratteristiche.

Il modello biologico essenziale, per capire il rapporto tra ambiente e genetica, è sempre stato quello dei gemelli. Dagli anni '70 in poi, si è cominciato a capire che ciò che non è scritto nei geni è dovuto all'ambiente. Due gemelli identici, che hanno quindi lo stesso DNA e condividono le stesse cose, divergono però in qualcosa, ciò che noi chiamiamo "concordanza" o "discordanza" nei gemelli.



Siamo andati avanti decenni con metodi e approcci statistici, matematici e metodologici che hanno fornito una serie di indicazioni importanti come, per esempio, i gruppi sanguigni in cui esiste una concordanza totale nei gemelli identici e una

concordanza un po' meno forte per i fratelli dizigotici, così come per patologie, dalla schizofrenia al diabete, la concordanza è molto forte. È come dire che il peso dei geni è più importante o lo è l'ambiente a seconda delle caratteristiche.

Concordance for Selected Traits		
Trait	Concordance in Identical Twins	Concordance in Fraternal Twins
Blood type	1.0	0.66
Eye color	0.99	0.28
Schizophrenia	0.69	0.28
Diabetes	0.65	0.19
Alcoholism (males)	0.41	0.17
Attention deficit disorder/ hyperactivity	0.58	0.31
IQ score	0.69	0.88

© Brooks/Cole, Cengage Learning

Un mio collega inglese sta studiando i gemelli in una maniera diversa, cercando di individuare le differenze in gemelli identici. Tre gemelli identici sono stati volutamente separati alla nascita: ognuno vive in un mondo suo, ognuno ha fatto esperienze diverse in tre diversi strati sociali. Così è stato studiato ogni singolo aspetto o patologia e in che modo siano state influenzate dall'ambiente.

In sostanza è un metodo per capire con strumenti innovativi, come quello della genomica, quanto l'ambiente può intervenire sul DNA e rispondere alle seguenti tre domande fondamentali:

- perché alcuni si ammalano e altri no?
- perché alcuni si ammalano di più e altri no?
- perché su alcuni un trattamento terapeutico funziona e su altri no o è addirittura nocivo?

Alcune di queste risposte dipendono esclusivamente da ciò che c'è scritto nel DNA. Ad esempio, nel caso della progeria, una drammatica patologia che causa un precoce invecchiamento per cui un bambino a soli 14 anni è come un anziano di 80 anni, una persona nasce con una mutazione sulle 3 miliardi di lettere che compongono il nostro codice genetico e questo succede con una frequenza di 1 ogni 8 milioni. Da un genitore (normalmente maschio) nasce in questi casi un bambino affetto da progeria e non esiste ad oggi alcuna terapia conosciuta.



## PROGERIA - HGPS



10 months



14 yr

Hutchinson, *Medicochirurgical Trans* 1886;69:473

Gifford, *Practitioner* 1904;73:188

Non c'è ambiente che possa ritardare questo fenomeno. Il cambiamento di una sola lettera del nostro codice genetico trasforma inevitabilmente una persona. Ciò accade anche per altre patologie, le cosiddette “patologie genetiche monogeniche” in cui c'è l'alterazione di un solo specifico gene.

Altre situazioni, invece, riguardano aspetti o patologie “complesse” o “multifattoriali” come, ad esempio, l'asma. E' ben noto che alcune comunità americane, come gli Hutteriti e gli Amish, condividono tra di loro le caratteristiche genetiche, essendo provenienti da stessi gruppi etnici, vivono in ambienti pressoché simili, in quanto si sposano tra di loro, non hanno industrializzazione, non hanno corrente elettrica e vivono di agricoltura. Non si riusciva a capire come il 21% degli Hutteriti fosse affetto dall'asma e un altro gruppo, gli Amish, solo per il 5%.

Hutteriti



21,3%

Amish



5%

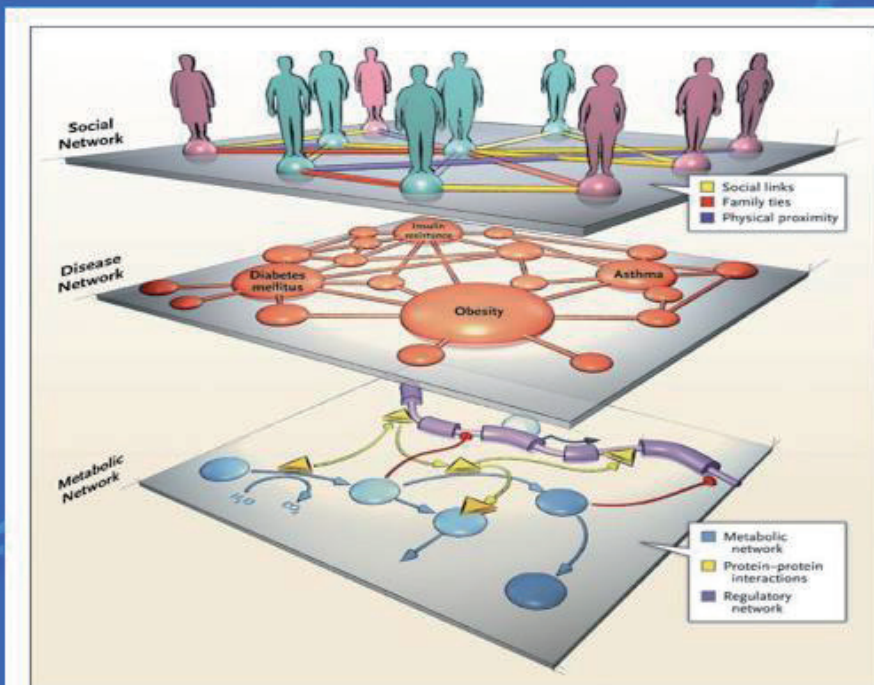
Innate Immunity and Asthma Risk in Amish and Hutterite Farm Children

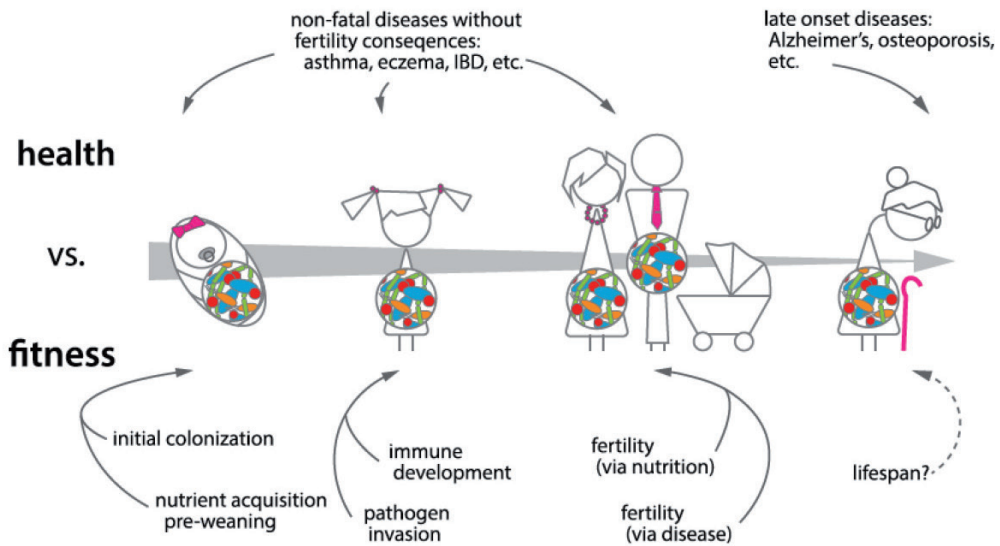
*N Engl J Med* 2016;375:411-21.  
DOI: 10.1056/NEJMoa1508749

Questo è stato un mistero per moltissimo tempo, finché non si è scoperto che la questione era abbastanza banale: all'interno delle loro case la pulizia era più o meno accurata. In un ambiente maggiormente polveroso possono svilupparsi alcuni particolari batteri e questo dipende dalla frequenza più o meno elevata di una variante polimorfa di alcuni geni che predispongono o aumentano la suscettibilità in un determinato contesto ambientale a sviluppare o meno l'asma. Questo è un sistema con cui oggi vengono studiate molte situazioni e si ha la possibilità di creare tre livelli di analisi complessa:

- i singoli genomi, cioè la singola persona, la singola popolazione;
- l'interazione tra di loro e le connessioni sociali tra le persone,
- il fatto che noi viviamo in un mare di DNA che non è solo il nostro ma è quello di batteri, virus, funghi e protozoi che vivono nel nostro corpo quotidianamente e lo influenzano.

## Network Medicine: a system of integrated complex networks

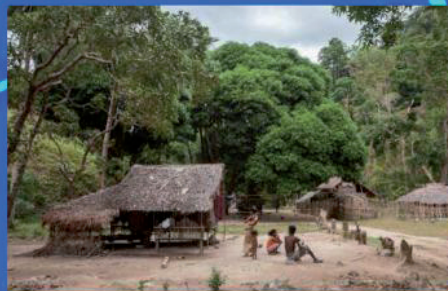




**Fig. 2.** Host-microbiome interactions can affect both health and fitness. Dysbiosis is associated with a number of negative *health* outcomes, including obesity, asthma, and certain cancers. Negative health outcomes are not sufficient evidence for coevolution of the microbiome and host, however. Not all of these diseases result in negative *fitness* consequences by limiting reproductive success. Microbiomes potentially impact host *fitness* at multiple stages of life by affecting survival through reproductive years or reducing fertility. In infancy, microbes extract energy from non-digestible components of milk, increasing nutrient acquisition at this vulnerable age. During childhood, a stable microbiome prevents invasion of deadly pathogens. In adulthood, the microbiome potentially influences fertility, either by altering nutrition or causing disease. Finally, the microbiome may be important for lifespan. Although lifespan after menopause will not result in more children, the grandmother hypothesis predicts that care of extended kin results indirectly in higher fitness [139]. *IBD* inflammatory bowel disease

Oggi abbiamo a disposizione nuove tecnologie che ci permettono di studiare le specie batteriche che costituiscono il nostro microbioma, batteri che influenzano la salute, la qualità della vita dal momento della nascita, forse anche prima, fino alla vecchiaia, in un cambiamento quotidiano dinamico legato a ciò che facciamo, che respiriamo e con cui veniamo a contatto.

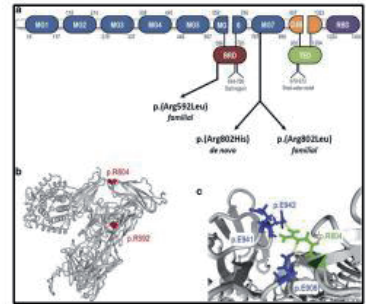
Alcuni esempi: l'otite, causa della spesa maggiore di antibiotici negli Stati Uniti con 5 miliardi di dollari all'anno, è un'infezione batterica di diverse specie. In una piccola popolazione delle Filippine si è scoperto un luogo dove si ha una prevalenza di otite del 50%.



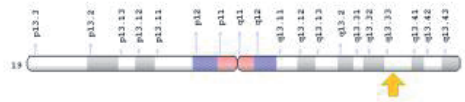
**The indigenous Filipino population has a ~50% prevalence of otitis media.**

La domanda è: perché il 50% si ammala e l'altro 50% no? E' resistente al batterio o perché assume antibiotici? La risposta è stata trovata identificando variabili genetiche su due geni importanti: A2ML1, macroglobulina, proteina di membrana dell'epitelio che impedisce ad alcune forme di batteri di entrare più o meno facilmente e il secondo è il FUT2, una fucosiltransferasi, che regola la secrezione delle molecole che determinano i nostri gruppi sanguigni del gruppo A B 0. Anche se può sembrare strano è stato dimostrato che il FUT2 ha anche a che fare con l'otite.

**A2ML1 encodes  $\alpha$ -2 macroglobulin-like-1 protein or A2ML1, which is localized to middle ear mucosal epithelium**



**FUT2 is a Golgi stack membrane protein that is involved in the creation of a precursor of the H antigen, which is required for the final step in the soluble A and B antigen synthesis pathway. This gene is one of two encoding the galactoside 2-L-fucosyltransferase enzyme**



[Eva Maria Cutiongco-De La Paz \(University of the Philippines Manila, Philippines\)](#) personal communication

Altro esempio di interazioni tra geni e ambiente è dato da una malattia poco conosciuta: la podoconiosi, che colpisce i piedi della persona affetta. Sembrava fosse causata da un parassita, un platelmite e si è studiata in alcune zone dell'Etiopia. Colpisce 1 agricoltore su 20 e solo questa categoria di persone si ammala. Come mai non tutti e perché colpisce solo gli agricoltori?



Podoconiosis is a tropical lymphedema resulting from long-term barefoot exposure to red-clay soil derived from volcanic rock

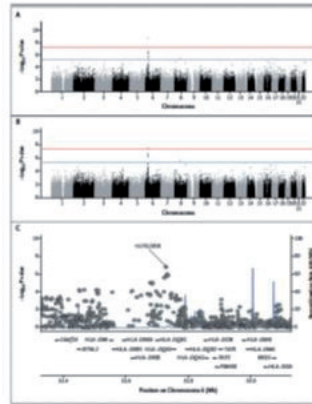


Un gruppo di studiosi inglesi ha scoperto una variante di un gene importante, l'HLA, che regola i nostri trapianti di cellule, di tessuti o di organi e che è differente in alcune persone rispetto ad altre. Chi possiede questa variante, che non determina apparentemente nessun problema e va a lavorare scalzo in un terreno pieno di argilla si ammala di podoconiosi. Basterebbe mettere gli stivali o i sandali per non ammalarsi. Ci sono voluti quasi 20 anni per debellare questa malattia, finchè qualcuno ha capito che era dovuta a questa interazione geni-ambiente e che per prevenirla bastava andare a lavorare nei campi con stivali e non scalzi. Questo è solo un esempio di come la “genetica ambientale” possa essere applicata alle patologie.

## Sandals instead of genetics....

HLA Class II (HLA-DQA1)  
Locus and Susceptibility to  
Podoconiosis

N Engl J Med 2012;366:1200-8.



Se prendiamo malattie come il cancro, bisogna andare indietro fino alla fine del '700 quando si sono avute le prime prove documentate del tumore al testicolo, molto più frequente negli spazzacamini che quotidianamente venivano a contatto con polvere e cenere.



1775 -Pott-increased occurrence of scrotal and nasal cancer among chimney sweeps -concluded: chimney soot was the causative agent

1892 -Butlin-duration of exposure influences risk of developing cancer



All'inizio del '900 si sono avute le prime prove sperimentali che hanno dimostrato come molti tumori, siano dovuti a cause ambientali, a molecole chimiche che interagiscono con il DNA e lo rendono in qualche modo suscettibile. Le persone sviluppano il tumore perché si altera la costituzione delle basi del DNA.

## Environmental exposure and cancer

- **1915 -Kennaway, Hieger** - dibenz(a,h)anthracene in coal tar is skin cancer inducing agent
- **1928-Meisel** – chemicals may be mutagens
- **1951-1983 Miller et al.** - established the role of chemical structure and metabolism in the “activation” of carcinogens
- **1988 - Heim et al.** - proved: accumulation of chromosomal damage needed for cancer formation
- **1981-2011 - Pfeifer, Walker, Yang et al.** – tobacco smoking and lung cancer, aniline dyes and bladder cancer; asbestos and mesothelioma; benzene and leukemia

Non esiste alcun dubbio sulla correlazione tra spazzacamino e tumore al testicolo o tra il fumo di sigaretta e il tumore ai polmoni, ma la domanda è perché non tutti i soggetti a rischio sviluppano il tumore o viceversa? E perché soggetti non a rischio lo sviluppano ugualmente? Questo paradosso è stato spiegato dai risultati della ricerca genetica. Ci sono varianti genetiche che se presenti predispongono maggiormente alla patologia oncologica. Se compaiono certe varianti ti ammali di più e se non ce l'hai sei più protetto e ti ammali di meno. Questa differenza oggi la conosciamo benissimo, la pesiamo, sappiamo individuarla, studiarla, caratterizzarla e definire per ognuno di noi un profilo specifico di rischio sulla base dei polimorfismi che ognuno di noi si porta dietro. È quello che è chiamato un “profilo genomico specifico selettivo” caratteristico della persona o anche del tumore di quella determinata persona.

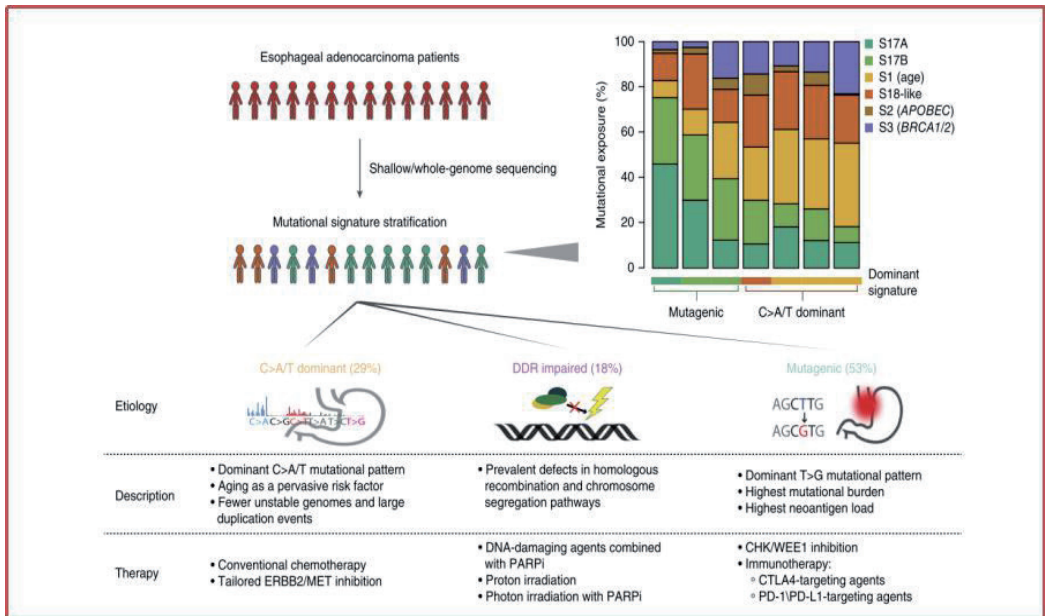
### Single Nucleotide Polymorphism (SNP) of the Genes in Carcinogenesis

Some discrepancies:

1. More than 80% of bronchial (lung) carcinoma patients = cigarette smokers  
less than 15% “heavy” cigarette smokers = bronchial carcinoma patients
2. 66% urinary bladder carcinoma patients = cigarette smokers  
less than 10% cigarette smokers = urinary bladder carcinoma patients
3. More than 90% gastric carcinoma patients = infected by *Helicobacter pylori*  
5 % infected by *Helicobacter pylori* = gastric carcinoma patients

Quello con cui nasci è il DNA che hai ereditato dai tuoi genitori con qualche mutazione già presente mentre quello presente nei tumori è un DNA alterato o per cause naturali o per cause dovute ad azione dell'ambiente. Questi profili mutazionali come vengono definiti, hanno cambiato anche il modo di fare diagnosi, prognosi e terapia del cancro.

## Mutational signatures provide a rationale for therapies in Esophageal Adenocarcinoma



Come nell'esempio di adenocarcinoma dell'esofago, in cui attraverso i profili specifici di ogni adenocarcinoma differente si può decidere se trattare in un luogo o in un altro con farmaci diversi. Quello che per noi una volta ero lo stesso tumore, ma che lo stesso tumore non è, oggi con il profilo molecolare di caratterizzazione si può a priori decidere di trattare con certi farmaci o con altri o con la radioterapia. Questo sta rivoluzionando in maniera completa e totale il modo di fare terapia, questo per ragioni che hanno portato alla conoscenza definitiva di quella che è l'anatomia molecolare di un cancro. Ogni tumore è diverso, ogni paziente che ha un cancro è diverso e ogni tumore ha una sua impronta molecolare caratteristica, un profilo di rischio determinato.

Sulla base di questo oggi è possibile stabilire quali sono i profili di rischio individuali senza e con i tumori o anche come trattare quel tipo di tumore sulla base della presenza di mutazioni o di polimorfismi di variabili genetiche che sono più o meno presenti. Questo è un modo oggi che ci permette di dire in maniera caratteristica, di selezionare e classificare i tumori che sono più o meno forti, determinanti, dominanti o potenzialmente più pericolosi da altri un po' meno pericolosi e un po' più tollerabili. Su questo si può discute-

re moltissimo e solo sulla base di questo riusciamo a definire quelle che sono varianti ad alto rischio (segnate in rosso nella figura precedente) con cui un individuo nasce e quelli a più basso rischio, più comuni e frequenti distribuiti nelle diverse popolazioni. Oggi cominciamo ad avere cataloghi interi che ci dicono esattamente per ogni tumore qual è il suo profilo, la sua anatomia. Questo è stato possibile grazie a studi estesi tra molti laboratori e molte popolazioni e grazie a studi internazionali che hanno permesso, nel giro di qualche anno - anche perchè sono crollati i costi di analisi del genoma - di fare grandi analisi su una grande quantità di genomi. Per esempio, da poco è stato fatto l'atlante di tumori di persone esposte al tabacco, che quindi naturalmente fumavano, in cui sono state trovate cellule di tumore nei polmoni che hanno permesso di ricavare dei profili che gli informatici traducono in maniera precisa e accurata e che ci permettono di individuare i diversi profili di singoli tumori, che hanno quindi anatomia molecolare diversa per ogni singolo tumore. Questo naturalmente esiste per il tumore in sé, ma la domanda era, fino a poco tempo fa, i carcinomi ambientali che tipo di profilo danno? La risposta l'abbiamo avuta in un articolo apparso recentemente su Science dove si riportano i risultati di una ricerca su mutazione del genoma. Sono stati identificati i profili molecolari di 80 carcinomi da esposizione di altrettanti fattori ambientali, come ad esempio radiazioni UV o radiazioni gamma. In un modello in vitro cellule staminali, cloni delle stesse cellule, sono state esposte agli stessi fattori esogeni e comparando le alterazioni del DNA si è trovata nella maggior parte dei profili studiati, la corrispondenza che tutti noi aspettavamo. Un determinato profilo è associato a un determinato fattore cancerogeno e questo dimostra l'effetto sul genoma dei fattori ambientali. Un passo importante nella definizione dei processi patogeni che cambia la storia delle malattie ed è determinante anche nella scelta di farmaci specifici. Cambiano le curve di sopravvivenza tra chi ha un farmaco mirato e chi riceve un farmaco a largo spettro senza alcuna conoscenza come si faceva qualche anno fa. Vorrei portarvi un esempio concreto di come la conoscenza del genoma ha cambiato non solo la storia di una malattia, ma la politica e l'aspetto sociale di un paese, di una regione e di una soluzione, questo è bellissimo per voi che studiate i rapporti tra geni e ambiente e lo dobbiamo ad un italiano, Michele Carbone, che da molti anni studia il mesotelioma, un tumore di cui si sapeva poco o niente. C'è un paese in Turchia, un villaggio in cui c'erano frequenze del mesotelioma del 50% nella popolazione, ma non si è mai capito perchè ci fosse una così alta incidenza di questo tumore raro in quella popolazione. Dopo essere stati lì ed aver speso molto tempo si è capito che questo era dovuto a delle caratteristiche biologiche intrinseche di questa popolazione, ma soprattutto alla presenza di un minerale, l'erionite, che è presente in quella zona e con cui venivano costruite le case degli abitanti di quella zona. L'erionite è un minerale fibronico e nelle popolazioni esposte a questo minerale si sviluppavano questo tipo di tumori. Sulla base di questo dato scientifico, il dott. Carbone ha convinto il governo turco a spostare la popolazione da un villaggio ad un altro e a ricostruire le case senza l'utilizzo del minerale erionite. Da quel momento in poi l'incidenza del mesotelioma è pari a zero. La stessa esperienza è stata ripetuta in North Dakota (USA) dove il dott. Carbone



ha rilevato la presenza dello stesso minerale (erionite) nell'asfalto stradale, respirato durante le fasi di posizionamento del manto stradale e successivamente dalla sua erosione determinata dal traffico veicolare. Modificata la composizione dell'asfalto è visibilmente diminuita l'incidenza del mesotelioma.

In conclusione, possiamo dire indubbiamente che il DNA fornisce la musica, come ha detto J. Craig Venter, ma le nostre cellule e l'ambiente procurano l'orchestra. Allora noi dobbiamo capire bene la musica e l'orchestra prima di poter fare qualcosa altrimenti non otteniamo niente. Quali sono gli strumenti con cui oggi possiamo fare della prevenzione molecolare e studiare appunto i profili di rischio senza ovviamente usare i metodi invasivi? Uno strumento oggi c'è, ed è quello che si chiama "biopsia liquida" attraverso lo studio del DNA circolante nel sangue si possono recuperare i profili di DNA che vengono dalle cellule tumorali che li liberano nel nostro torrente sanguigno. Oggi questa tecnica è di routine, si può fare molto bene e permette di recuperare quantità di DNA libero plasmatico in diverse situazioni di tumore con un semplice prelievo di sangue, senza fare alcuna biopsia. Oggi si può studiare in maniera precisa e personalizzata il profilo di rischio e non solo, si può fare qualcosa di più: come Università di Tor Vergata siamo coinvolti direttamente in un progetto per lo studio di un primo programma di prevenzione primaria utilizzando la biopsia liquida sulle persone sane e non sui malati, su cui già funziona abbastanza ed è usato nelle terapie per stabilire qual è il profilo di rischio in termini di carico mutazionale (se è basic, se cambia). Da una base stabile se osserviamo che dopo un po' di tempo aumenta il profilo mutazionale o incrementa l'instabilità genomica, che è sintomo di tumore, allora si approfondisce con altre tecniche e con altre metodiche.

In sostanza la biopsia liquida fornisce oggi un sistema diagnostico di caratterizzazione del tumore e di trattamento eventualmente terapeutico o il farmaco giusto e prognostico. Non abbiamo dati sui sani se non quelli prospettici, abbiamo fatto un primo studio sulla biopsia liquida sul plasma circolante di soggetti sani e come viene recuperato nei donatori sani un certo quantitativo mentre nei donatori con cancro accertato cambia notevolmente: ce n'è di più. Oggi questo è quantificabile e studiabile in maniera tale che si può addirittura stabilire il Cancer Management, la gestione di rischio sulla base di queste conoscenze, per arrivare infine a quella che è la medicina di precisione personalizzata: 8 pazienti, 8 terapie diverse. Ognuno è diverso e dobbiamo avere una terapia giusta per la giusta persona.

Noi siamo il risultato dei nostri geni e non è tutto ovviamente descritto dal nostro DNA, ma certamente in esso c'è molto. La genetica non ha il completo controllo di noi ma non possiamo sfuggire all'influenza di questo suo effetto e il meglio che possiamo fare è imparare a convivere con il nostro DNA e cercare di conoscerlo per cercare di prevenire le patologie.

## **Giorgio Pranterà**

*Docente di Genetica e Direttore del Dipartimento di Scienze Ecologiche e Biologiche,  
Università degli studi della Tuscia*

### **Le scienze omiche nella ricerca epidemiologica ambientale - I**

L'epidemiologia è nata all'inizio dell'800 e ha riguardato all'inizio l'epidemiologia delle malattie infettive, in cui c'è una diretta correlazione causa-effetto tra la malattia e l'agente patogeno qualunque esso sia (un virus, un batterio, un fungo, un protozoo ecc.). Nel caso invece delle malattie non trasmissibili, in cui è presente una forte componente ambientale che interagisce con la struttura genetica degli individui, si presentano situazioni più complesse e complicate da decifrare sia poiché in un ambiente vi è una molteplicità di fattori, chimici o fisici (multifattorialità), sia per il verificarsi di esposizioni ripetute a sostanze inquinanti. Per esempio, in una persona, che vive e lavora in un ambiente particolarmente inquinato come quello industriale e allo stesso tempo conduce uno stile di vita non raccomandabile (fumo, alimentazione non regolare e non bilanciata ecc.), è difficile isolare le singole componenti ambientali, potenziali cause di una condizione patologica. In questi casi è necessario cercare d'isolare gli eventi causali di una patologia, partendo dalle modificazioni che questi eventi provocano sul panorama molecolare e sui pattern metabolici di un individuo.

L'epidemiologia delle malattie non trasmissibili solleva notevoli problematiche: la prima, è la differenza tra danno somatico e danno germinale perché, naturalmente, mentre il danno somatico ha effetto sull'individuo, il danno germinale ha un effetto trasmissibile, quindi ha un effetto sulla progenie e a cascata sulla specie (in questo caso quella umana). L'altra problematica è passare da una generica associazione tra un'esposizione e una patologia, alla individuazione della causa reale del rapporto tra l'esposizione e le suddette modificazioni.

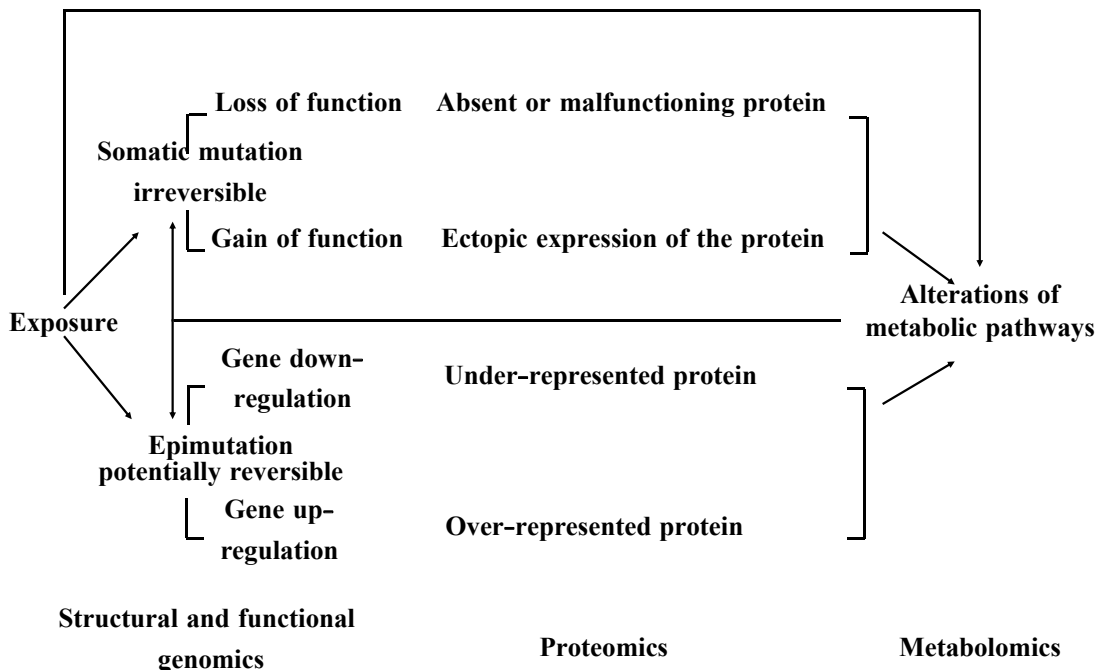
Le ricerche epidemiologiche, poi, sono generalmente condotte a livello di popolazione (lavorando quindi su delle medie): quanto un singolo individuo, sottoposto alla stessa esposizione, sia destinato o meno allo sviluppo della patologia, non può essere previsto con sicurezza, ma gli può solo essere assegnata una probabilità statistica.

Ultimo punto è la necessità di identificare singoli, possibilmente precoci, eventi molecolari che portano allo sviluppo di una patologia.

Un caso tipico è quello dell'esposizione ad agenti inquinanti che provocano uno stress ossidativo, una cascata infiammatoria e poi lo sviluppo di una patologia.

Un riassunto di quelli che sono i rischi dell'esposizione ambientale si trova nella Figura 1 che mette anche in evidenza il ruolo delle diverse scienze omiche nei differenti passaggi del flusso dell'informazione genetica.

**Epitome of the environmental exposure risks and the role of the OMICS: Searching for exposure early marks**



*Figura 1. Diagramma dei rischi di esposizione ambientale e il ruolo delle “omiche”.*

Qui, abbiamo il risultato di che cosa l’esposizione può provocare (a parte la risposta da parte dell’individuo, quindi un’attivazione di pattern di espressioni geniche peculiari): mutazioni somatiche che sono irreversibili e mutazioni della regolazione epigenetica, che sono potenzialmente reversibili, aspetti studiati da quella che si chiama “genomica strutturale e funzionale”, quindi il sequenziamento del DNA e il sequenziamento dell’RNA messaggero che è quello che dà l’idea dell’attività di un gene.

Tutto ciò provoca alterazioni della produzione quantitativa e qualitativa delle proteine, che alla fine può essere monitorata attraverso gli studi di proteomica. Tutto ciò va poi a ricadere sull’alterazioni dei pattern metabolici, studiati attraverso la metabolomica; l’alterazione dei pattern metabolici a sua volta, può provocare mutazioni somatiche o epimutazioni, innescando quindi un circolo vizioso.

Un tipico esempio può essere quella di una mutazione indotta dall’ambiente che altera una proteina chiave del metabolismo cellulare e della progressione del ciclo cellulare, quale P53, la quale, una volta alterata, porta a sua volta all’insorgenza di instabilità genomica e quindi a nuove mutazioni.

Le principali scienze omiche sono la genomica, la genomica funzionale, la proteo-

mica, la metabolomica e l'ultima nata, l'epigenomica che si sviluppa sulla scia del riconoscimento dell'importanza dell'epigenetica (Figura 2).

## The Omics

### **GENOMICS:**

• *The whole sequence of the genome (genotype) of a living or an extinct prokaryote or eukaryote*

### **FUNCTIONAL GENOMICS:**

• *the complete repertoire of expressed genes from a prokaryote or from a given tissue of an eukaryote*

### **EPIGENOMICS :**

• *the global pattern of the molecular signatures (histone modifications, DNA methylation) that regulate epigenetic phenomena*

### **PROTEOMICS :**

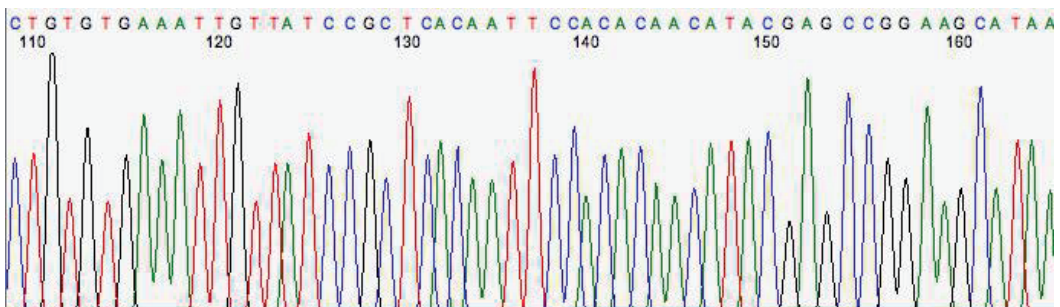
• *the whole set of proteins, whether modified or not, produced from the cells of a given tissue of a given organism*

### **METABOLOMICS :**

• *the complete landscape of the metabolites present in the cells of a given tissue of an organism*

*Figura 2. Le “omiche”*

La prima omica è la genomica, che nasce quando Frederick Sanger, che aveva già preso il premio Nobel per il sequenziamento delle proteine, decise di prendere il secondo Nobel scoprendo come si può decifrare la sequenza del DNA. In Figura 3 si può vedere la versione automatizzata del suo metodo, in cui ad ogni picco colorato diversamente (ferogramma), corrisponde una base. Quindi in base alla successione dei colori si ricostruisce la sequenza del DNA che stiamo analizzando.



**Figura 3.** Una sequenza di DNA ottenuta con il metodo Sanger automatizzato di sequenziamento del DNA.

Successivamente ci sono stati dei perfezionamenti della tecnica originale ma anche delle innovazioni, come il sequenziamento Illumina che utilizza una PCR a ponte, che produce dei cloni di sequenze tutte uguali, fatte replicare e a loro volta sequenziate colorate con l'incorporazione di nucleotidi fluorescenti.

Una delle ultime tecnologie è il sequenziamento nanopore, che non richiede la duplicazione del DNA tramite PCR, il DNA entra in un nano-poro, viene frammentato in singoli nucleotidi, provocando delle piccolissime alterazioni, nano-variazioni del campo elettrico che possono essere registrate. Dato che ogni nucleotide provoca una variazione diversa, in questo modo è possibile ricostruire la sequenza in questo modo.

Lo strumento che può operare il sequenziamento nanopore è grande quanto un cellulare, è portatile e quindi adatto per l'applicazione sul campo.

Tutto ciò ha portato ad un aumento dell'efficienza del sequenziamento del DNA. La Figura 4 riporta il tasso di aumento del numero di genomi umani sequenziati; il 2008 è il momento di svolta, quando viene introdotto il sequenziamento automatico.



Source: Stephens ZD, Lee SY, Faghri F, Campbell RH, Zhai C, Efron MJ, et al. (2015) Big Data: Astronomical or Genomical? PLoS Biol 13(7).

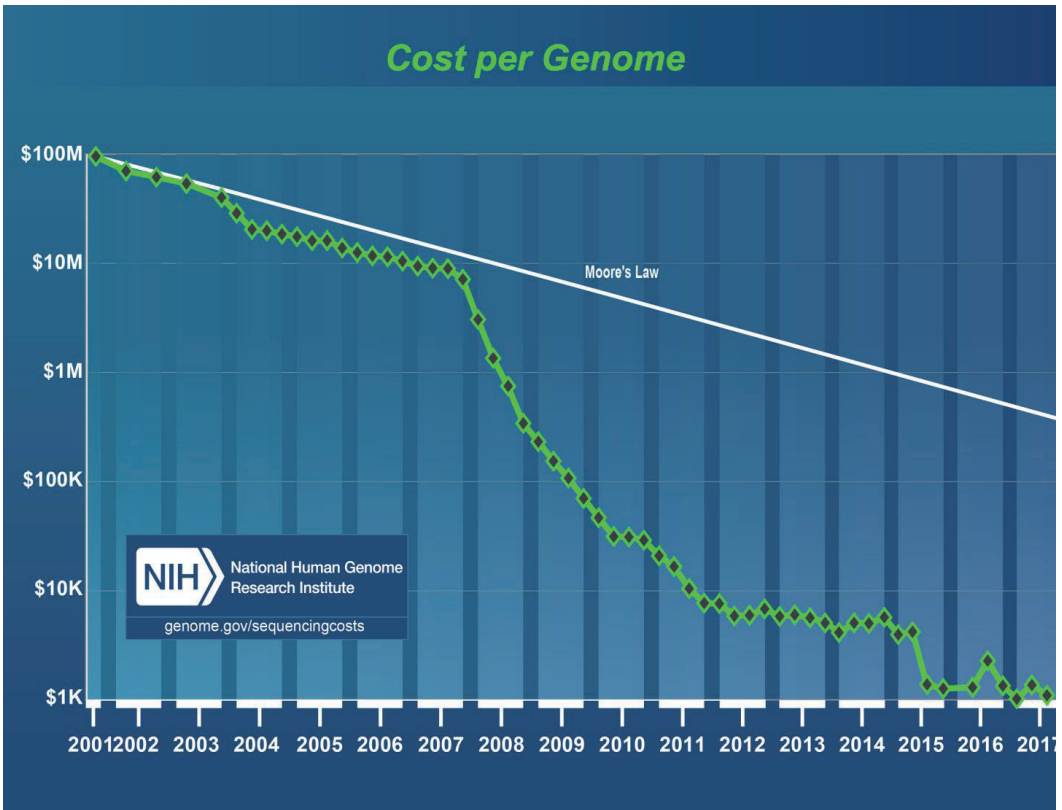
Figura 4. Tasso di crescita dell'efficienza del sequenziamento del DNA.

Questo è accompagnato da una diminuzione impressionante dei costi di sequenziamento: si parte nel 2001, anno in cui veniva completato il sequenziamento del genoma umano, partendo da 10000 \$ per megabase (un milione di coppie di basi sequenziate) per arrivare nel 2017 a meno di un decimo di dollaro.

Quindi questo accompagna anche la diminuzione del costo di sequenziamento del genoma dai 100 milioni di dollari del primo sequenziamento, che ha visto l'inve-

stimento di un Consorzio Pubblico Internazionale (Human Genome Project) e di una Compagnia privata (Celera genomics), a meno di 1000\$ per un sequenziamento (Figura 5).

Attualmente, ci sono delle ditte che possono fare il sequenziamento a poche centinaia di dollari. Se andate sul sito di queste ditte ci sono delle offerte speciali, ad esempio se entro due giorni una persona manda la propria saliva ad un costo ridotto di 200\$, si possono scegliere vari tipi di prestazione di analisi della sequenza (standard, premium, diagnostica) come quando con la macchina si scelgono gli allestimenti. Adesso abbiamo la stessa cosa per il DNA, comprando l'allestimento che si preferisce con naturalmente costi diversi.



**Figura 5.** Il decremento del costo di sequenziamento di un genoma dai primi anni del secolo ai giorni nostri.

Tuttavia, proprio adesso che abbiamo a disposizione mezzi veloci e relativamente poco costosi per determinare la sequenza del DNA, ci si è dovuti confrontare con il fatto che conoscere la sequenza del DNA, non vuol dire conoscere automaticamente come questa controlli i fenomeni biologici complessi. Questo perché si è andato affermando sempre più il ruolo della modulazione dell'espressione genica attraverso la regolazione epigenetica, che definisce quei fenomeni ereditari che

non dipendono dalla sequenza del DNA per se. Molto spesso l'epigenetica viene invocata anche per fenomeni che sono semplicemente di regolazione genica, ma alla maggior parte di essi non si può applicare la definizione di fenomeni ereditari, e quindi non rientrano nella definizione di epigenetica. Al contrario delle mutazioni, i fenomeni epigenetici sono reversibili e quindi sono buoni candidati per essere i bersagli dell'interazione gene-ambiente. Questa interazione gene-ambiente, intervenendo sull'espressione dei geni, mette in evidenza il ruolo dei contaminanti ambientali sulla vita di un individuo.

L'epigenomica consiste nella mappatura fine dei marchi (signatures) epigenetici, su di un genoma. Si basa su varie tecnologie, una di queste è la cosiddetta "Chromatin immunoprecipitation" (ChIP), seguita dal sequenziamento del DNA (ChIP seq), che permette di mappare le modificazioni istoniche su specifiche sequenze di DNA. Le altre tecniche epigenomiche di elezione sono quelle che permettono di determinare i pattern di metilazione del DNA sull'intero genoma.

Tornando a quello che dicevamo all'inizio, qual è il ruolo specifico e innovativo delle tecniche omiche?

Ci aiutano a capire più direttamente qual è la relazione tra l'esposizione e lo sviluppo finale di una patologia, perché ci danno una misura in maniera analitica e precisa, di quello che è il panorama di mutazioni, il panorama di alterazione d'espressione genica, il panorama di modificazioni delle proteine e del metaboloma, susseguenti appunto all'esposizione ad un insulto ambientale.

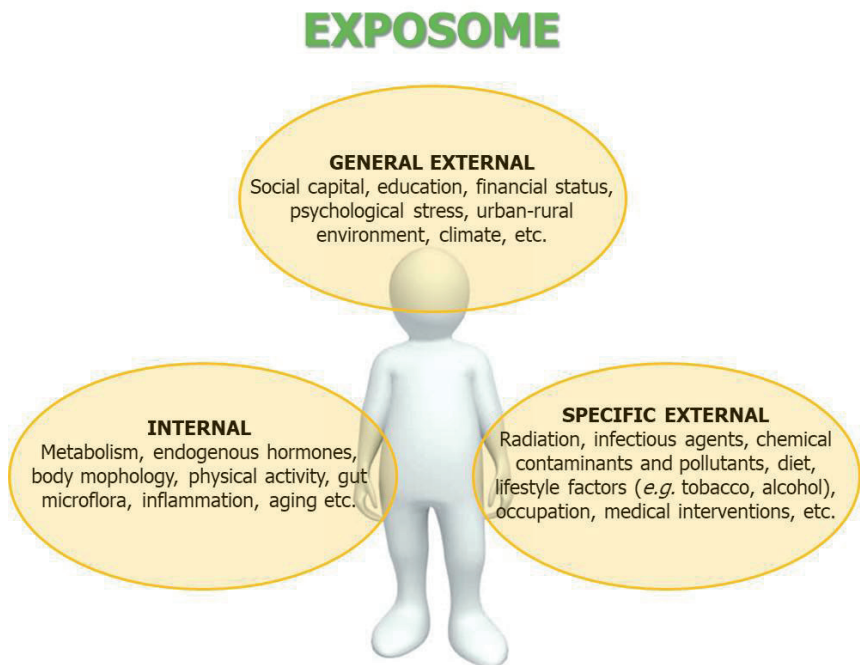
Anche queste tecniche di analisi ci aiutano inoltre a trovare dei biomarker precoci delle alterazioni prodotte dalle esposizioni ambientali, che permettono di prevedere e anche intervenire terapeuticamente su quelli che sono i danni dovuti all'esposizione ambientale. Questo da ultimo potrebbe portare allo sviluppo di terapie mirate per arrivare a quella che è la "medicina personalizzata", come approccio sempre più efficace alla diagnosi precoce e alla cura di queste malattie non trasmissibili.

**Sara Rinalducci**

*Docente di Biologia Molecolare, Dipartimento di Scienze Ecologiche e Biologiche, Università degli studi della Tuscia*

## **Le scienze omiche nella ricerca epidemiologica ambientale - II**

Il termine di esposoma umano, inteso come complemento al genoma, fu proposto all'incirca nel 2005 ed è diventato ormai un concetto ben definito, come nuovo paradigma per capire e valutare quali sono i fattori determinanti per la salute umana. In generale, l'esposoma può essere definito come l'insieme di tutte le esposizioni, sia endogene che esogene, a cui un individuo è sottoposto durante tutta la sua vita; esposizioni che ovviamente non sono soltanto di carattere chimico, come si può vedere in Fig.1, ma si possono considerare anche fattori di stress non chimici come per esempio lo stesso stress psicologico, le radiazioni, il clima, etc.

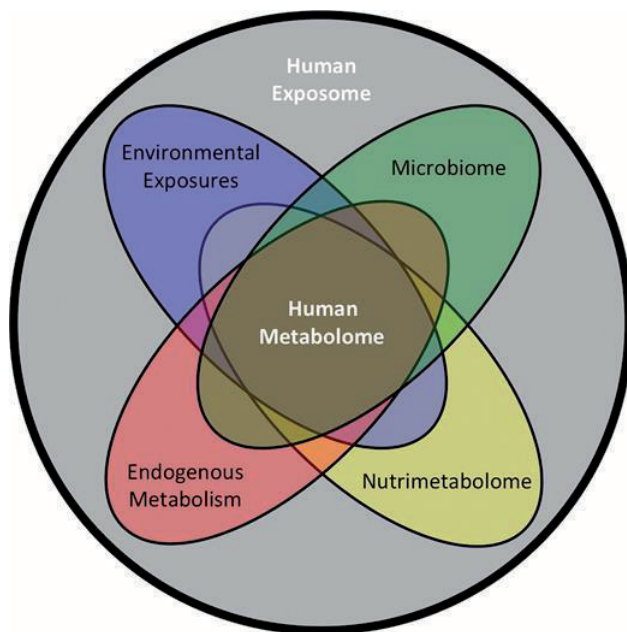


*Figura 1. adattata da Siroux et al. Int. J. Epidemiol. 2012, 41, 24–32*

L'esposoma descrive quindi i fattori non genetici che possono influenzare il nostro fenotipo. A parte le successive rifiniture che ci sono state rispetto alla definizione originale del termine "esposoma", in realtà una forte spinta si è avuta verso la comprensione del ruolo del metaboloma umano come parte dell'esposoma, in quanto il metaboloma umano può essere considerato come l'insieme, non solo dei



metaboliti endogeni prodotti sotto il controllo della biochimica umana, ma anche di un ampio spettro di sostanze di natura esogena (quindi xenobiotiche), composti prodotti da microbi presenti all'interno del nostro intestino, o sostanze che possono derivare dall'assunzione di alimenti (Fig. 2).



*Figura 2.* tratta da Athersuch and Keun *Mutagenesis* 2015, 30, 755-762.

Quindi in realtà il nostro metaboloma è una sorta di crogiolo metabolico in cui avvengono tante trasformazioni chimiche e tante interazioni; basti pensare che una sostanza introdotta dall'esterno può essere appunto metabolizzata dai nostri enzimi, ma nello stesso tempo può essere magari coniugata con composti di origine microbica e questo può portare eventualmente al suo riassorbimento e/o alla sua escrezione di nuovo verso l'ambiente. Ne consegue che l'interazione tra il nostro metabolismo e quello che è l'ambiente multifaccettato di un individuo è sicuramente una relazione altamente dinamica.

Esistono diverse tecnologie per poter valutare l'esposoma. In generale, però, si può parlare di strategia bottom-up o top-down (Fig. 3). L'approccio "bottom-up" si usa per identificare quale è la fonte di esposizione attraverso misurazioni ambientali basate su sistemi d'informazione geografica, o attraverso l'utilizzo di dispositivi mobili, di questionari o la realizzazione di immagini per documentare lo stato della dieta, il consumo di droghe, l'utilizzo di farmaci o cosmetici. Nell'approccio "top-down", invece, i livelli circolanti dei composti parentali e/o dei suoi metabo-

liti e sottoprodotti vengono valutati individualmente attraverso analisi biologiche nel sangue, nelle urine o in altri tessuti, andando essenzialmente a ricercare quelli che possono essere definiti come *biomarkers*. Questo approccio “dall’alto verso il basso” non mi consente però di avere informazioni dirette sulle fonti ambientali e sui livelli di esposizione, anche se questo può essere affrontato, in un certo senso, attraverso i modelli farmacocinetici basati sulla farmacologia.

## TOOLS FOR THE ASSESSMENT OF THE EXPOSOME

### BOTTOM-UP AND TOP-DOWN APPROACHES

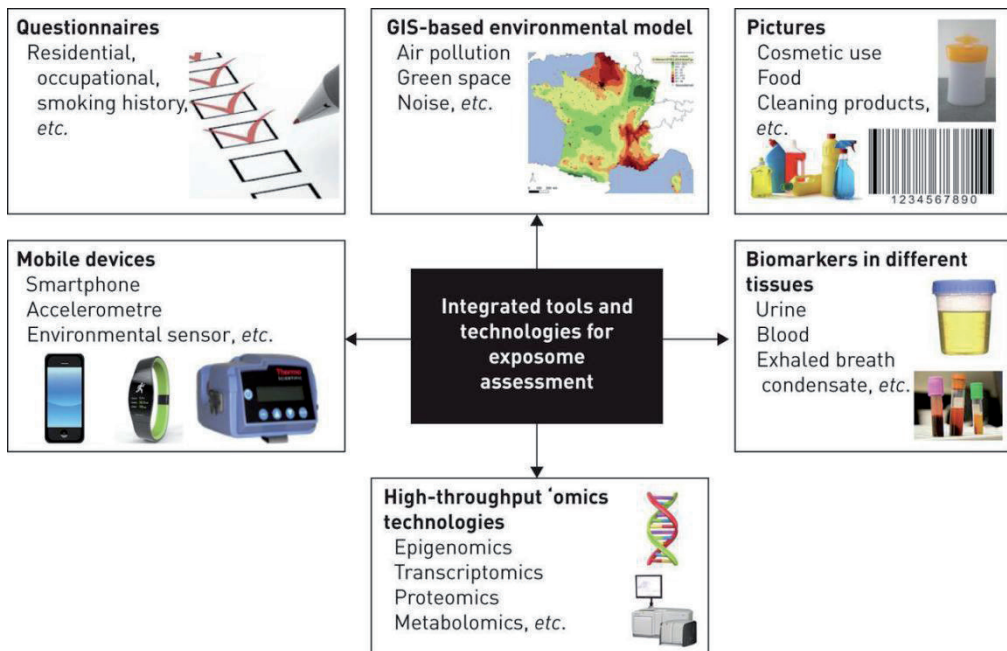
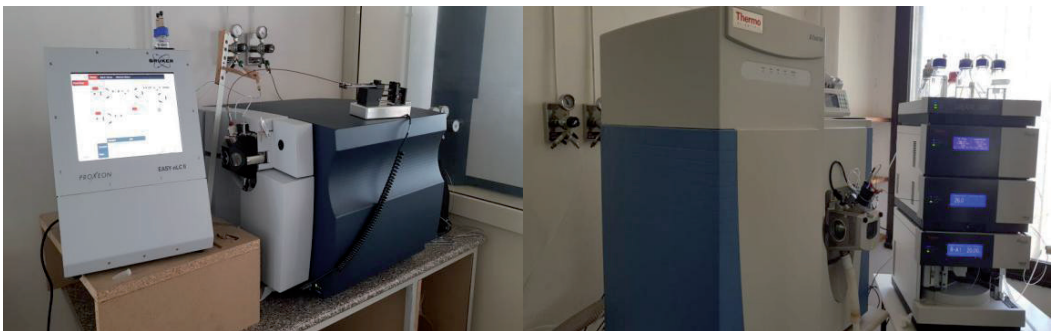


Figura 3. tratta da Siroux et al. *Eur. Respir. Rev.* 2016, 25, 124-129.

L’individuazione di biomarcatori è affrontata oggi grazie alle tecnologie omiche, che includono l’epigenomica, la trascrittomica, ovviamente la proteomica, ma anche la metabolomica. Di tutte le entità molecolari, (geni, trascritti, proteine e metaboliti), i metaboliti sono quelli che più rispecchiano il fenotipo in maniera diretta, in quanto sono il prodotto finale dell’elaborazione biochimica a monte. Quindi la metabolomica sta sicuramente diventando lo strumento di eccellenza per questo tipo di ricerche in campo epidemiologico.

La metabolomica si avvale principalmente di due piattaforme analitiche: la spettroscopia di risonanza magnetica nucleare (NMR) che identifica i metaboliti sulla base delle loro frequenze di risonanza in un campo magnetico, e la spettrometria di massa (MS) che misura il loro rapporto massa/carica. Entrambe le tecniche hanno

numerosissimi vantaggi, tra cui il fatto di poter lavorare con piccolissime quantità di campione, in più sono metodi analitici che consentono una forte riproducibilità dei risultati. Tuttavia, la spettrometria di massa è ormai diventata la tecnica di elezione per l'analisi dell'esposoma e ciò è dovuto al fatto che la spettrometria di massa ha una più alta sensibilità rispetto all'NMR. Questa caratteristica è molto importante se pensiamo che un metaboloma complesso può anche contenere livelli molto bassi di piccole molecole che invece possono avere ampi effetti sul fenotipo. Non in ultimo, la spettrometria di massa può essere collegata facilmente con sistemi di separazione a monte, per esempio sistemi di gas-cromatografia (GC) o liquido-cromatografia (LC), permettendo di ridurre la complessità della matrice di partenza, ma anche di rivelare un vasto range di diverse classi di composti chimici. In Fig. 4 sono riportati due differenti spettrometri di massa accoppiati con sistemi LC che operano a nano-flussi, consentendo per l'appunto di raggiungere sensibilità elevatissime. Inoltre sono state sviluppate diverse tecnologie a livello degli spettrometri di massa con analizzatori che raggiungono risoluzioni pari a 200.000 ed accuratze anche dell'ordine di 1-2 ppm; avere un'accuratezza di massa alla quarta cifra decimale significa poter determinare la composizione elementare di uno ione, quindi massima caratterizzazione di una sostanza.



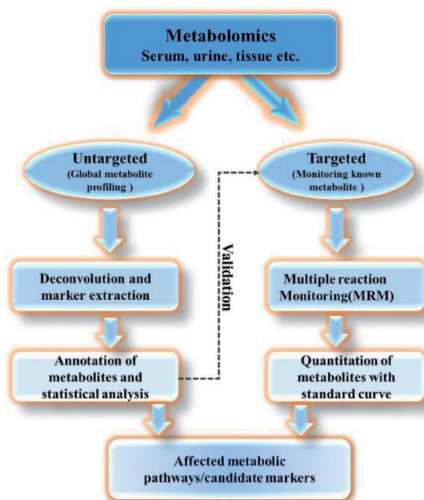
*Figura 4. Differenti tipologie di spettrometri di massa on-line con sistemi nano-LC.*

Come si lavora in metabolomica?

Si può adottare un approccio *targeted* quando si ha un'ipotesi a priori e si vuole andare ad identificare un metabolita di interesse conosciuto, oppure un approccio *untargeted* quando si vuole misurare il numero più alto possibile di metaboliti, andando quindi a fare un *profiling* generale di quello che è all'interno del campione biologico di partenza (Fig. 5).

Ovviamente procedere senza avere un'idea iniziale di quello che si va a cercare, produrrà una mole di dati enorme, rendendo più difficile la misurazione quantitativa; tuttavia tale approccio apre interessanti scenari perchè la possibilità di “vedere” contemporaneamente più metaboliti consente la scoperta di numerose, e magari nuove, interazioni che possono correlare con una determinata patologia. I due approcci (*targeted* e *untargeted*) possono essere combinati, anzi una loro combinazione sequenziale è fortemente consigliata, anche per una validazione dei risultati ottenuti.

**Non-targeted** analyses allow to simultaneously detect and quantify a wide range of compounds that fall under various chemical classes. This makes it the perfect platform for measuring exposure, especially considering that endogenous metabolites are included in the exposome definition.



**Targeted** profiling identifies a limited number of specific metabolites of known identity and is a more hypothesis-driven approach.

Figura 5. tratta da More et al. J. Proteomics 2015, 127(Pt A), 73-79

Può essere stimato che nell'uomo esistano più di 500 metabolomi cellulari dinamici, perciò lo step di identificazione diventa il collo di bottiglia più grande in un esperimento di metabolomica. Per risolvere questo problema, vengono in aiuto molteplici databases (Fig. 6), alcuni dei quali sono stati ampliati ultimamente, includendo diverse sostanze xenobiotiche. Questo ci permette di facilitare il biomonitoraggio delle esposizioni, ma anche di migliorare la valutazione del conseguente impatto biologico.

**Table 1**

Mass spectrometry metabolite databases for identification of environmental exposures

Database name	Description	URL
Human metabolome database (HMDB)	114,113 xenobiotic and endogenous metabolites with chemical, biochemical, and clinical information.	<a href="http://www.hmdb.ca/">http://www.hmdb.ca/</a> [33]
Toxic exposome database (T3DB)	3767 toxic compounds, targets and gene expression data, part of the HMDB suite.	<a href="http://www.t3db.ca/">http://www.t3db.ca/</a> [28]
METLIN	961,829 xenobiotic and endogenous metabolites with chemical information. Contains information from DSSTox.	<a href="https://metlin.scripps.edu/">https://metlin.scripps.edu/</a> [34]
Exposome-Explorer	692 dietary and pollutant biomarkers, with concentration values measured from biospecimens with intra class correlation coefficients.	<a href="http://exposome-explorer.iarc.fr/">http://exposome-explorer.iarc.fr/</a> [32]
Madison-Qingdao Metabolomics Consortium Database	20,300 xenobiotics and endogenous metabolites, with chemical information	<a href="http://mmcd.nmrham.wisc.edu/">http://mmcd.nmrham.wisc.edu/</a> [35]
Drugbank	10,513 drug entries with drug target information, part of the HMDB suite	<a href="https://www.drugbank.ca/">https://www.drugbank.ca/</a> [36]
PubChem	93,977,784 compounds, xenobiotic and endogenous metabolites but also peptides, and chemically altered macromolecules. Data is derived from hundreds of sources.	<a href="https://pubchem.ncbi.nlm.nih.gov/">https://pubchem.ncbi.nlm.nih.gov/</a> [37]
CompTox Chemistry Dashboard	758,000 xenobiotics with chemical information compiled from multiple sources; PubChem, and US EPA's DSSTox, ACToR, ToxCast, EDSP21, and CPCat.	<a href="https://comptox.epa.gov/dashboard">https://comptox.epa.gov/dashboard</a> [38]

Figura 6. tratta da Rattray et al. Human Genomics 2018, 12, 4.

In Fig. 7 è mostrato il tipico workflow di un esperimento metabolomico di tipo *untargeted*. Per trarre una qualsiasi conclusione, soprattutto in questo caso dove si ottiene una grossa quantità di dati, è necessario trattare questi dati sia in maniera matematica che statistica. Infatti, è fondamentale ottenere una clusterizzazione in gruppi dei nostri campioni che consenta la loro caratterizzazione in base a quelli che sono i segni metabolici distintivi.

### Flow chart of the steps in data analysis of metabolomic data for epidemiologic studies

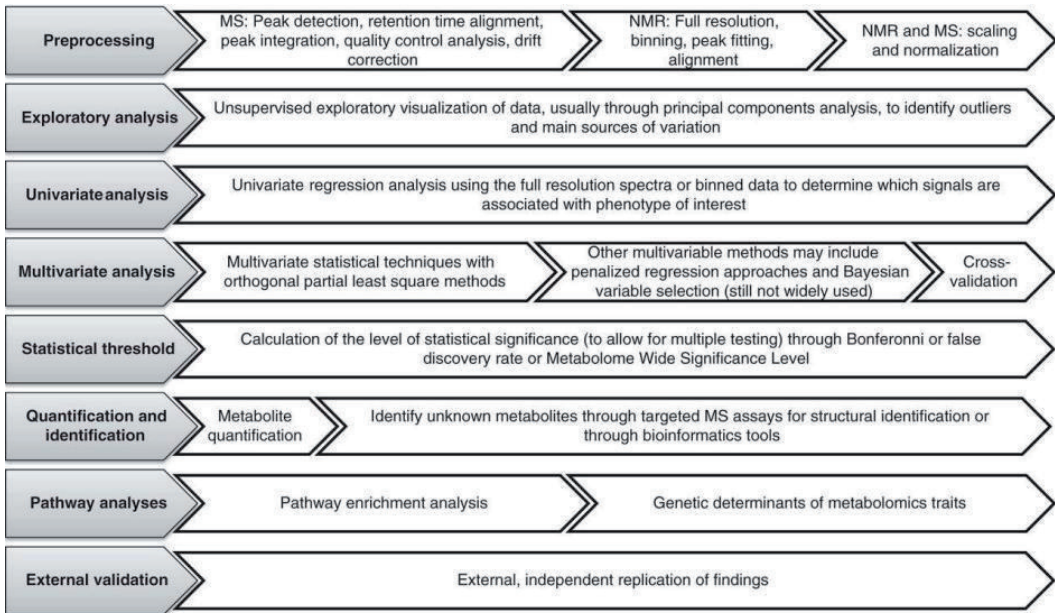


Figura 7. tratta da Tzoulaki et al. *Am. J. Epidemiol.* 2014, 180, 129-139.

Inizialmente i dati richiedono una pre-elaborazione che prevede l'allineamento dei cromatogrammi, selezione e merge dei picchi (segnali corrispondenti agli ioni ottenuti (Fig. 8).

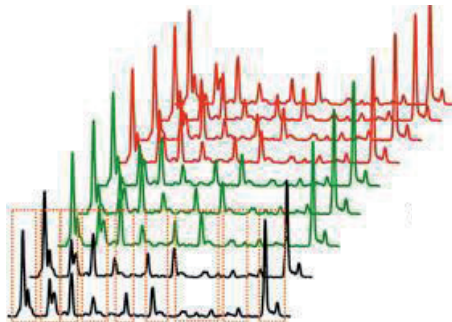


Fig. 8 Allineamento dei profili cromatografici

Spesso è necessario eseguire uno step di normalizzazione per minimizzare la variazione tra campione e campione. A questo segue poi l'approccio statistico che può essere di tipo univariato, come il "volcano-plot" (Fig. 9, pannello di sx) che ci permette di avere subito un'idea delle discriminanti fra le due tipologie di campioni, oppure multivariato attraverso PCA e/o PLS-DA (Fig. 9, pannello di dx), così da effettuare una riduzione dei dati multidimensionali.

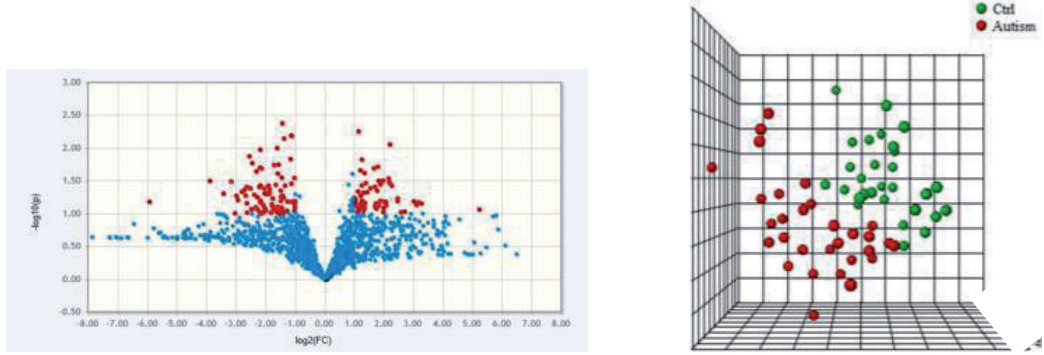


Figura 9. Esempi di volcano-plot e PLS-DA.

Si possono poi applicare diversi approcci per la selezione delle variabili più influenti (VIP score; Fig. 10), o test di correzione della molteplicità (FDR, Bonferro-ni). Dopo l'identificazione dei biomarcatori, è necessario eseguire test di verifica della performance per convalidarli (analisi della curva ROC).

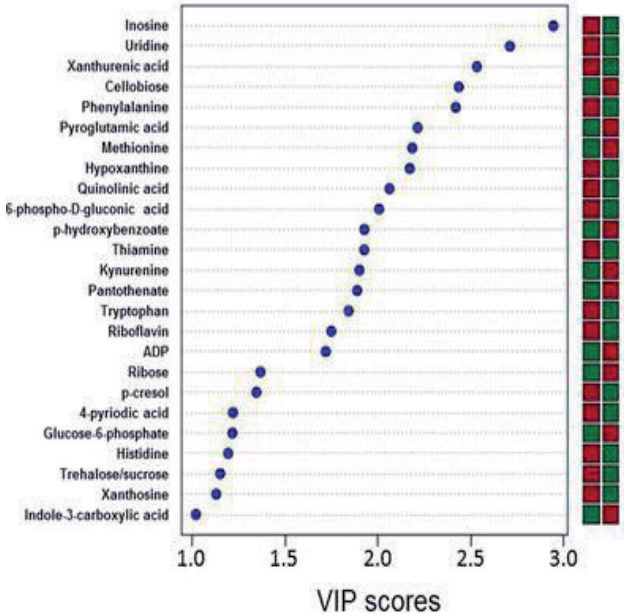
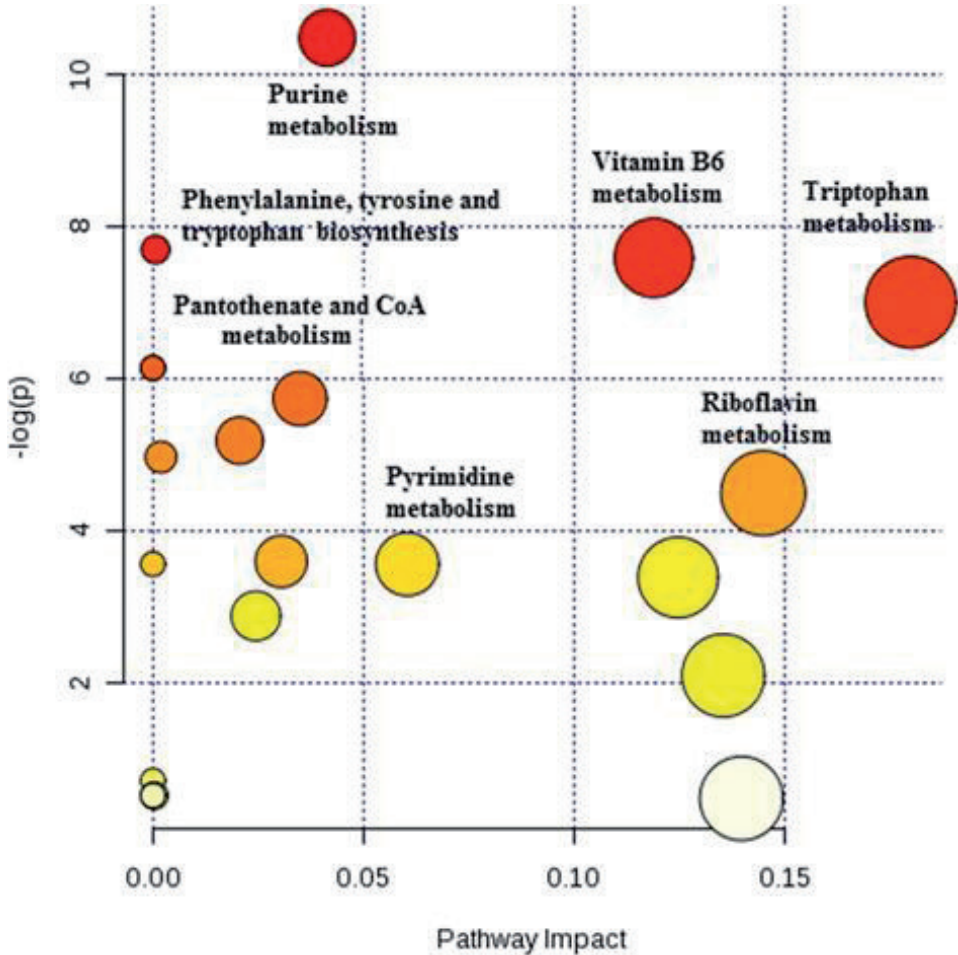


Figura 10. Esempio di determinazione dell'importanza delle variabili (VIP, Variable Importance in Projection).

Infine, ma non per importanza, si procede all'analisi dei pathways e dei networks metabolici (Fig. 11), aumentando sicuramente la rilevanza clinica della nostra indagine ed anche l'impatto biologico.



*Figura 11. Esempio di analisi dei pathways metabolici (MetPA) attraverso il software MetaboAnalyst ([www.metaboanalyst.ca](http://www.metaboanalyst.ca)).*

Come tutti gli approcci metodologici, anche gli esperimenti di metabolomica hanno in sé delle criticità: ci sono fattori sia di tipo analitico che di tipo biologico di cui dobbiamo tenere conto (Fig. 12).

<u>Biological Factors</u>	<u>Analytical Factors</u>
<p>Accounting for the high biological background arising from inter-subject variability is an essential part of experimental design in metabolomics. Other sources of variability can include:</p> <ul style="list-style-type: none"> <li>➤ Dietary effects</li> <li>➤ Body mass index</li> <li>➤ Age</li> <li>➤ Sex</li> <li>➤ Subjects at high risk of disease</li> <li>➤ Subjects at high risk of exposure</li> <li>➤ Observer selection</li> <li>➤ Timing of subject sampling</li> <li>➤ Gene/Environmental effects</li> <li>➤ Circadian/temporal effects</li> </ul>	<p>Sample preparation and analytical procedures used within metabolomics experiments vary greatly across labs. Sources of variability that should be considered for study design include:</p> <ul style="list-style-type: none"> <li>➤ Collection and storage procedures</li> <li>➤ Between sample and batch variance</li> <li>➤ Between instrumental comparisons</li> <li>➤ Lack of global “Gold Standard” methodology driven by constantly evolving instrumentation (annually) from multiple vendors</li> <li>➤ Selection of extraction solvent</li> <li>➤ Selection of column chemistry</li> <li>➤ Selection of eluent composition</li> <li>➤ Ionization technique and source geometry</li> <li>➤ Signal processing algorithms</li> <li>➤ Compound library</li> </ul>

*Figura 12. tratta da Rattray et al. Human Genomics 2018, 12, 4.*

### *Applicazione della metabolomica allo studio dell'autismo.*

L'autismo è una malattia complessa, caratterizzata da gravi disturbi della comunicazione, del comportamento e dell'interazione con gli altri. Nelle forme più gravi, le persone affette non parlano, tendono ad isolarsi, presentano comportamenti stereotipati e disabilità intellettuali (Fig. 13). Se da un lato, la ricerca neuropsicologica tenta di capire la patologia dell'autismo attraverso studi di competenze cognitive, emozionali e sociali dei soggetti affetti, dall'altro, esistono diverse linee di ricerca volte alla scoperta delle basi neurobiologiche di questa sindrome, confermando la presenza di anomalie nello sviluppo di alcune strutture cerebrali, con alterazioni nei livelli di connessione tra certe aree nel cervello, ma anche disfunzioni immunologiche, nonché disturbi metabolici. Questi ultimi possono essere studiati attraverso la metabolomica.

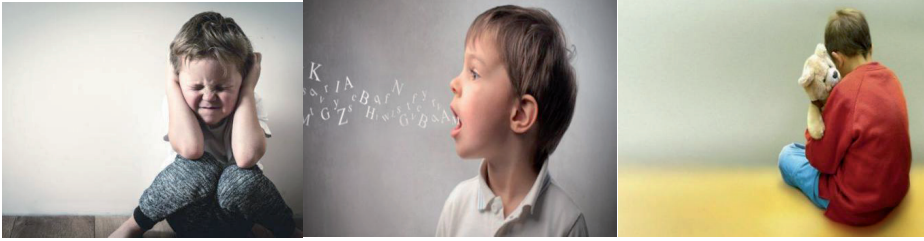
Sebbene le cause non siano state ancora individuate con precisione, sappiamo che esiste una grossa componente genetica, ma anche una forte influenza di diversi fattori ambientali (Fig. 14). Ciò risulta evidente esaminando i gemelli monozigotici



## Autism spectrum disorder (ASD)

Autism is categorized as a mental development disorder that occurs within the child's early years and, if left untreated, gets worse

- **Repeating words or phrases (echolalia)**
- **Repeated motions (rocking or spinning)**
- **Getting upset by minor changes**
- **Avoiding eye contact or physical touch**
- **Delays in learning to talk**



*Figura 13. Principali disturbi legati all'autismo.*

(stesso patrimonio genetico); infatti sembra che nel 20% dei casi, soltanto uno dei gemelli risulti affetto da autismo. Molti ricercatori sono concordi nel ritenere che la gravidanza ed i primi mesi di vita del bambino costituiscano i momenti più critici dove le esposizioni ambientali, come a metalli pesanti e pesticidi, giocano un ruolo predominante nel determinare l'insorgenza dell'autismo in quanto il sistema immunitario del neonato non è ancora sufficientemente sviluppato; il cervello del neonato infatti assorbe le sostanze tossiche e non riesce ad eliminarle. Altri fattori di rischio sono da ricercarsi nello stile di vita, di salute e nutrizionale della madre.

## ASD: environmental factors

Electromagnetic  
Radiation toxicity



Heavy metal toxicity



Life style



- **Pollution, nutrition and psychophysical state in which the child grows contribute to the pathophysiological mechanisms at the ASD basis**
- **The most sensitive periods for these substances are pregnancy and the first months of life, as the fetus or newborn is very vulnerable and does not yet have a sufficiently developed immune system. These substances are very easily absorbed by the brain, which often fails to eliminate them**

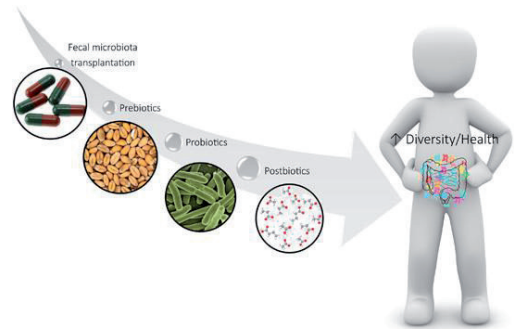
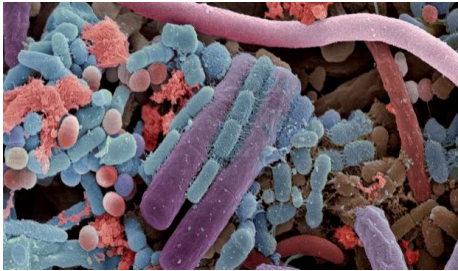
*Figura 14. Principali fattori ambientali che influenzano l'insorgenza dell'autismo.*

Un altro aspetto da considerare è l'influenza del microbiota intestinale (Fig. 15): sembra infatti che un'alta percentuale di soggetti affetti da autismo presenti disbiosi intestinale (alcuni batteri intestinali infatti possono influenzare diverse attività del cervello attraverso la secrezione di sostanze o la regolazione del sistema immunitario, e non a caso si parla spesso di "asse intestino-cervello").

## ASD: microbiota

### The human gastrointestinal tract is colonized by $10^{14}$ microorganisms

The microbiota commensals play an important role in human health by acting as a barrier against pathogens, exerting metabolic functions and stimulating the development of the host immune system



The microbiota is a complex ecosystem consisting of multiple species that interact with each other and with the host to generate a dynamic community

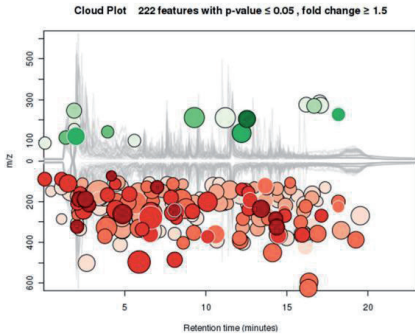
*Figura 15. Ruolo del microbiota nello sviluppo della patologia dell'autismo.*

Recentemente, alcune anomalie metaboliche sono state associate a varie malattie che comportano disturbi comportamentali associati all'autismo. Diversi studi hanno dimostrato la presenza di difficoltà di apprendimento in soggetti affetti da alterazioni metaboliche, evidenziando la presenza di un progressivo declino neuropsicologico in aree di intelletto, attenzione, memoria e funzioni esecutive. La perturbazione nell'omeostasi metabolica potrebbe portare ad alterazioni nei livelli di specifici metaboliti che possono essere valutate con la metabolomica.

Per questi motivi, abbiamo effettuato uno studio di tipo metabolomico su campioni di urina di 60 bambini, di cui 30 affetti da autismo e 30 di controllo. La coorte di pazienti è stata selezionata attentamente per escludere il più possibile fenomeni di variabilità; l'età dei bambini è stata ristretta da 2 a 7 anni e sono stati esaminati 22 maschi e 8 femmine (l'autismo ha un'incidenza più elevata nel genere maschile). Attraverso un'analisi metabolomica di tipo *untargeted* sono emersi circa 16000 metaboliti alterati, di cui 222 sono stati ritrovati up- o down-regolati in maniera statisticamente significativa (Fig. 16).

# Metabolomic approach for the study of ASD

## XCMS Metline RESULTS

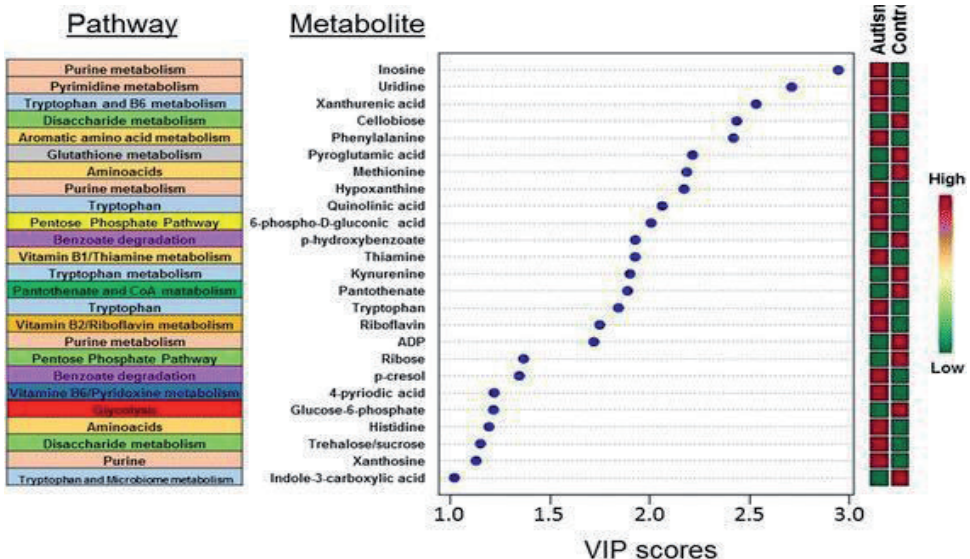


A considerable number (16704) of urinary altered metabolites were recorded. Among these, **XCMS identified 222 statistically significant metabolites** that were **up- or down-regulated** (red and green circles)

*Figura 16. Identificazione dei metaboliti alterati nei soggetti affetti da ASD attraverso il software XCMS Metline.*

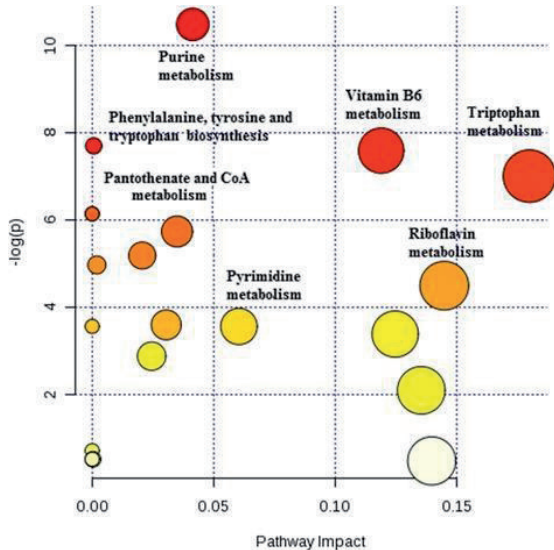
Successivamente siamo andati a vedere quali tra questi 222 metaboliti fossero i più discriminanti, ovvero quelli che potrebbero, in un certo senso, essere considerati come biomarkers della patologia. In Fig. 17 sono mostrati i 25 metaboliti distintivi più significativi attraverso l'utilizzo del VIP score; per ciascun metabolita è riportato anche il pathway metabolico di appartenenza.

The top 25 most discriminating metabolites between ASD cases and controls, ranked by **variable importance in projection** (VIP) scores, and their KEGG biochemical pathway



*Figura 17. Classificazione mediante VIP score dei metaboliti discriminanti (Gevi et al. Mol Autism 2016, 7, 47)*

In particolare, l'analisi statistica dei pathways metabolici ha evidenziato tre macro gruppi maggiormente alterati: metabolismo delle purine, metabolismo del triptofano e metabolismo della vitamina B6 (Fig. 18).



Color intensity (white to red) reflects increasing statistical significance, while circle diameter covaries with pathway impact.

- The main affected pathways can be subdivided into three macro groups:
- **purine metabolism:** pyrimidine deoxyribonucleosides; purine degradation; biosynthetic pathway
  - **tryptophan metabolism:** biosynthesis of eumelanin; serotonin and melatonin biosynthesis and degradation
  - **vitamin B6 metabolism**

*Figura 18. Analisi con il software MetPA dei pathways metabolici alterati (Gevi et al. Mol Autism 2016, 7, 47).*

Prendiamo ad esempio il metabolismo del triptofano (Fig. 19): il 90-95% di esso viene metabolizzato attraverso la via della chinurenina, circa l'1-2% finisce nel pathway della serotonina, mentre circa il 4,6 % viene metabolizzato dai batteri del nostro intestino.

Per ciò che riguarda il pathway della chinurenina (Fig. 19a), si può notare come nei bambini autistici ci sia un forte aumento di acido chinolinico e acido xanturenico, mentre invece c'è una forte diminuzione di acido chinurenico e chinurenina. Gli enzimi che portano alla formazione di acido chinolinico e acido xanturenico sono espressi principalmente nella microglia e nei macrofagi, riflettendo una produzione periferica di questi composti, mentre la produzione di chinurenina e acido chinurenico è concentrata a livello degli astrociti. Un punto focale è rappresentato dalla 3-idrossichinurenina, metabolita in grado di superare la barriera emato-ence-

# TRYPTOPHAN PATHWAY

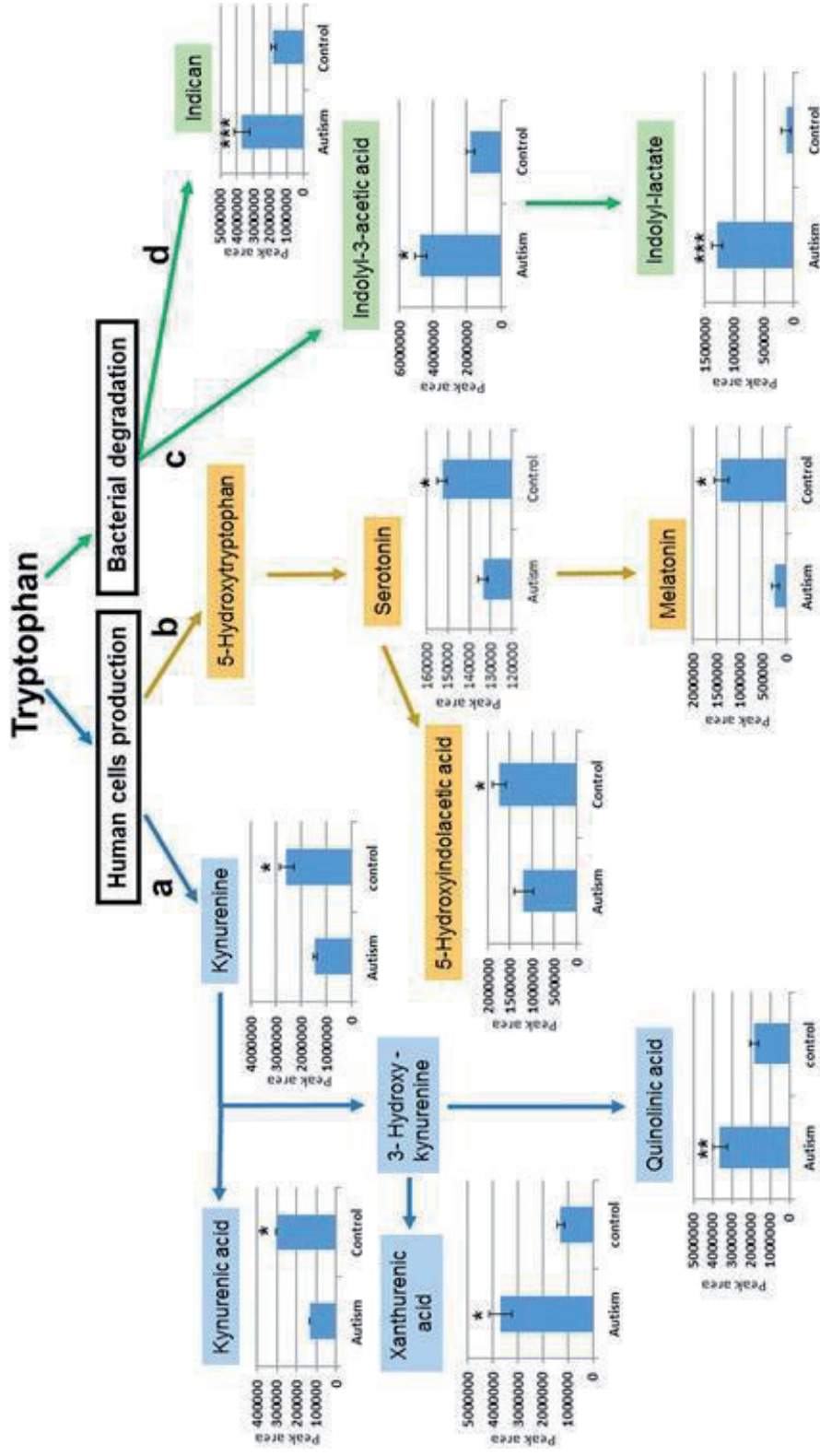


Figura 19. Variazioni quantitative dei metaboliti appartenenti al metabolismo del triptofano (Gevi et al. Mol Autism 2016, 7, 47).

falica. Dal momento che tutti i metaboliti a valle di questa sostanza (acido chinolinico e acido xanturenico) hanno un forte aumento nei soggetti affetti ad autismo, mentre quelli che si trovano a monte (acido chinurenico e chinurenina) presentano in realtà una diminuzione, si potrebbe ipotizzare un deflusso della 3-idrossichinurenina dal sistema nervoso centrale verso la circolazione sistemica.

Andando invece ad analizzare il pathway della serotonina (Fig. 19b), risulta rilevante la forte diminuzione della melatonina e questo si correla bene con l'evidenza che i bambini affetti da autismo, specialmente all'inizio della malattia e durante la prima infanzia, soffrono frequentemente di disturbi del sonno che sono molto responsivi a terapie farmacologiche con melatonina.

Infine, analizzando le variazioni dei metaboliti correlati alla degradazione batterica (Fig. 19c, d), si evidenzia un aumento significativo di indolano ed altri indoli nei bambini affetti da autismo. Ovviamente, è prevedibile che i composti di origine batterica ritrovati in altre indagini condotte su urine di soggetti con ASD possano differire nello specifico, ma questo può dipendere dalla diversa demografia dei campioni di partenza (è noto per esempio che l'etnia esercita profonde influenze sul microbioma). Tuttavia, a prescindere dalle singole molecole ritrovate, l'aumento di questo tipo di composti batterici è confermato nei diversi studi, avvalorando le alterazioni della flora batterica intestinale riscontrate nei bambini con questa patologia.

Seppure molto interessanti, sarà cruciale verificare queste osservazioni nel liquido cerebrospinale, perché questo scenario metabolico potrebbe avere almeno due importanti implicazioni cliniche: (i) l'acido chinolinico agisce come gliotossina, mediatore proinfiammatorio e molecola pro-ossidante, aumentando lo stress ossidativo andando a stimolare la microglia a rilasciare grandi quantità di NO e superossido; (b) l'acido chinolinico esercita effetti eccitotossici agendo come agonista del recettore per l'N-metil-D-aspartato (NMDA), stimolando il rilascio di glutammato, bloccando la ricaptazione del glutammato negli astrociti e riducendo l'attività della glutammina sintasi; invece, l'acido chinurenico esercita effetti antiossidanti e genera neuroprotezione attraverso l'antagonismo sul sito di legame della glicina del recettore NMDA.

In sintesi, lo squilibrio metabolico da noi documentato nelle urine, se presente anche nel sistema nervoso centrale, potrebbe favorire il potenziamento dello stress ossidativo ed il ben noto squilibrio eccitazione/inibizione presente nell'ASD, che contribuisce allo sviluppo di convulsioni in circa il 20% degli individui autistici.

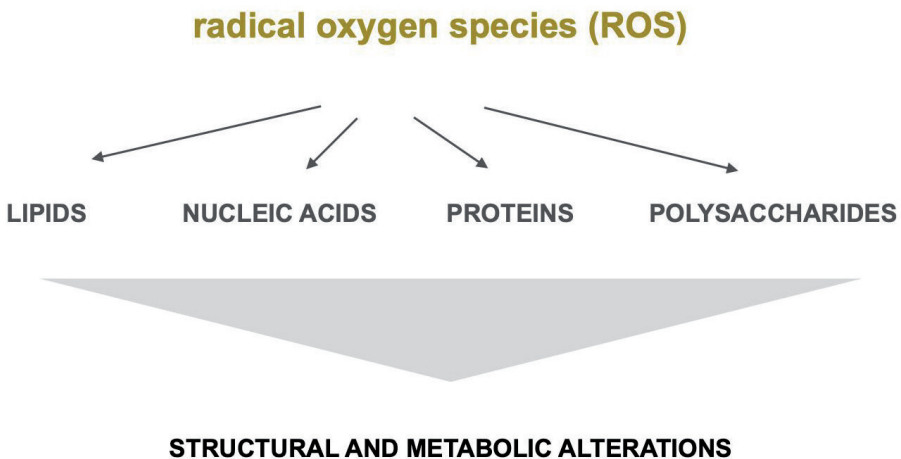
## Luca Proietti De Santis

*Docente di Genetica dell'Invecchiamento, Laboratorio di Genetica Molecolare dell'Invecchiamento, Dipartimento di Scienze Ecologiche e Biologiche, Università degli studi della Toscana*

### **Le variazioni dell'espressione dei geni della riparazione del DNA e dei geni di risposta allo stress ossidativo come biomarcatori di esposizione ai metalli pesanti: progetto di ricerca presso il Termovalizzatore di San Zeno – Arezzo**

In questo progetto, abbiamo introdotto l'analisi di alcune componenti molecolari del DNA e del suo metabolismo, per potenziare i sistemi di indagine utilizzati a tutela della salute dei lavoratori. Nello specifico valuteremo l'eventuale esposizione ai metalli pesanti dei lavoratori del termovalizzatore di Arezzo e il potenziale effetto nocivo sull'organismo analizzando marcatori biologici. Prima di entrare nel merito avrei piacere di presentarvi alcune diapositive che illustrano cos'è lo stress ossidativo e perché è tanto pericoloso per il nostro organismo.

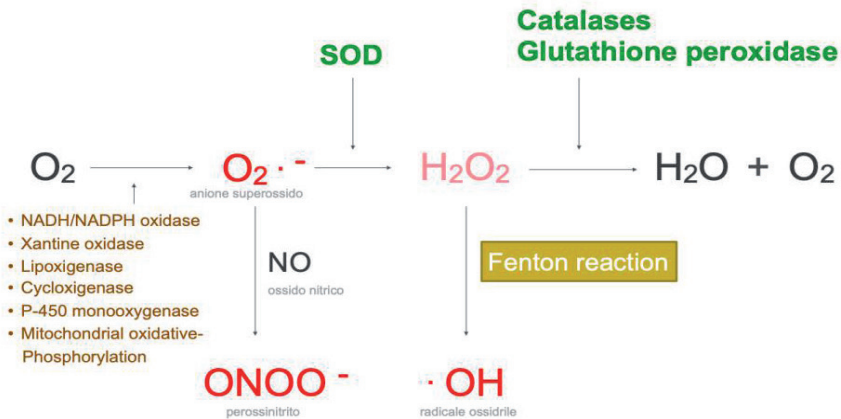
I radicali liberi sono degli atomi particolarmente avidi di elettroni capaci pertanto di formare dei legami di tipo covalente con atomi appartenenti ad altre molecole incluse le macromolecole biologiche quali gli acidi nucleici (DNA e RNA), le proteine, i polisaccaridi ed i lipidi.



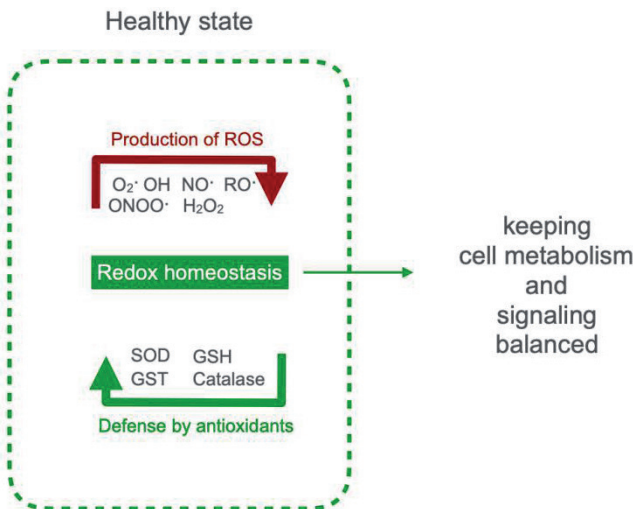
Quando queste macromolecole subiscono tali modificazioni la loro funzione può risultare alterata e determinare uno squilibrio metabolico. La fonte maggiore di produzione di radicali liberi è endogena ed è rappresentata dalla fosforilazione ossidativa mitocondriale. Questo processo, che si svolge all'interno dei mitocondri, è essenziale per il nostro organismo in quanto è il processo attraverso il quale la cellula immagazzina energia sotto forma di ATP. Tale processo prevede un trasporto di elettroni, tra vari atomi e molecole, che però determina dei sottoprodotti di re-

azione dell'ossigeno: infatti i radicali liberi sono atomi o molecole di ossigeno. In primis si forma una molecola chiamata anione superossido che è un radicale libero molto tossico perchè appunto in grado di attaccare le macromolecole biologiche. La cellula però è dotata di sistemi di detossificazione rappresentati da molecole o enzimi capaci di riconvertire i radicali liberi in specie meno reattive.

### radical oxygen species (ROS)



Ad esempio l'anione superossido è convertito dall'enzima superossido dismutasi in perossido di idrogeno che successivamente viene convertito dalla catalasi e dalla perossidasi in H<sub>2</sub>O e O<sub>2</sub>, due molecole sicuramente non tossiche. Però può succedere che durante queste reazioni, il perossido d'idrogeno venga a contatto con alcuni metalli inclusi i cosiddetti metalli pesanti dando vita alla cosiddetta "reazione di Fenton" durante la quale il perossido d'idrogeno viene convertito nella specie radicalica più tossica per l'organismo, il radicale ossidrilie. Quest'ultimo è la specie radicalica più reattiva. Inoltre non esiste un enzima in grado di convertirla in una forma meno reattiva.



Ad ogni modo come appena detto la produzione di radicali liberi fisiologica è tamponata da un sistema di detossificazione che va a tamponare la loro produzione.



In condizione fisiologiche, pertanto, la cellula permane in una situazione di equilibrio detta omeostasi redox.

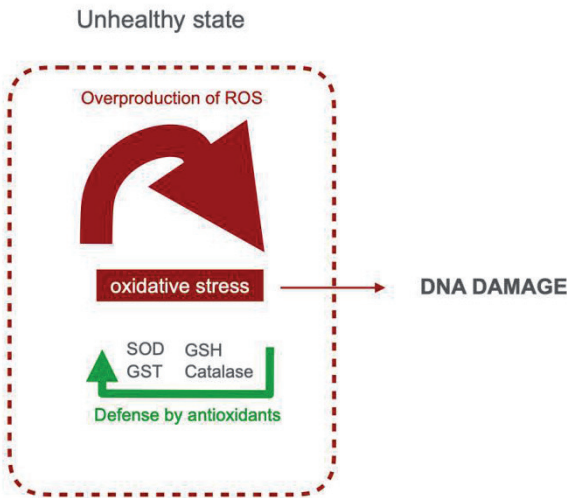
Può succedere però in alcune circostanze come, un processo di infiammazione, un intenso sforzo muscolare oppure a seguito di esposizione ad agenti inquinanti

come i metalli pesanti, può avvenire una intensa produzione di ROS che non è tamponata dal network di de-tossificazione costituito dalle molecole e dagli enzimi antiossidanti.

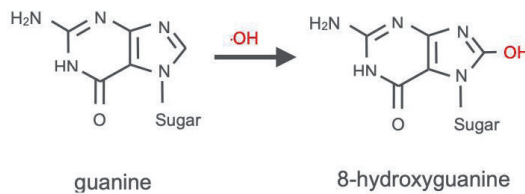
In questo caso la cellula viene a trovarsi in una situazione di “stress ossidativo”, caratterizzato da una produzione di radicali liberi maggiore della capacità antiossidante della cellula stessa.

Come già detto i radicali liberi possono legarsi a macromolecole. Ma poco importa se ad essere attaccate sono le proteine, i carboidrati o gli

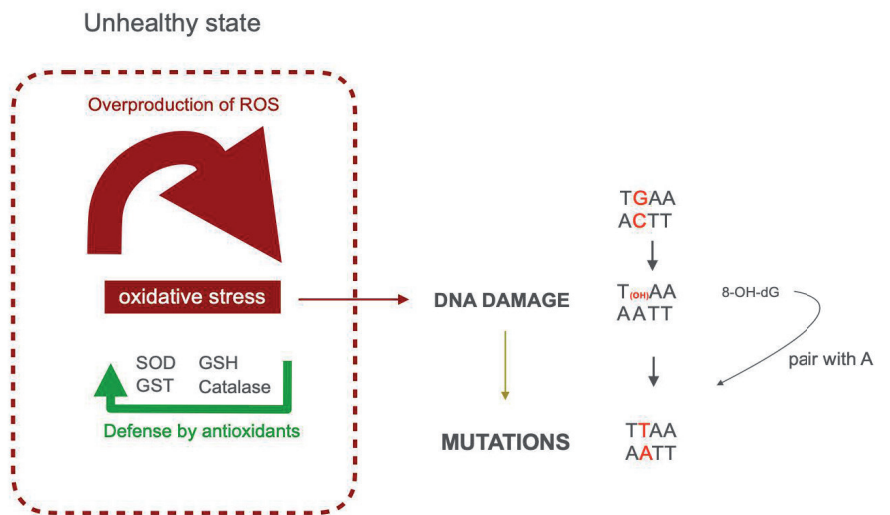
acidi grassi perché queste molecole possono essere sostituite producendone di nuove. Il problema si presenta quando il danno riguarda il DNA perché come vedremo il rischio è che venga persa in maniera permanente l’informazione genetica con eventi che possono essere catastrofici. Il radicale che abbiamo prima descritto infatti (radicale ossidrile), è capace di ossidare una delle basi azotate del DNA, la guanina, trasformandola in 8-idrossiguanina.



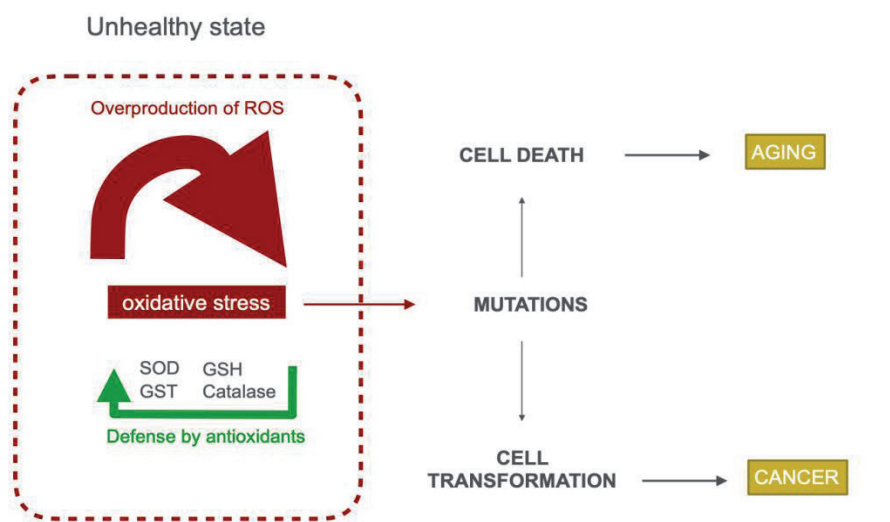
### DNA DAMAGE



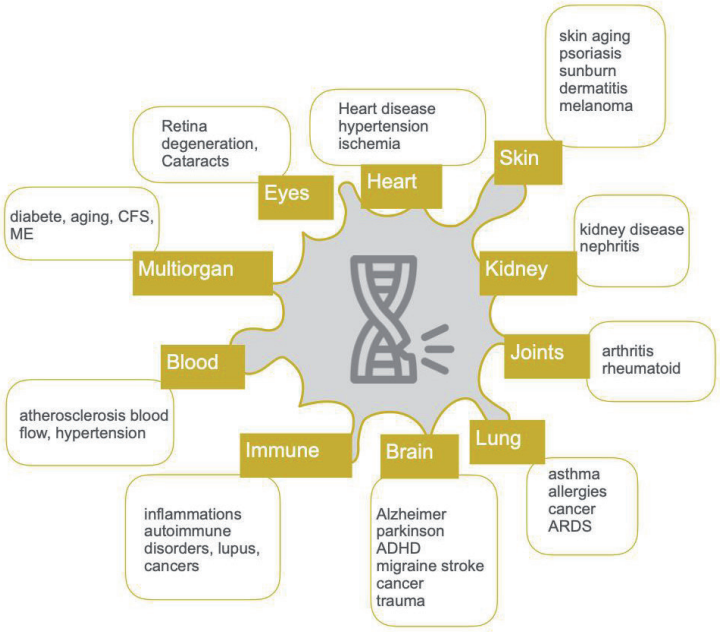
Perché questa trasformazione è così pericolosa? Il processo di replicazione del DNA si basa sulla complementarietà dell’elica, ossia durante il processo di duplicazione la timina si lega obbligatoriamente all’adenina e la guanina alla citosina e questo permette durante la replicazione del DNA, che avviene quotidianamente nelle nostre cellule, che l’informazione genetica venga conservata in maniera esatta (obbligata). Se però è presente una forma di guanina ossidata, viene riconosciuta dal sistema come timina per cui le si assocerà un’adenina invece che una citosina e nel successivo ciclo di duplicazione si avrà una sostituzione completa della coppia G-C in coppia T-A.



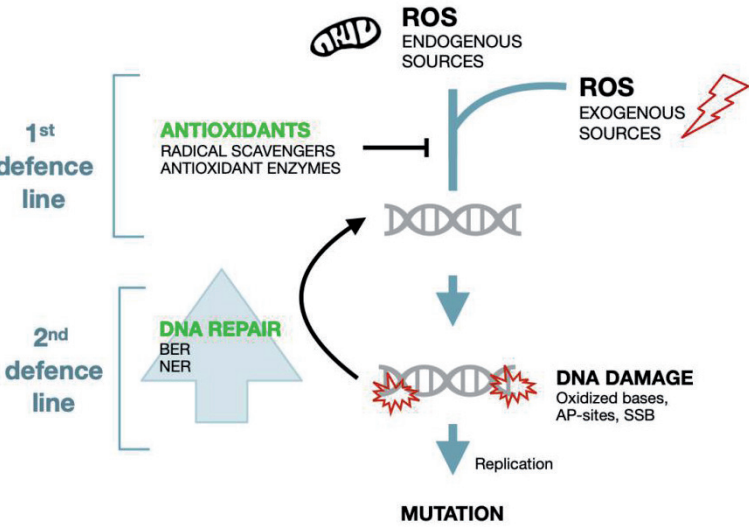
Questo determina la formazione di una modificazione permanente dell'informazione genetica (mutazione). Anche una singola mutazione del DNA può determinare esiti catastrofici: a) per la cellula, perché se la mutazione riguarda un gene essenziale per la funzione metabolica può determinare la morte cellulare. Se questo evento è cronico (quindi se l'organismo è esposto in maniera continua ad un'esposizione di tipo ossidativo) questo processo si enfatizza e numerose cellule vanno incontro a morte determinando una perdita di cellularità del tessuto e conseguente atrofia tissutale, quindi un danno organico che è la base del processo di invecchiamento.



Se la mutazione tocca dei geni coinvolti nella tumorigenesi (oncogeni o geni onco-repressori), può determinare la trasformazione della cellula normale in una cellula tumorale, e quindi avviare il processo neoplastico. In entrambi i casi pertanto la formazione di una mutazione genica può avere un esito catastrofico. In effetti mutazioni legate al danno ossidativo sono state collegate a patologie che riguardano tutti gli organi del nostro corpo incluse diverse forme neoplastiche.



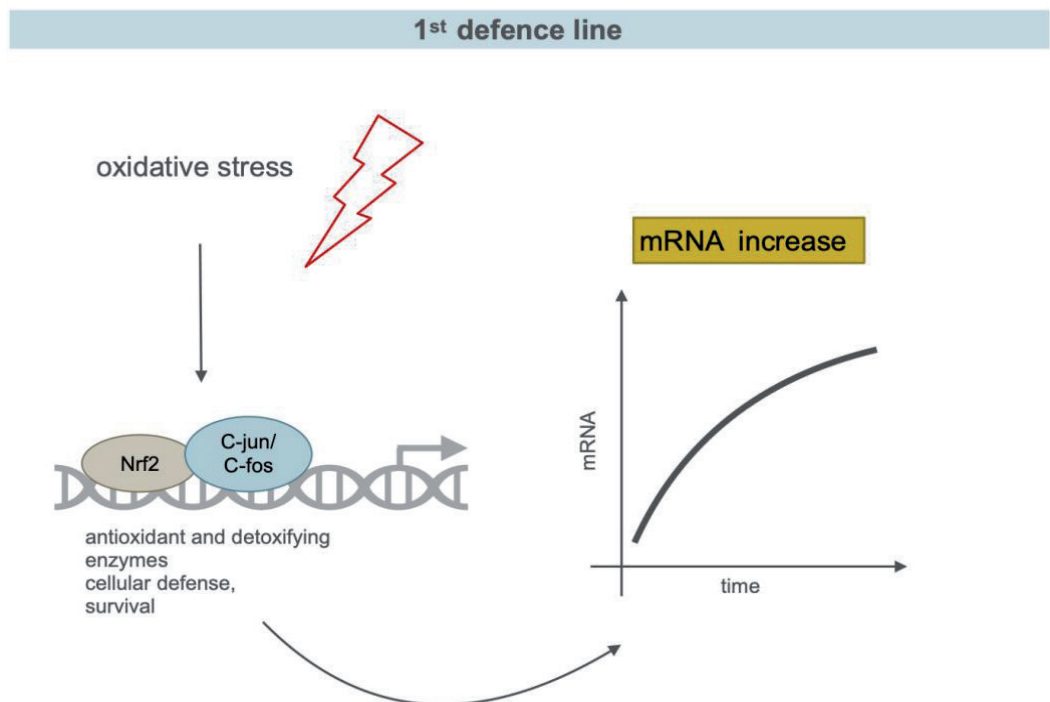
Però è anche vero che la cellula non accetta passivamente un eventuale attacco ossidativo. Essa può reagire attivando due linee di difesa.



Nel momento in cui c'è una sovrapproduzione di ROS, la cellula può attivare la prima linea di difesa potenziando la rete antiossidante mediante l'aumento di espressione di proteine di detossificazione al fine di tamponare l'aumento dei ROS.

La seconda linea di difesa è costituita dal sistema di riparazione del DNA: esistono degli enzimi capaci di riconoscere se la base azotata del DNA è stata ossidata dai radicali liberi e di ripristinare la base azotata normale prima che il processo di replicazione del DNA trasformi la lesione ossidativa in una mutazione permanente. Uno di questi è l'enzima OGG1 che riconosce l'8-idrossiguanina e la elimina dal DNA al fine di permettere il ripristino della base azotata corretta.

Noi possiamo misurare la risposta della cellula in caso di attacco ossidativo, andando a rilevare marker molecolari che ci indicano l'attivazione della prima e/o della seconda linea di difesa. Ad esempio in caso di attacco si attivano dei fattori di trascrizione che si associano ai promotori dei geni e ne determinano l'attivazione. Questo determina la produzione di specifici RNA messaggeri (mRNA) codificanti per proteine antiossidanti. Pertanto la misurazione della quantità di tali mRNA codificanti per proteine antiossidanti, ad esempio nelle cellule del sangue, indica in tempo reale che la cellula sta reagendo ad un attacco ossidativo.



Come abbiamo detto esistono enzimi (come OGG1) capaci di eliminare le basi azotate ossidate che saranno poi escrete con le urine. Pertanto misurare la presenza di basi azotate ossidate nell'urina indica in tempo reale che la cellula si trova in condizioni di stress ossidativo e sta riparando il DNA.

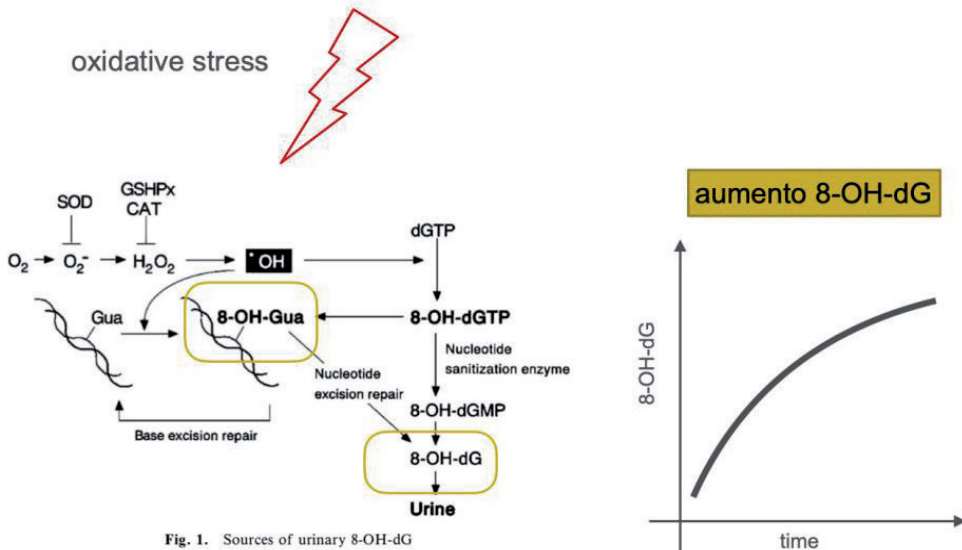


Fig. 1. Sources of urinary 8-OH-dG

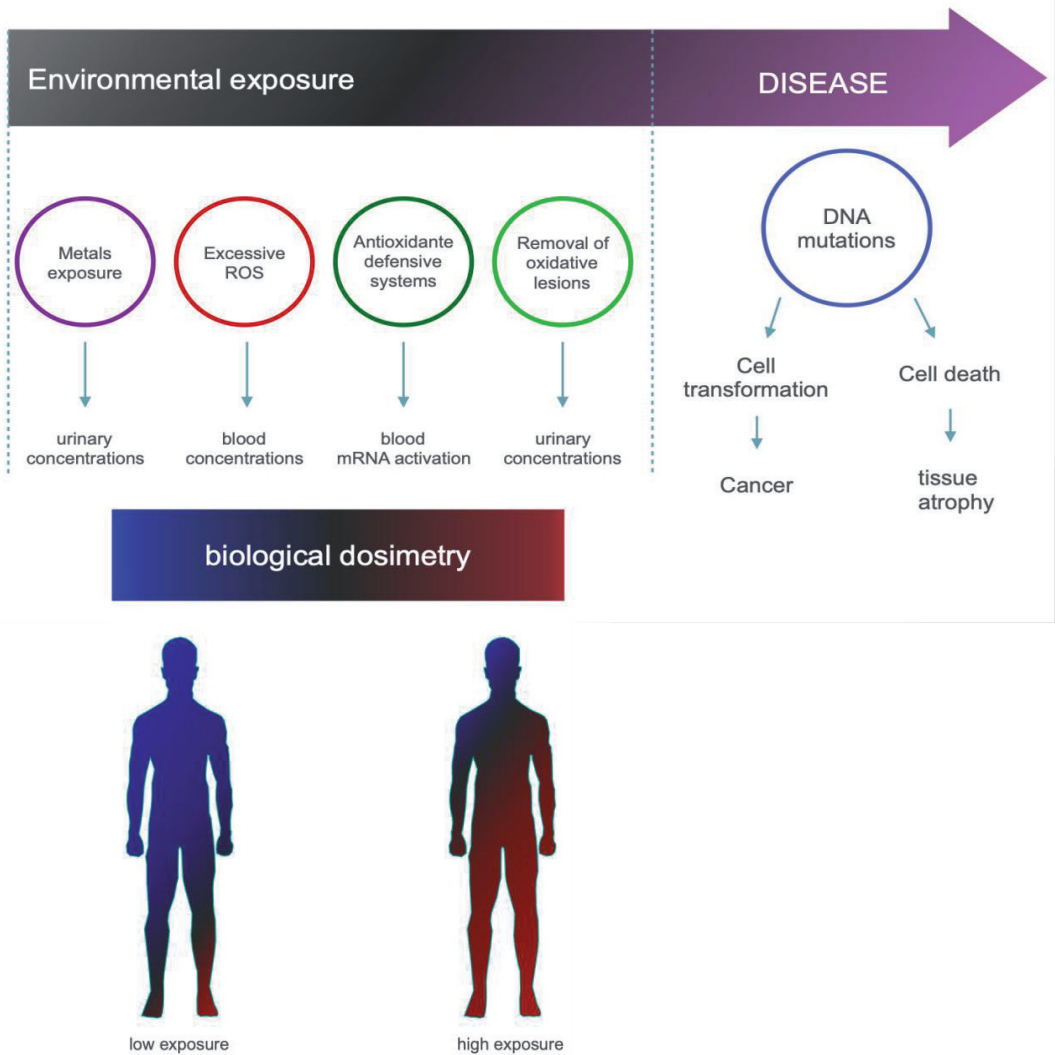
### DNA repair

Oltre a misurare la risposta della cellula all'attacco ossidativo possiamo anche analizzare l'effettiva esposizione ai metalli pesanti ricercandoli mediante spettrometria di massa nelle urine. Altresì possiamo misurare l'incremento dei ROS nelle cellule del sangue mediante saggi enzimatici (tipo Elisa). Infine è anche possibile analizzare le mutazioni presenti nel DNA delle cellule del sangue.

Riguardo quest'ultimo marker molecolare però dobbiamo fare un distinguo. Conoscere il profilo mutazionale del genoma è essenziale perché ci permette di comprendere se c'è un rischio maggiore di sviluppare determinate malattie, incluse le neoplasie. In tal caso le persone possono seguire dei programmi di screening mirati volti a diagnosticare l'eventuale insorgenza della malattia il più precocemente possibile.

Però l'insorgenza di una mutazione del DNA oltre ad essere un evento irreversibile, non è possibile datarla. Pertanto è impossibile collegarlo ad un determinato fattore o alla presenza di un agente di rischio nell'ambiente o nell'attività lavorativa. Al contrario, l'analisi dei primi 4 parametri (concentrazione di metalli urinari, quantità di ROS nelle cellule del sangue, attivazione dei sistemi di detossificazione e di riparazione del DNA nelle cellule del sangue e presenza di sottoprodotti di riparazione del DNA nelle urine), ci permette di analizzare la situazione in tempo reale, ovvero nel momento esatto in cui quegli eventi stanno accadendo. In tal modo possiamo effettuare **dosimetria biologica**, ossia utilizzare questi parametri per rilevare immediatamente un'eventuale esposizione ad agenti capaci di altera-

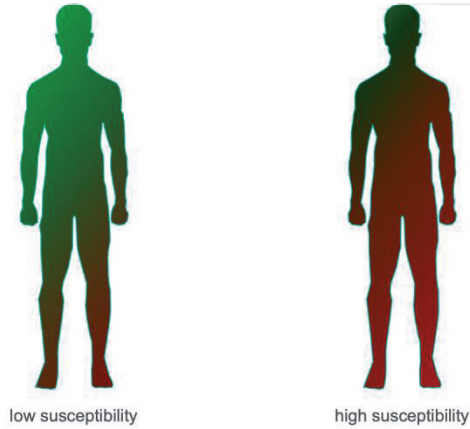
re l'omeostasi redox nelle cellule dell'organismo. Il vantaggio è quello di poter attuare programmi immediati di protezione nei confronti degli organismi esposti.



Questo tipo di approccio ci permette anche di individuare la diversa suscettibilità interindividuale poiché non tutti gli organismi rispondono con la stessa efficacia ad un eventuale insulto come può essere lo stress ossidativo.

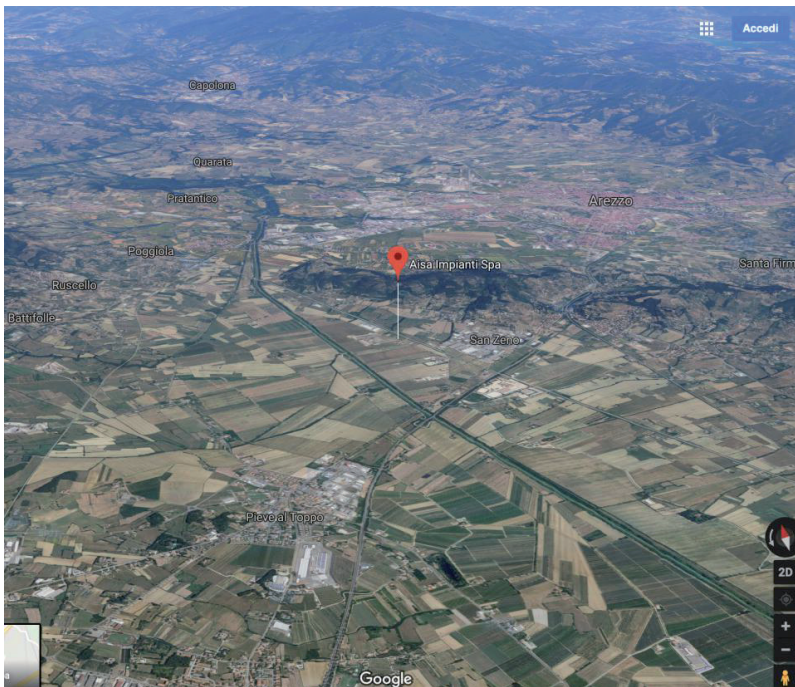
Anche in questo caso quindi si possono ipotizzare programmi di protezione nei confronti di questi individui.

## individual susceptibility

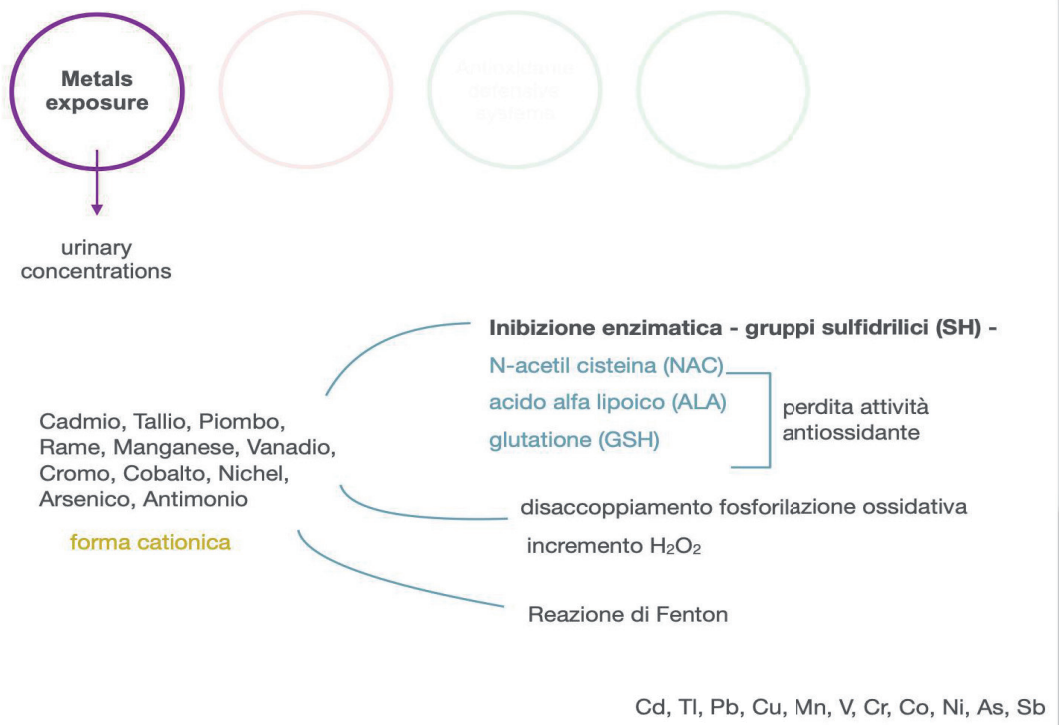


### **Il progetto del termovalorizzatore di San Zeno ad Arezzo.**

In questo progetto valuteremo l'eventuale esposizione ai metalli pesanti in 30 individui che lavorano presso il termovalorizzatore di Arezzo. Abbiamo anche individuato una popolazione di controllo costituita da individui che lavorano e vivono ad almeno a sei chilometri di distanza dal termovalorizzatore.



Misureremo: la concentrazione di metalli pesanti nelle urine. Focalizzeremo la nostra analisi su 11 metalli pesanti che sono quelli a cui sono maggiormente esposti gli operai degli impianti di incenerimento e smaltimento rifiuti. Tali metalli determinano un incremento dei ROS intracellulari per una serie di cause. Ad esempio è nota la loro capacità di legarsi ai gruppi sulfidrilici delle proteine o delle molecole antiossidanti inibendone la loro azione. Allo stesso tempo possono alterare la conversione dei radicali liberi in specie meno reattive durante il processo della fosforilazione ossidativa mitocondriale determinando un innalzamento della produzione di ROS.



Inoltre misureremo l'attivazione della risposta antiossidante. Nello specifico ci concentreremo su quattro enzimi analizzando l'espressione di OGG1 (enzima di riparazione capace di riconoscere le lesioni nel DNA) associato ad un aumento di stress ossidativo; la sovrapproduzione di MT1A che è capace di legare i metalli pesanti e portarli ai reni per la loro eliminazione; l'aumento di NQO1 E ST13 che partecipano alla rimozione di radicali liberi.

Altro parametro che analizzeremo mediante spettrometria di massa la presenza di base azotate ossidate nelle urine che come detto in precedenza indicano un attacco ossidativo al DNA e l'attivazione dei sistemi di riparazione deputati a ripristinare le basi azotate normali.



Ivan Arisi

Ricercatore Bioinformatico

Bioinformatics - Genomics Facility, Fondazione EBRI "Rita Levi Montalcini"

## Statistica e Machine Learning per dati ambientali

Vorrei iniziare con un'introduzione molto didattica alle metodiche di Machine Learning utilizzando come esempio dei semplici dati ambientali. I primi dati che voglio mostrare sono quelli di una centralina di rilevamento inquinanti dell'ARPA e ho scelto come esempio gli anni 2011-2012. Questi dati sono liberamente disponibili online.



Dati: centralina ARPA, Bologna, Giardini Margherita  
O<sub>3</sub> (Ozono), NO<sub>2</sub> (Biossido di azoto)



Campionamento della «popolazione» delle concentrazioni di O<sub>3</sub> e NO<sub>2</sub> ogni giorno alle ore 16 ☺ serie temporale di concentrazioni (µg/m<sup>3</sup>)

Campioni	2011			2012		
	Data	O3	NO2	Data	O3	NO2
	1-gen-11	5	57	1-gen-12	5	57
	2-gen-11	4	59	2-gen-12	4	59
	3-gen-11	31	22	3-gen-12	31	22
	...	...	...	...	...	...
	...	...	...	...	...	...
	...	...	...	...	...	...
	29-dic-11	14	39	29-dic-12	14	39
	30-dic-11	34	30	30-dic-12	34	30
	31-dic-11	41	29	31-dic-12	41	29

Campioni = Osservazioni

In clinica  
Campioni = casi = pazienti

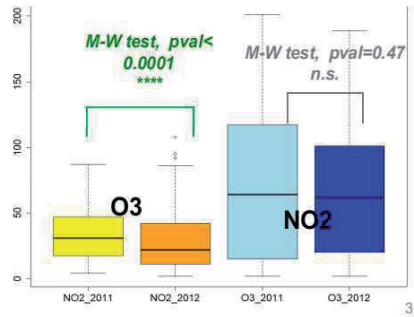
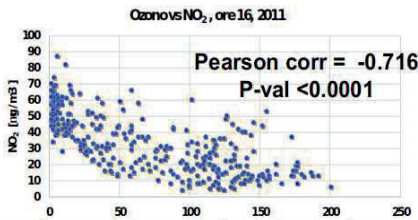
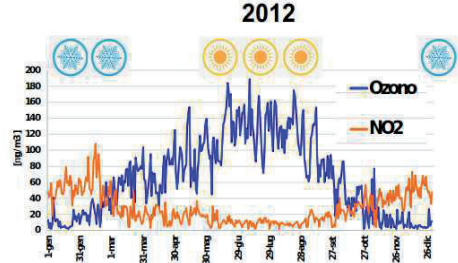
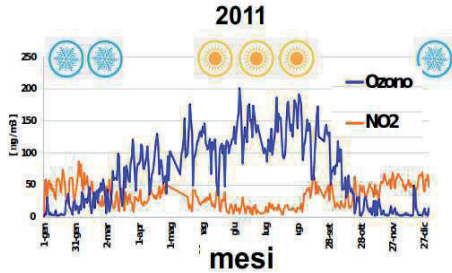
[Fonte dati: <https://dati.arpae.it/dataset/qualita-dell-aria-rete-di-monitoraggio>]

1

Questa centralina di Bologna misura ozono e biossido di azoto, quindi siamo in presenza di una serie temporale di concentrazioni in cui ho campionato quello che si chiama "popolazione" in termini statistici, cioè l'insieme delle infinite concentrazioni nel tempo di questi inquinanti in quel particolare luogo. Sulle colonne ho le variabili e sulle righe i campioni. Naturalmente con un primo approccio, quello di statistica standard, guardo i dati e mi accorgo, come è ben noto, che l'ozono è molto più alto in estate ed è antiparallelo al biossido di azoto che è invece più alto in inverno. Le due variabili in termini quantitativi sono piuttosto anti-correlate,

quindi posso fare naturalmente dei test in cui posso attribuire una significatività alle correlazioni e qui con un test parametrico posso vedere che vi è una differenza significativa fra le concentrazioni di ozono tra il 2011 e il 2012. Diciamo che questa è una statistica semplice, standard.

## Statistica descrittiva e inferenziale sui dati



	2011			2012		
	Media	Mediana	DevStd	Media	Mediana	DevStd
Ozono	70.4	64.0	54.9	65.6	62.0	65.6
NO2	33.3	31.0	18.0	27.6	22.0	19.9

La statistica standard è basata pesantemente su dei modelli predefiniti. Cosa intendo per modelli? Se per esempio facciamo un T-test, dietro al T-test ho la distribuzione gaussiana e la distribuzione T, quindi delle funzioni ben determinate. Anche se faccio il test non-parametrico di Mann-Withney la distribuzione è ben nota: in questo caso è la distribuzione U.

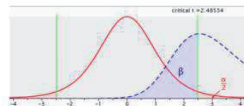
### La statistica inferenziale e descrittiva è basata su modelli

#### predefiniti

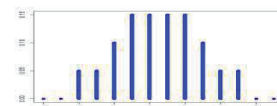
Media campionaria  $\bar{x} = \frac{1}{N} \sum_{i=1 \dots N} x_i$

Dev. standard  $s(X) = \sqrt{\frac{1}{N-1} \sum_{i=1 \dots N} (x_i - \bar{x})^2}$

T-test: distribuzioni gaussiana e T-student

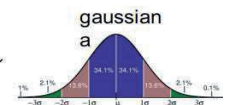


Mann-Withney test: distribuzione U



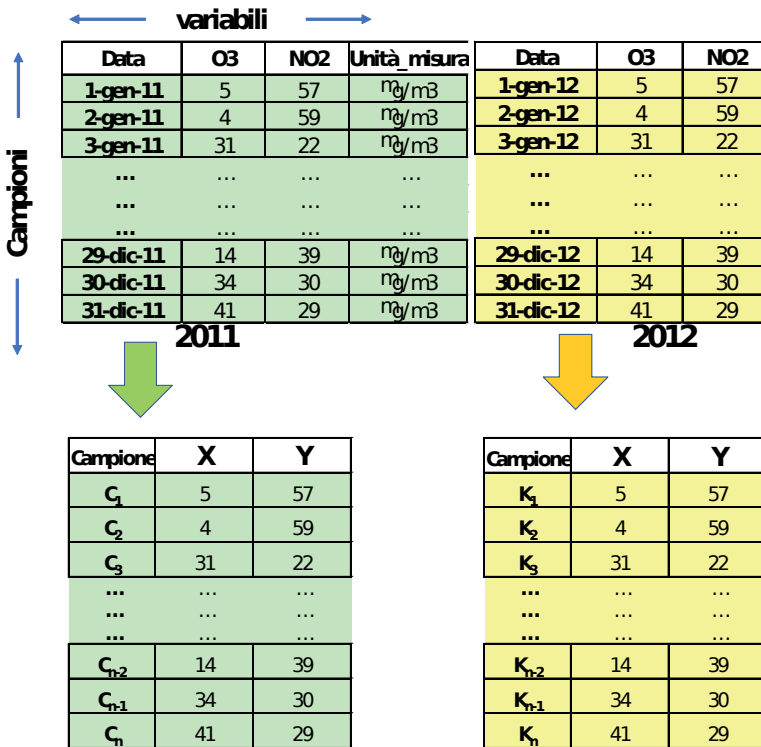
Pearson Corr =  $r = \frac{\sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{X})^2 \sum_{i=1}^n (y_i - \bar{Y})^2}}$

con  $z = \frac{1}{2} \log_e \left( \frac{1+r}{1-r} \right) \sim$



Anche per attribuire una significatività ad una correlazione, posso trasformarla e associarla ad una distribuzione gaussiana che, come si dice, è la madre di tutte le distribuzioni statistiche.

Per cambiare approccio, bisogna però riportare i dati in un modo un po' diverso, in una matrice in cui compaiano i campioni sulle righe e le variabili sulle colonne: l'ultima colonna la chiamiamo variabile "risposta" (in questo caso biossido di azoto) e il resto delle variabili "predittori".



X=variabile predittore  
(possono essere tante)

Y=variabile risposta

Cerco un **modello**:  
NO2 ~ Ozono  
Y ~ X

**Basato sui dati**

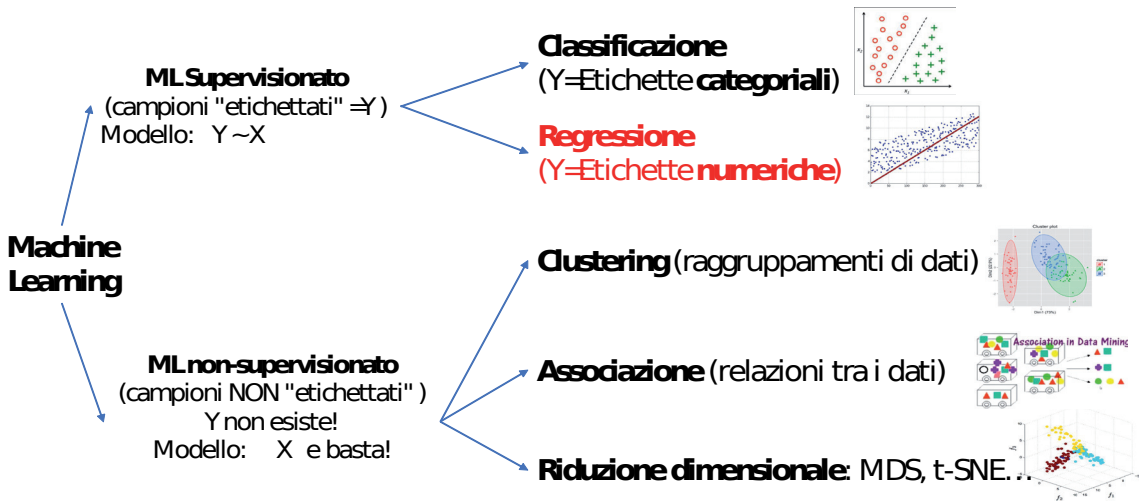
Vorrei cercare un modello in cui appunto la mia risposta  $Y$  è spiegata dalle “ $X$ ”, quindi un modello basato esattamente sui dati e definisco in modo operativo cos’è il “Machine Learning” (ML): è stato definito come un insieme di metodologie che permettono di costruire dei modelli e descrivere i risultati apprendendo dai dati, senza cioè fornire in modo esplicito e dettagliato una legge che colleghi le “ $X$ ” alle “ $Y$ ”.

Caratteristica del Machine Learning è che l’apprendimento migliora con l’esperienza, il che in termini informatici significa fornire dei nuovi dati al sistema.

### Definizione operativa di Machine Learning (ML)

Un insieme di metodologie, che permettono al computer di costruire **modelli** unicamente apprendendo dai **dati**, senza fornire esplicitamente tali modelli.

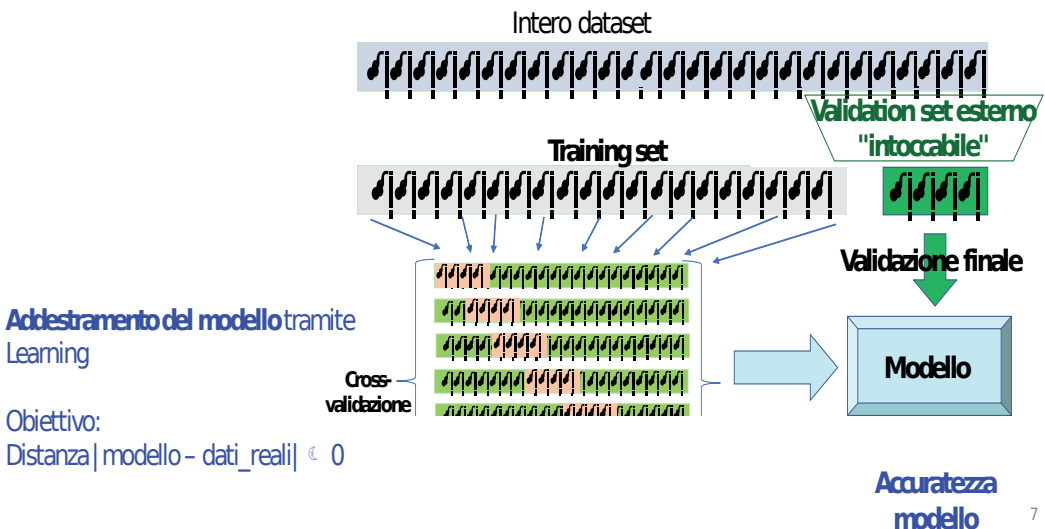
**I modelli migliorano fornendo nuovi dati ("esperienza") al sistema.**



Posso dividere l'ambito del Machine Learning in due grandi macro-ambiti: il Machine Learning non supervisionato in cui essenzialmente non ho la colonna delle X, la variabile risposta non ce l'ho e quindi non ho dati etichettati, posso in quel caso raggruppare i casi, posso trovare associazioni, ma la cosa più importante è l'attività di Clustering, di raggruppamento; il Machine Learning supervisionato, dove ho quindi un'etichetta, una variabile risposta. Nel caso che la risposta sia categoriale (ghiaccio, nebbia, sole e pioggia) posso appunto fare una classificazione e quindi costruire un modello che classifichi i miei campioni. Oppure posso fare quello di cui parlerò oggi, una regressione: se questa variabile risposta è un numero come nel nostro caso in cui è una concentrazione quotidiana di biossido di azoto. In generale un modello mi serve per interpretare meglio i fenomeni, per stimare nei dati ambientali dei valori mancanti nella matrice dei predittori e soprattutto per fare previsioni.

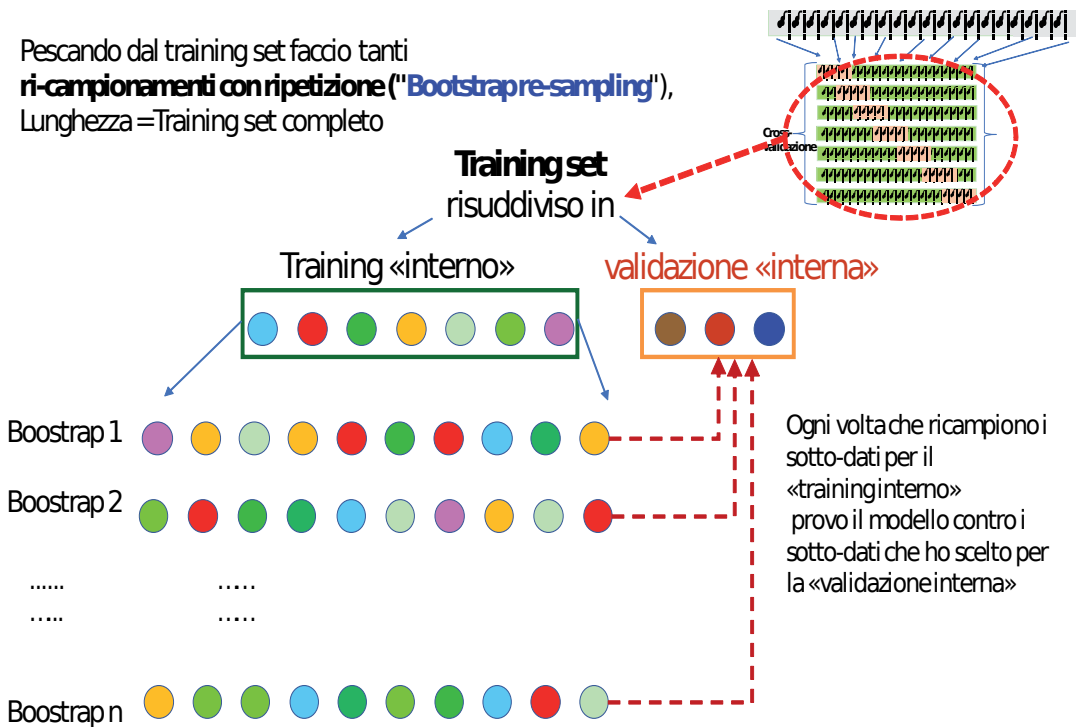
Come funziona il Machine Learning?

Divido i miei dati matriciali in due parti: una parte più grande ("training set") che mi serve per l'addestramento del modello, e una parte più piccola ("validation set") che non tocco e userò alla fine per validare il mio modello.



Naturalmente l'obiettivo a livello statistico e informatico è che la distanza tra la previsione del modello e i dati reali tenda a zero, che sia la più piccola possibile. Questo è quello che si impone anche dal punto di vista numerico, e questo processo di apprendimento è basato sulla cross-validazione, cioè su un sistema di ricampionamento in cui ripesco tante volte i dati dal mio insieme di addestramento suddividendoli ulteriormente all'interno, in una parte di addestramento, che ricambio tante volte con ripetizioni (quindi posso avere anche dei dati doppi), e una piccola parte di cross-validazione.

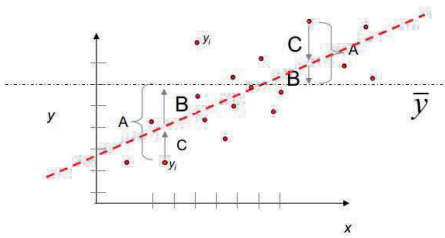
Ogni volta che campiono il dataset, costruisco un primo modello che andrò a testare contro questo piccolo set di validazione interna, lo ricampiono e continuo il processo finché arrivo ad un punteggio soddisfacente.



Con questo ricampionamento con ripetizione (detto "bootstrap") faccio finta che in un qualche modo il mio set di addestramento sia una nuova popolazione quindi ripesco tante volte da lì.

Il modello più semplice di machine learning è la retta, la regressione lineare in cui si dà una forma funzionale, però i due parametri A e B sono ignoti.

Il tipo più semplice di ML: **Regressione lineare** 2 soli parametri **a, b**  
**Modello:**  $\text{NO}_2 \sim \text{Ozono}$   $\text{NO}_2 = a \cdot \text{Ozono} + b$  **2011={Training set}**  
**2012={Validation set}**



$\hat{Y}_i =$  fitted Y values  $= ax_i + b$

$$\sum_{i=1 \dots N} (y_i - \bar{y})^2 = \sum_{i=1 \dots N} (\hat{y}_i - \bar{y})^2 + \sum_{i=1 \dots N} (\hat{y}_i - y_i)^2$$

**A<sup>2</sup>**  
**SS<sub>total</sub>**  
 Total Squared distance of observations from mean of y

**B<sup>2</sup>**  
**SS<sub>reg</sub>**  
 Distance from regression line to mean of y. Variability due to x

**C<sup>2</sup>**  
**SS<sub>residual</sub>**  
 Variance around the regression line. Additional variability not explained by X

2

Misuro l'accuratezza del modello con:

$$R^2 = \% = \frac{\text{Var\_spiegata\_modello}}{\text{Var\_totale}}$$

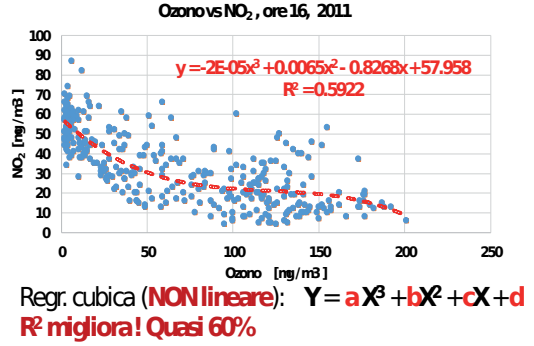
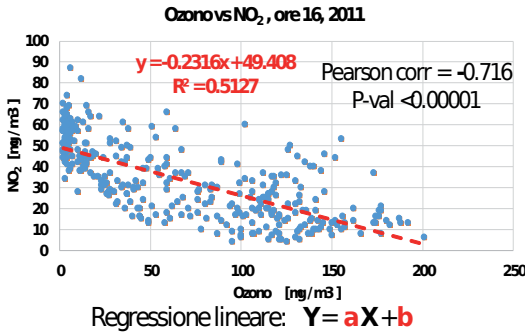
$$\begin{aligned} R^2 &= B / A = 1 - C / A \\ &= \text{SS}_{\text{reg}} / \text{SS}_{\text{total}} \\ &= 1 - \text{SS}_{\text{RES}} / \text{SS}_{\text{total}} \end{aligned}$$

$R^2 \ll 1$  significa che il modello spiega interamente i dati, ma non succede quasi mai e di solito è un problema, si parla di **Overfitting**

Ho scelto il caso di questa centralina come se fosse l'insieme di addestramento e il 2012 il mio insieme di miei dati di validazioni, quelli che voglio prevedere. Come misuro l'accuratezza? Come do un punteggio a questo modello? Con il famoso R2 (questo posso usarlo in generale per qualunque tipo di modello), spiego quanto il mio modello è in grado di descrivere la variabilità totale del sistema.

Naturalmente più è vicino ad 1 meglio è, ma se è proprio 1 è un problema ed è quello che si chiama "overfitting", in cui ho un modello troppo preciso e schiacciato su questi dati, e non va bene perché non è qualcosa di realistico, non si può applicare in generale ed è qualcosa da cui gli informatici si guardano. Con la regressione lineare non ho una grande performance e arrivo al 51%, però posso usare questo primo modellino per colmare i vuoti e quindi stimare i dati mancanti. Se faccio poi una regressione un po' più complessa, provando con una polinomiale di terzo grado, una regressione cubica, questa volta arrivo ad un R2 del 60%.

A questo punto valido il mio modello: qui vedete che in rosso mattone ho la previsione del modello cubico e in giallo ho il biossido d'azoto reale e diciamo quindi che per l'anno 2012 riesco a spiegare il 55% della varianza con questo modello polinomiale molto semplice.



Uso la regressione lineare anche per stimare i dati mancanti.

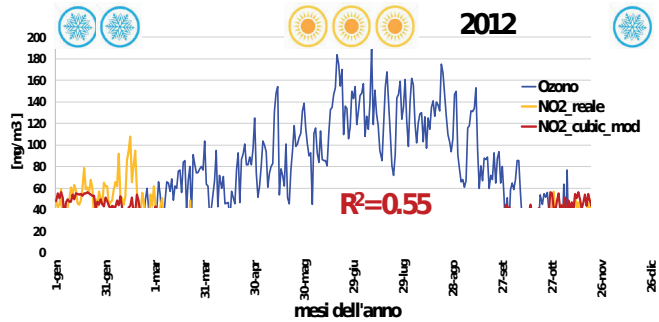
La regressione cubica ottiene risultati migliori

Dopo aver addestrato il modello sui dati 2011.

faccio una

previsione di NO<sub>2</sub> per il 2012 (validation set)

**R<sup>2</sup> = 0.55** € il modello cubico spiega il 55% della variabilità



Adesso complico un po' i dati: uso un'altra centralina che misurava ben otto inquinanti.



**Dati:** altra centralina ARPA,

Reggio Emilia, viale Timavo

Misure: **PM10, NO<sub>2</sub>, NO, NO<sub>x</sub>, CO, Benzene, Toluene, Xylene**

**Campionamento** di n=8 inquinanti (variabili), medie giornaliere

2015-2017: Training set

Giorno	X							Y
	PM10	NO <sub>2</sub>	Nox	CO	Benzene	Toluene	NO	Xylene
Campione	X1	X2	X3	X4	X5	X6	X7	Y
1-gen-15	56.00	54.29	175.75	1.01	2.87	6.41	79.33	2.13
2-gen-15	62.00	61.96	245.50	1.43	3.89	9.32	119.9	3.23
3-gen-15	55.00	53.17	152.52	1.12	3.23	6.62	65.09	2.08
...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...
29-dic-17	4.00	52.39	150.04	0.86	1.88	4.47	63.83	1.25
30-dic-17	42.00	52.74	153.17	0.93	2.26	5.02	65.87	1.34
31-dic-17	42.00	39.13	101.87	0.84	2.20	3.61	41.04	0.89

2018: Validation set

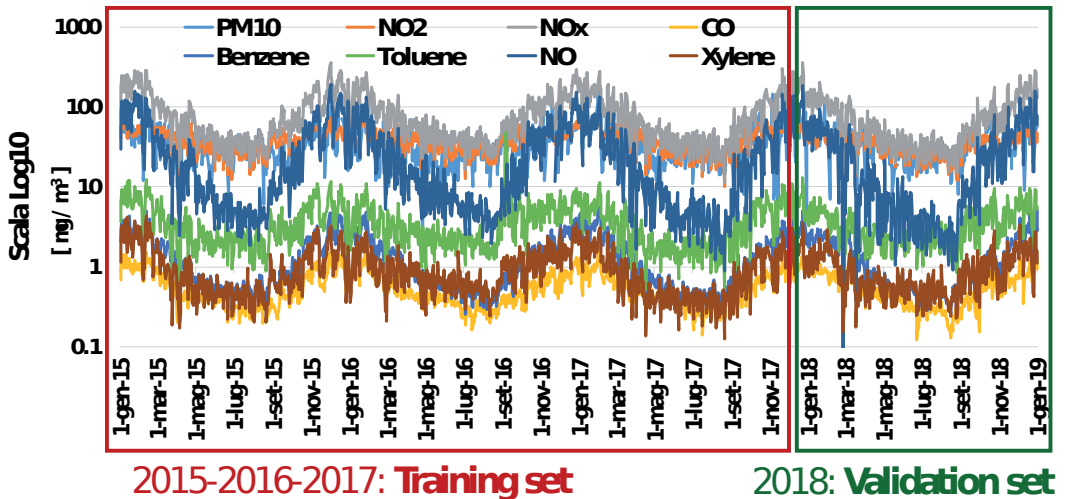
Giorno	X							Y
	PM10	NO <sub>2</sub>	Nox	CO	Benzene	Toluene	NO	Xylene
Campione	X1	X2	X3	X4	X5	X6	X7	Y
1-gen-18	49.00	35.58	100.00	0.92	2.61	3.21	42.18	0.74
2-gen-18	62.00	45.29	103.92	0.77	1.73	3.23	38.2	0.84
3-gen-18	40.00	53.48	157.19	0.96	2.38	5.87	68.2	1.60
...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...
29-dic-18	56.00	44.00	164.75	0.96	3.22	5.93	78.9	1.69
30-dic-18	57.00	46.50	166.54	0.93	2.85	5.52	78.4	1.53
31-dic-18	59.00	36.00	128.00	1.20	5.10	5.30	60.0	1.10

[Fonte dati: <https://dati.arpae.it/dataset/qualita-dell-aria-rete-di-monitoraggio>]

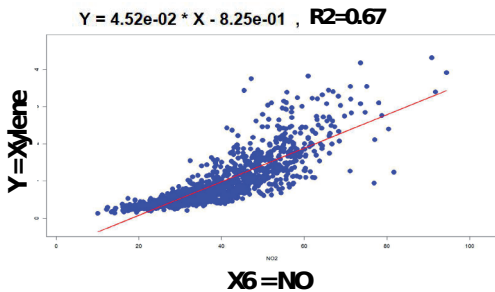


Questa volta scelgo come variabile risposta lo xilene e tutte gli altri microparticoli li considero variabile predittrici. Per l'insieme di addestramento ho scelto come campionamento le medie giornaliere, giorno per giorno, dal 2015 al 2017 e il 2018 è l'insieme di validazione, quello per cui voglio costruire un modello che più si avvicina, appunto, a come sono andati i dati l'anno scorso.

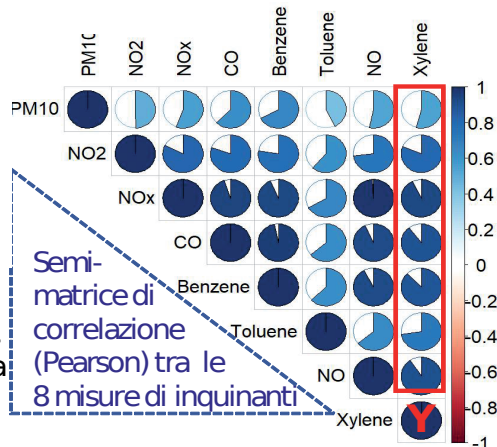
Tutti i gas hanno andamenti ciclici con periodo annuale



Naturalmente i dati hanno andamento ciclico su base stagionale (questo è su scala logaritmica) e alcuni di loro sono piuttosto correlati: queste che vedete sono matrici di correlazione tra questi inquinanti che corrispondono ad uno scatter plot che ha un trend più o meno lineare, però non mi voglio accontentare facendo il modello dello xilene in funzione di uno solo di questi inquinanti.



Queste variabili sono correlate tra loro, ma cerco un modello più flessibile della regressione tra 2 sole variabili



Notazione: "+" NON è algebrico

**Xylene** ~ PM10 + NO2 + NOx + CO + Benzene + Toluene + NO

**Y** ~ X1 + X2 + X3 + X4 + X5 + X6 + X7

Allora cerco un modello in cui lo xilene dipenda da tutto il resto e questa  $Y \sim \dots$  non è una notazione algebrica: si indica così il modello che dipende da tutte queste cose, da tutte le altre variabili.

Allora provo un altro modello, introduco gli alberi decisionali che sono utilizzati in tanti altri campi, anche nella pratica clinica.

Per creare un modello più complesso che:

- > includa tutte le variabili,
- > Integri la non-linearità del sistema

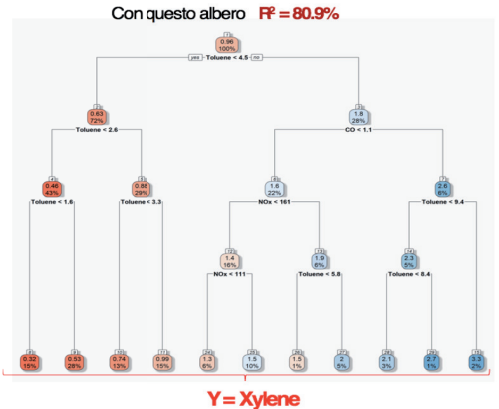
Scelgo un metodo di regressione basato su

**Alberi decisionali**  
 ("decision tree")

Una struttura a rami ("branches"), decisioni binarie:  $X_1 \leq \dots$ ?  $X_4 < \dots$ ? Sì o No?

Ipotizziamo che ogni campione sia un "gettone"= $\{X_1, X_2, X_3, X_4, X_5, X_6, X_7\}$  che cade. Lo infilo in cima ("root node"), ad ogni ramificazione ("decision node") il gettone sceglie destra o sinistra in base ai valori delle variabili fino a raggiungere una foglia (leaf) in fondo.

Le foglie sono i valori di Y stimati dal modello



Questo è un albero decisionale ed è un albero al contrario in cui in ogni biforcazione ho una decisione binaria: immaginate che ognuno dei miei campioni, in questo albero, sia un gettone che contenga le mie sette variabili (monossido di azoto ecc.). Ad ogni biforcazione questo gettone deciderà se infilarsi a destra o a sinistra in base alla domanda: esempio: "toluene < 4.5", se sì andrà a sinistra, se no andrà a destra; "monossido di carbonio < 1", se sì andrà a sinistra se no andrà a destra, finchè non arrivo alle cosiddette "foglie" che non contengono più domande ma i valori numerici finali.

Lo spazio dei valori di Y è partizionato dall'albero.

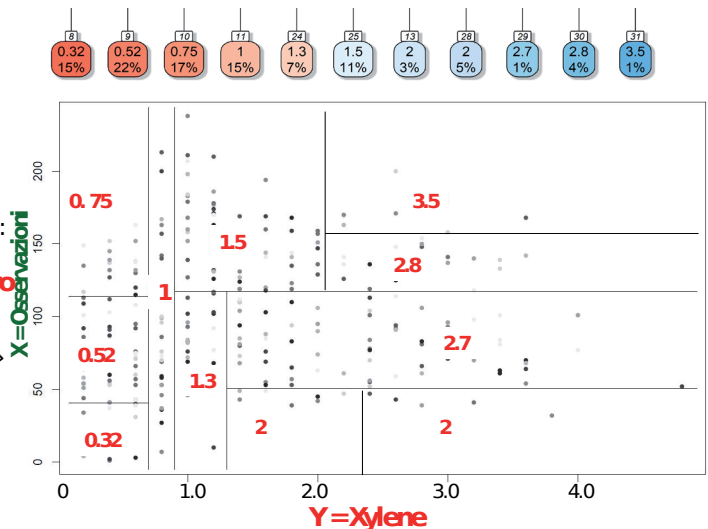
**Ogni foglia corrisponde a un riquadro** di valori di Y.

Per ogni osservazione  $\{X_1, X_2, X_3, X_4, X_5, X_6, X_7\}$  il modello a singolo albero calcola:

$Y = \text{media dei punti del riquadro}$

Domanda ovvia: «solo 11 valori possibili per Y?»

Risposta: «Certo, con UN solo albero, ma se ne uso molti...» «metodi "Ensemble"»



E questa è la mia previsione, le misure risultate dal modello con singolo albero decisionale. Cosa sono questi valori? In realtà sono delle medie, perché con un albero io ho solo partizionato lo spazio di tutti i valori possibili della variabile xilene, della variabile risposta, in tanti rettangoli; attraverso l'albero il mio gettone entra in uno di questi rettangoli e prende il valore medio dei punti che contiene. E questa è la previsione. Allora vi chiederete: è vero, è un modello semplice, però soltanto 11 valori per una variabile che ne potrebbe assumere tanti su un range continuo? A questo punto ne posso usare tanti e nascono i metodi *ensemble*. Uno di questi, molto usato in campo ambientale è il "random forest", una selva di alberi decisionali, in genere almeno 500 tutti diversi tra loro: così ottengo una risposta continua del modello e non solo per questi 11 valori.

Ogni albero è un modellino costruito come vi ho spiegato prima. Quindi ognuno di essi farà una previsione e alla fine farà una media del risultato di tutti alberi. Infilo il mio campione in cima a tutti gli alberi e alla fine faccio una media dei punti del rettangolo corrispondente: questa è una regressione fatta con random forest.

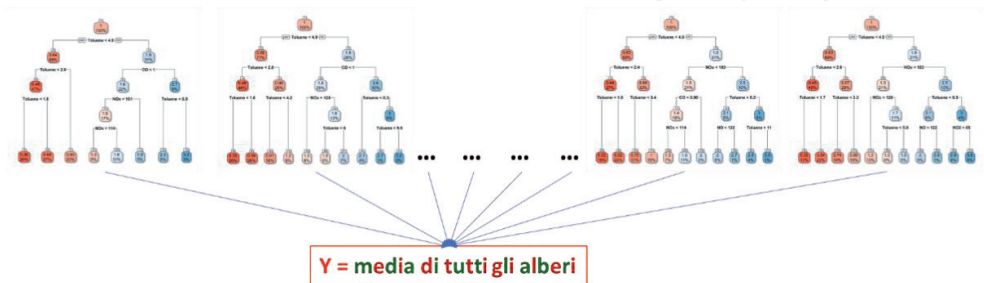
## Foresta Casuale ("Random Forest"): RF

La regressione è basata su una vera foresta di alberi decisionali rovesciati. Quanti?



Almeno  $n=500$  alberi in generale diversi tra loro  
Sono generati da **ricampionamenti casuali dei dati** così il modello diventa quasi una funzione continua, ed è in grado di descrivere comportamenti complessi. Servono **centinaia-migliaia di casi per addestrare il modello**.

Come combino gli alberi per la regressione ?



In un modello in cui sono stati usati 1000 alberi invece di 500, con i dati reali del 2018 dello xilene e quelli previsti siamo arrivati ad una performance dell'86%. Abbiamo migliorato molto ma riesco a prevedere un comportamento non lineare durante l'arco dell'anno?

Per paragone ho provato un modello molto più semplice, diciamo l'equivalente della regressione lineare multidimensionale, un "iperpiano" come si dice in geometria. Qui nel modello abbiamo una somma algebrica di tutte le variabili predittrici.

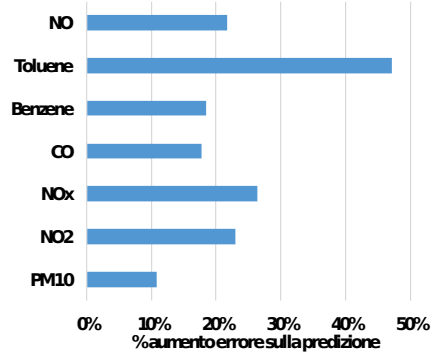
Posso interpretare facilmente il modello multi-lineare, attraverso il valore dei coefficienti

$$Y = -0.00028 \cdot X_1 + 0.0078 \cdot X_2 + 0.0068 \cdot X_3 + 0.0035 \cdot X_4 - 0.1076 \cdot X_5 + 0.075 \cdot X_6 + 0.036 \cdot X_7 - 0.14$$

I più "pesanti" sono qui nell'ordine: X5(benzene), X6 (toluene), X7 (NO).

Ma come interpretare la selva del **RF**? E' troppo **complesso**. Il modello è certamente più performante, ma pago il prezzo della ridotta decifrabilità. Ci sono migliaia di parametri che descrivono gli alberi. Questa difficoltà è vera in generale per tutti gli approcci avanzati di Machine Learning, in particolare per le "reti neurali".

Per i metodi ad alberi decisionali, posso però estrarre un equivalente dei pesi delle variabili ("**Variable Importance**"). Un sistema è eliminare una variabile per volta: più l'errore aumenta, più vitale risulta la variabile.



Con questo modello multilineare, vado abbastanza bene ma non riesco mai ad andare oltre l'83%, quindi non riesco a migliorare. Più cerco di arrivare alla soglia del 100% più il miglioramento è complicato. Quindi questo modello molto più complesso spiega meglio i dati, ma appunto uno dei problemi principali è che è molto complicato. Come posso andare ad interpretare il modello lineare? Andando a vedere i valori assoluti dei miei coefficienti: il peso dei miei coefficienti mi dice quali sono le variabili più pesanti del modello (in questo caso il benzene e il toluene).

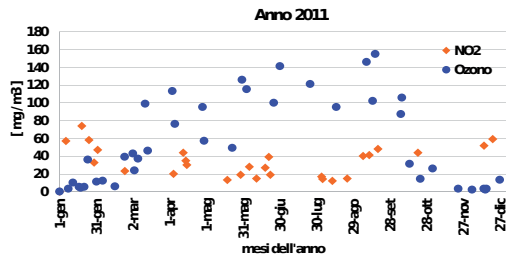
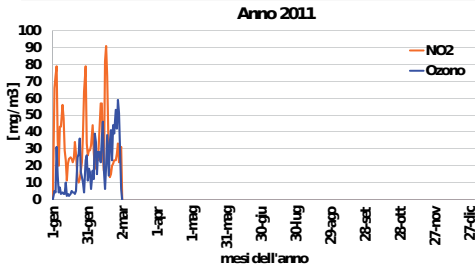
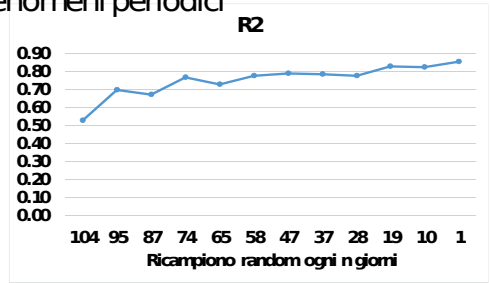
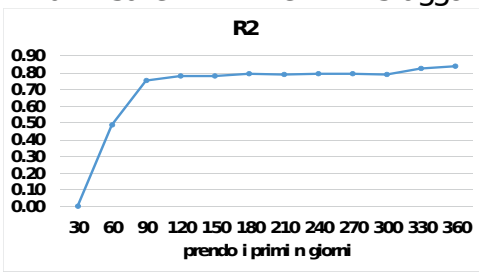
Come posso interpretare questi modelli ensemble con tutti questi alberi? Appunto con una "variable importance" che è già stata citata prima.

In generale, il "prezzo" di questi modelli Machine-Learning è che hanno delle ottime performance ma il tasso di interpretabilità è molto basso: migliore il punteggio ma riesco ad interpretarlo meno.

La "variable importance" è uno dei metodi ed è molto semplice: tolgo una variabile direttrice per volta, ricalcolo il modello e vedo quanto peggiora il punteggio. Più peggiora più quella variabile risulta fondamentale per il modello. Se non cambia nulla vuol dire che quella variabile non contava. In questo caso al primo posto ho il toluene che se lo tolgo peggiora del 50% la performance e al secondo posto ho gli ossidi di azoto.

Questo slide è solo per sottolineare l'importanza del sotto-campionamento: non posso fornire dati sotto-campionati, questo in generale, perché la performance risulta bassa, in questo caso avendo dei fenomeni ciclici e periodici piuttosto che campionare fitto fitto per 2/3 mesi è meglio campionare piuttosto una volta al mese ma distribuito lungo l'arco dell'anno (ogni 20 giorni) e si vede che le performance sono più alte perché questo naturalmente si porta dietro la periodicità del sistema.

Servono molti dati per avere un buon modello, **NON** posso **sotto-campionare** Meglio campioni rari e distribuiti nel tempo, piuttosto che “concentrati” €  
**anni** di misure... Altrimenti mi sfuggono i fenomeni periodici



Cito due lavori, il primo romano e il secondo statunitense, che hanno valutato la distribuzione spazio-temporale degli inquinanti su territorio nazionale, o sulle principali città. Perché spazio temporale? Io dei dati solo nel tempo ma tipicamente i dati ambientali ed epidemiologici sono nello spazio e nel tempo quindi sono molto più complessi.

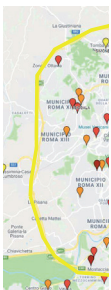
**Modelli nel tempo**  **e nello spazio** 

Invece di una sola centralina, raccolgo le misure dall'intera rete che copre:

**Una città (Roma)**

**Italia**

**USA**



**STAGE 3:**  
 calibrating PM with  
 AOD, meteorology  
 and land-use data



Avrò centinaia di serie temporali di dati da tutte le centraline, ognuno con le sue coordinate geografiche {Lat., Long.} € un grande **dataset spazio-temporale**

Come per le previsioni del tempo! Fonte: <https://www.cittadiniperlaria.org/no2-anno-roma-2018/>  
 Stafoggia M et al, *Environ Int.* 2019 Mar;124:170-179  
 Meng X et al, *Environ Int.* 2018 Dec;121(Pt 2):1137-1147

Chiaramente basta pensare ai dati di meteorologia: questi modelli sono basati sul *Machine-Learning* con il *random forest*. A vari stadi il random forest viene usato sia per colmare tutti i vuoti (come vedete in azzurro) o perchè le centraline non sono distribuite in modo uniforme sul territorio o perchè i dati satellitari mancavano a causa delle nuvole.

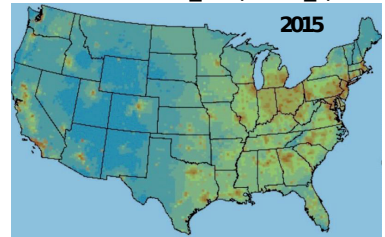
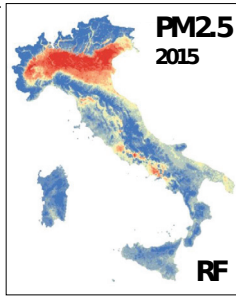
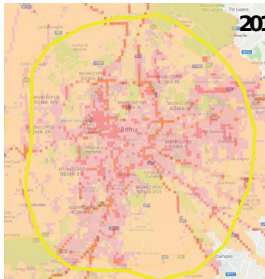
La distribuzione delle centraline è eterogenea, inoltre mancano molti dati spaziali ☹ è necessario interpolare per stimare tutti i valori mancanti. Sempre attraverso modelli RF.

Poi posso stimare i valori degli inquinanti con altri modelli RF che includano tutte le variabili temporali (mese/giorno) e spaziali (popolazione, altitudine, strade, industrie, meteorologia,...) ☹ **LURF (Land Use Random Forest)**

$$NO_2 \sim \text{Var\_temp} + \text{Var\_spaz}$$

$$PM_{2.5} \sim \text{Var\_temp} + \text{Var\_spaz}$$

$$PM_{2.5} \text{ Nitati} \sim \text{Var\_temp} + \text{Var\_spaz}$$



<https://www.cittadiniperlaria.org/no2-anno-roma-2018/>  
 Stafoggia M et al, *Environ Int.* 2019 Mar;124:170-179  
 Meng X et al, *Environ Int.* 2018 Dec;121(Pt 2):1137-1147

Quindi prima devo cercare di riempire i buchi della matrice dati e poi alla fine posso costruire un modello finale in cui, ad esempio in questo caso i monossidi di azoto o il microparticolato, dipendono però da tantissime variabili, ad esempio variabili meteorologiche, la concentrazione di inquinanti, la popolazione, l'asperità del terreno ecc... Quindi risulta essere un modello complesso.

Anche qua hanno estratto la variabile importance e hanno visto che la temperatura, la pressione barometrica, l'altitudine del luogo erano le variabili più pesanti nello studio condotto in Italia mentre nello studio condotto negli Stati Uniti per le concentrazioni di nitrato nel microparticolato le variabili erano: la distanza tra le centraline, la densità di popolazione del luogo.

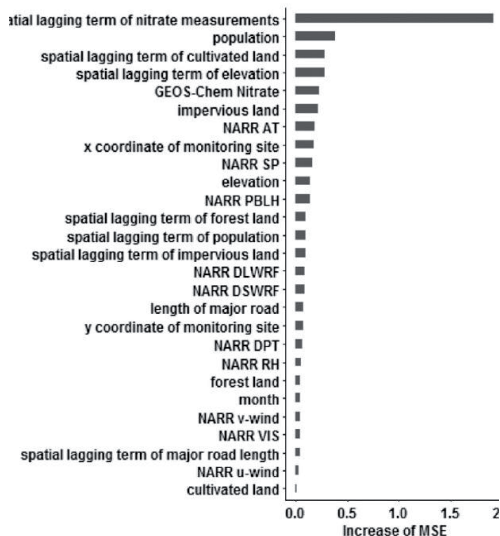
## “Variable Importance” dei predittori per:

### Italia: PM2.5

Relative importance (%) of the predictors in the stage 3 model for PM<sub>10</sub> and PM<sub>2.5</sub> (2013–2015).

Predictor	PM <sub>2.5</sub>			PM <sub>10</sub>		
	2013	2014	2015	2013	2014	2015
Air temperature	13.6	7.2	13.4	7.4	4.4	7.4
PBL (hh 00.00)	9.5	7.3	9.5	9.5	5.7	9.4
Julian day	9.7	10.7	9.2	1.8	8.9	1.8
Barometric pressure	7.9	12.9	7.8	10.2	11.5	9.9
Elevation	7.3	4.7	7.1	9.3	6.7	9.3
PBL (hh 12.00)	6.3	6.2	6.7	8.2	6.0	8.4
Wind (v component)	4.2	4.1	4.0	4.8	5.1	4.6
AOD (470 nm)	2.5	2.5	3.0	2.8	2.8	3.2
AOD (550 nm)	2.5	2.5	2.9	2.8	2.7	3.1
Month	2.9	3.0	2.7	5.1	2.5	5.0
Latitude	2.6	2.9	2.6	3.7	3.7	3.7
Administrative region	2.3	2.0	2.2	0.9	1.5	0.9
Precipitations	1.9	2.5	2.1	3.3	4.0	3.4
Longitude	2.2	1.9	2.1	2.1	1.9	2.1
Wind (u component)	2.0	2.4	2.0	2.6	2.6	2.6
Distance from sea	1.5	1.5	1.4	1.5	1.3	1.5
Resident population	1.4	1.4	1.4	1.5	1.5	1.6
Distance from emission points	1.4	1.9	1.4	1.3	1.4	1.2
Distance from highways	1.3	1.2	1.3	1.2	1.2	1.2

### USA: PM2.5 nitrati



## Conclusioni

### I metodi di Machine Learning:

- 🌐 sono ormai alla portata di tutti i PC
- 🌐 Sono estremamente flessibili e potenti
- 🌐 Riescono a descrivere comportamenti non lineari
- 🌐 Sono molto diffusi in epidemiologia e modellistica ambientale

⚠️ MA! necessitano di tanti [anni di...] misure ("BigData") per essere creati

**Grazie** alla SAGEN, all'ARPA  
e a tutti voi

**Gerry Melino**

*Direttore del Dipartimento di Biologia Molecolare, Facoltà di Medicina e Chirurgia Università degli studi di Roma “Tor Vergata”*

**Giuseppe Novelli**

*Rettore, Direttore U.O.C. Laboratorio di Genetica Medica, Università degli studi di Roma “Tor Vergata”*

## **Il rapporto tra ambiente e geni: dal rischio professionale alla medicina di precisione**

Questa che riporto è una copertina di “Science” che è uno dei tre giornali più importanti nell’ambiente, che ricorda come l’organismo umano può geneticamente ed epigeneticamente attivare dei processi per difendersi da questi insulti tossici. Vorrei spendere una parola a favore degli insulti tossici: se non ci fossero mutazioni non ci sarebbe evoluzione, però quando è troppo è troppo.



**Figura 1.** Copertina di Science del 7 ottobre 2013 (vol. 354; n. 6308) dedicata a tossici ambientali.

L’obiettivo non è “insulti tossici zero” ma capire che rapporto c’è tra i geni e l’ambiente e per ambiente non intendo solo quello che respiriamo, ma anche quello che mangiamo e le sostanze infettive che sono sia dentro di noi nel microbioma che nell’ambiente e quindi cercare di capire come gli insulti tossici che derivano da ambienti esterni quindi esogeni ma anche da sistemi endogeni. Lo stress replica-



tivo, quindi il meccanismo di divisione di una cellula, come il DNA viene duplicato, crea un numero di insulti tossici elevatissimo. Sia le sostanze endogene sia che quelle dell'ambiente quotidiano sono tutte quante riconosciute e ci sono dei meccanismi in cui il DNA danneggiato attiva dei processi che portano all'arresto dell'attività proliferative della cellula, ad alterazione del metabolismo e meccanismi di riparo. Se non ci sono meccanismi di riparo si attiva il meccanismo della morte cellulare per evitare che mutazioni, alterazioni possano essere propagate nella progenie.

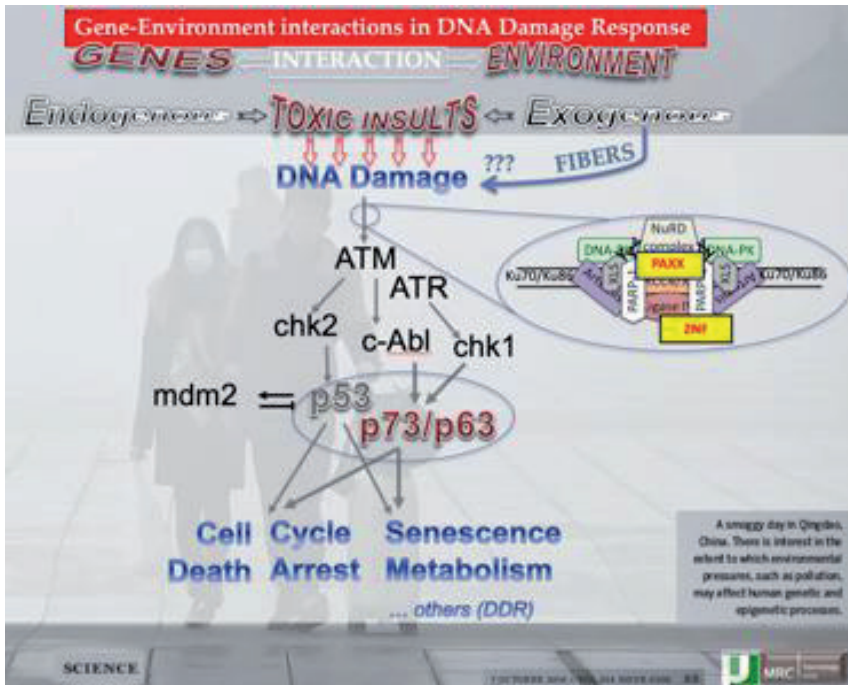
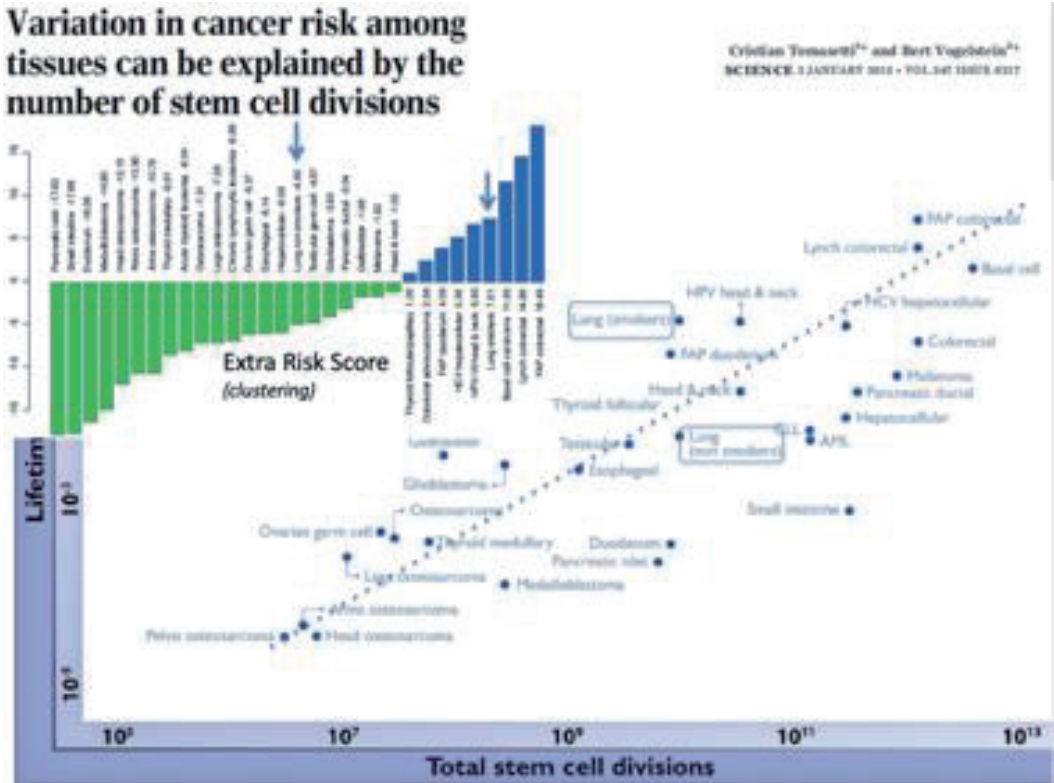


Figura 2. Principali meccanismi tossici endogeni ed ambientali.

A livello molecolare questo è abbastanza conosciuto poichè fino a pochi anni fa si pensava di sapere già tutto del meccanismo di riparo del DNA; però due anni fa è stata trovata una nuova molecola, PAXX, che è necessaria per il meccanismo di riparo ed adesso stanno per essere pubblicati dei fattori di trascrizione come ZNF281 che è in grado di regolare i meccanismi di riparo del DNA; tutto poi passa nella proteina p53 o geni simili (p63 e p73) che quindi vanno ad attivare processi di riparo. Le fibre, nanofibre, nanotubi che stanno per essere introdotti in maniera massiccia nell'industria dell'ambiente e di cui sappiamo abbastanza poco e sono simili come meccanismo di danno all'asbesto; le fibre di asbesto provocano l'attivazione di questi meccanismi che non è un danno al DNA "classico", hanno quindi meccanismi molecolari diversi, tant'è che è vero che i tumori da mesotelio-

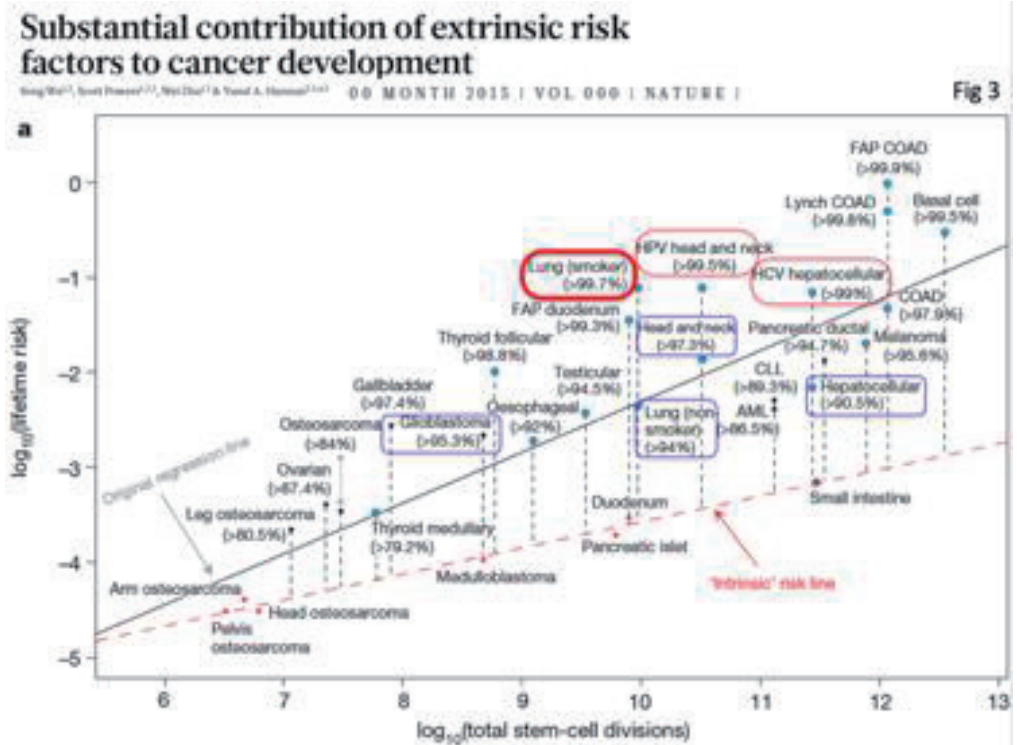
ma da asbesto derivano da un danno che dura 20/30 anni ma che alla fine vanno a sviluppare un risultato finale simile. Queste fibre, in cui un nostro collega, il professor Michele Carbone di Honolulu, ha identificato meccanismi genetici interni (BAP1) che se mutati favoriscono l'insorgenza del mesotelioma se non mutati ne inibiscono l'insorgenza, dimostrando la concomitanza di fattori genetici (BAP1) ed ambientali (asbesto).



**Figura 3.** Variazioni del rischio di cancro in rapporto alle divisioni delle cellule staminali di ogni tessuto. Dimostra una relazione lineare, indipendente dall'ambiente (teoria della "sfortuna"). Solo in alcuni tumori selezionati (insetto verde e blu) è possibile effettuare la prevenzione. Da: "Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention." Tomasetti C, Li L, Vogelstein B. *Science*. 2017; 355(6331): 1330-34. doi: 10.1126/science.aaf9011.

E' importantissimo oggi capire i meccanismi di base con cui l'ambiente in senso lato interagisce con i geni a livello genetico ed epigenetico per difenderci da questi danni. Vi mostro una diapositiva che illustra la teoria della "sfortuna" che mostra l'incidenza di rischio di contrarre un tumore nella nostra vita è perfettamente lineare con il numero di cellule staminali di quel tessuto, questo vuol dire che non c'è danno ambientale, se non ad esempio nel caso del polmone, di passare dal sotto della media ad un punto decisamente più elevato della media oppure di Papilloma

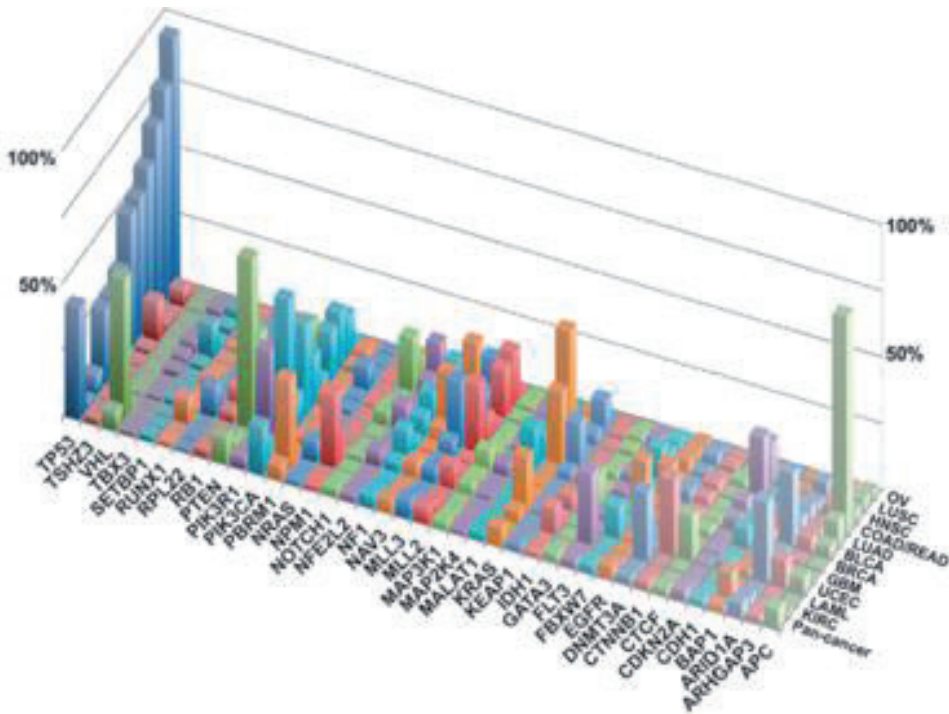
virus che possono elevare i tumori da testa-collo da basso ad alto rischio. Quindi, c'è un contributo dell'ambiente, ma primariamente è la nostra risposta genetica che porta allo sviluppo neoplastico. Questo lavoro ha portato ad identificare tutta una serie di tumori in cui i valori di rischio aggiuntivo è basso, per cui la prevenzione di questi tumori non è possibile farla mentre per altri tumori come per esempio quello del polmone da fumo, da fattori ambientali può avere un contributo significativo della prevenzione e questo va a selezionare l'intervento di tipo preventivo che noi dobbiamo fare.



**Figura 4.** Rispetto la linea rossa di rischio intrinseco delle cellule staminali, la linea grigia mostra la correlazione reale da rischio ambientale. Da: “ Substantial contribution of extrinsic risk factors to cancer development.” Wu S, Powers S, Zhu W, Hannun YA. Nature. 2016; 529(7584): 43-7. doi: 10.1038/nature16166.

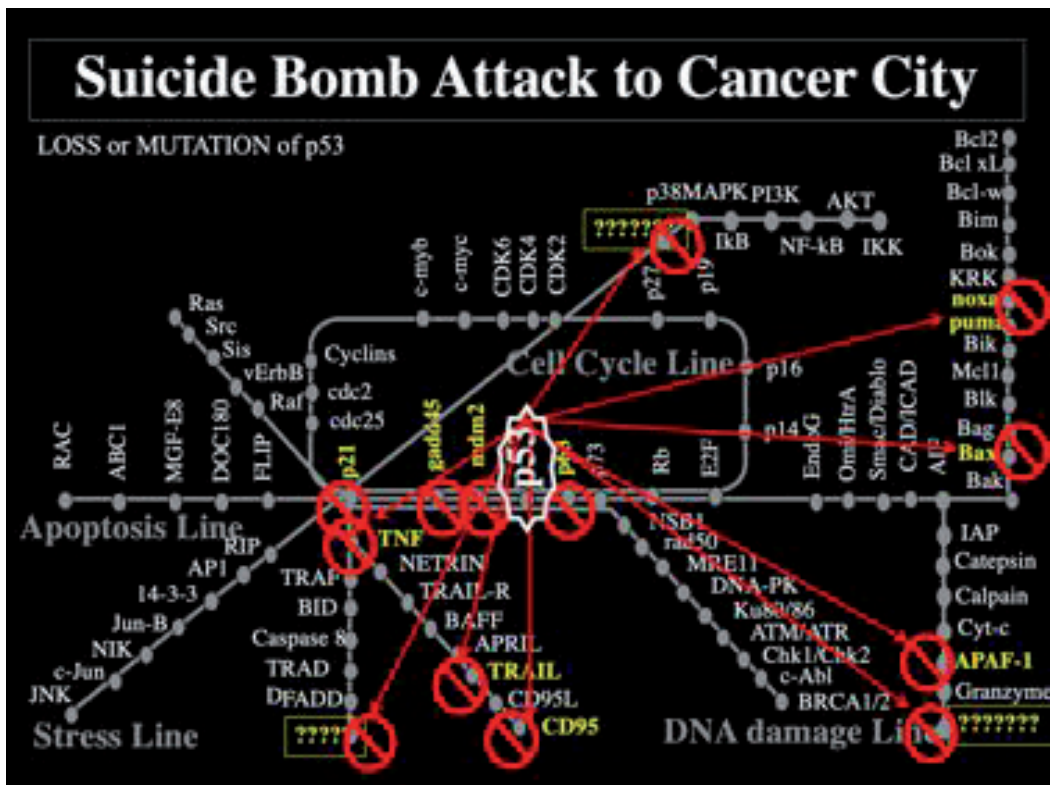
E' uscito un altro lavoro, in cui si afferma che la linea intrinseca del pericolo di rischio è collegata alle cellule staminali, però per alcuni tumori esiste un'altra linea di regressione in cui c'è aumento del rischio. In questo studio in realtà si sostiene che la teoria di base delle cellule staminali è molto meno importante dei fattori ambientali. Per esempio, nel del tumore al polmone dei fumatori il 99,7% è il contributo della parte ambientale, quindi un contributo elevatissimo; nell'HPV in questo caso un inquinamento da virus, il 99,5%. Anche qui però bisogna essere un po' cauti valutando la biologia dei rischi, per esempio: il tumore nel polmone dei

non fumatori, il rischio ambientale è del 94%. Fumare mi fa diventare il rischio dal 94% di fattori esogeni al 99%, solo il 5%? Allora posso fumare. Ma la domanda è: questo 94% di fattori ambientali da dove arriva? Quindi molti aspetti rimangono ancora da chiarire.



**Figura 5.** Frequenza di mutazioni in vari tipi di neoplasie umane. Il gene codificante per p53 (linea blu) raggiunge quasi il 100% nei tumori ovarici e comunque in tutti i tumori umani (pan-Cancer) raggiunge quasi il 50%, la frequenza più elevata di qualunque altro gene, dimostrandone l'assoluta importanza nella prevenzione dei tumori. Da: "TP53: an oncogene in disguise." Soussi T, Wiman KG. *Cell Death Differ.* 2015; 22(8): 1239-49. doi: 10.1038/cdd.2015.53.

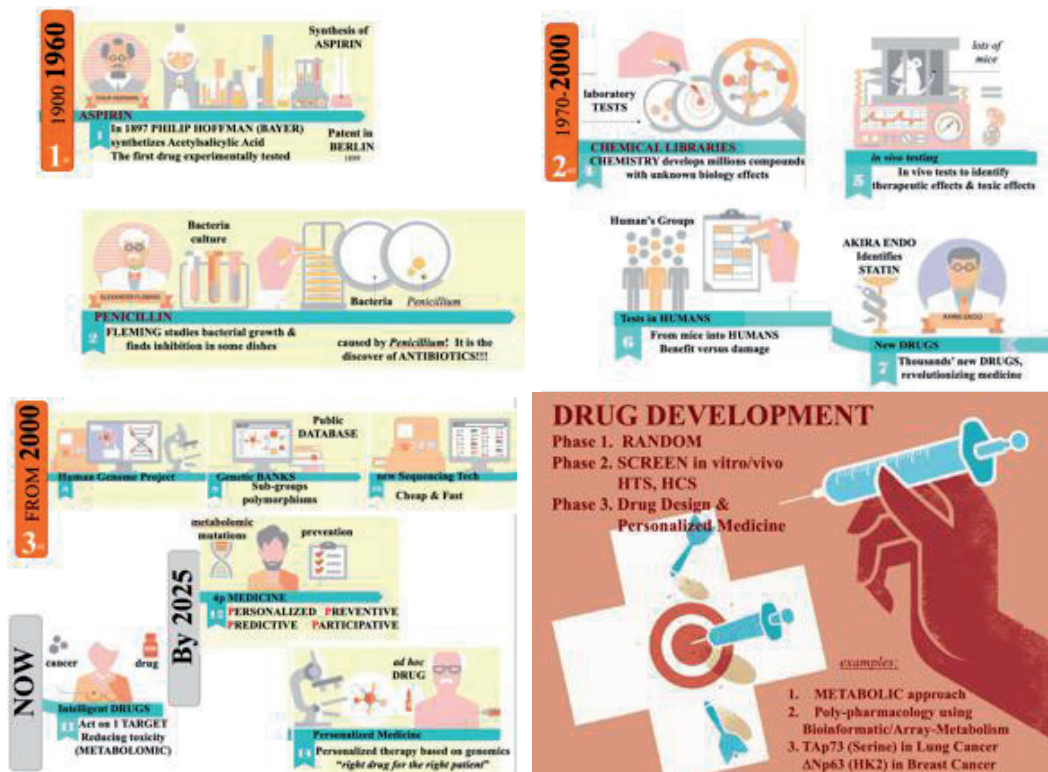
Lo studio tra il rapporto tra ambiente e geni è oggi non concluso ed è assolutamente importante portarlo avanti e cercare di capire quali sono i fattori predittivi che bisogna andare a monitorare. Uno di questi è p53: una proteina importante che risponde a stress, danno al DNA, stress ossidativo, fattori nutritivi e fa attivare tutta una serie di meccanismi, compreso il riparo ma anche la morte cellulare, l'apoptosi vera e propria. E' così importante che come vedete, in vari tipi di tumore, andiamo a guardare tutti i geni, oncogeni, oncorepressori, uno è frequentemente mutato ed è proprio p53, il più mutato in almeno il 50% di tutti i tipi di tumore. Per cui capire in un individuo com'è la funzionalità di questo gene e quindi i fattori di rischio che questi hanno è fondamentale. Non è del tutto conosciuto il suo funzionamento, ve l'ho rappresentato in una maniera poco scientifica: immaginiamo che questa proteina p53 controlli il ciclo cellulare, controlla anche la risposta allo stress, il meccanismo di riparo del DNA e la morte cellulare. Sta quindi in un crocevia delle funzioni essenziali di tutta la cellula. Questa è la



**Figura 6.** Rappresentazione poco ortodossa (ricalcando la linea di metropolitana di ... Londra) dell'azione di p53, al centro della regolazione del Ciclo cellulare, Danno al DNA, Stress, Apoptosi. Quando è mutato, inattiva gran parte delle funzioni cellulari, facilitando la formazione e sviluppo dei tumori. Da: Molecular mechanisms of cell death: recommendations of the Nomenclature Committee on Cell Death 2018. Galluzzi L,.... Melino G, Kroemer G. Cell Death Differ. 2018; 25(3): 486-541. doi: 10.1038/s41418-017-0012-4.

mappa della metropolitana di Londra (!?!): se p53 non funziona, tutta la rete funziona tranne una stazione; ma in realtà una deformazione non solo non funziona ma va a fare delle cose che normalmente non dovrebbe fare, acquisendo nuove funzioni. Bisogna capire quali "stazioni" vengono alterate e quali sono i meccanismi. Si trasforma quindi una stazione non funzionante in una "suicide bomb attack to cancer city". La ricerca di base deve riuscire ancora a capire come determinate proteine regolino meccanismi di interazione, di fosforilazione, di degradazione, di apoptosi. Abbiamo anche fatto uno screening per cercare farmaci che possano intervenire terapeuticamente: questa molecola che abbiamo trovato per esempio riesce a bloccare il meccanismo di riparo difettoso di queste cellule.

E' importante sia per lo sviluppo dei farmaci sia per lo screening della popolazione: lo sviluppo di farmaci ha subito tre fasi nell'evoluzione storica, una random, la seconda facendo degli screening e la terza, che sarà il futuro, è la medicina personalizzata sia per individuare persone a rischio e fare determinate mansioni sia fare un aggiustamento delle terapie.



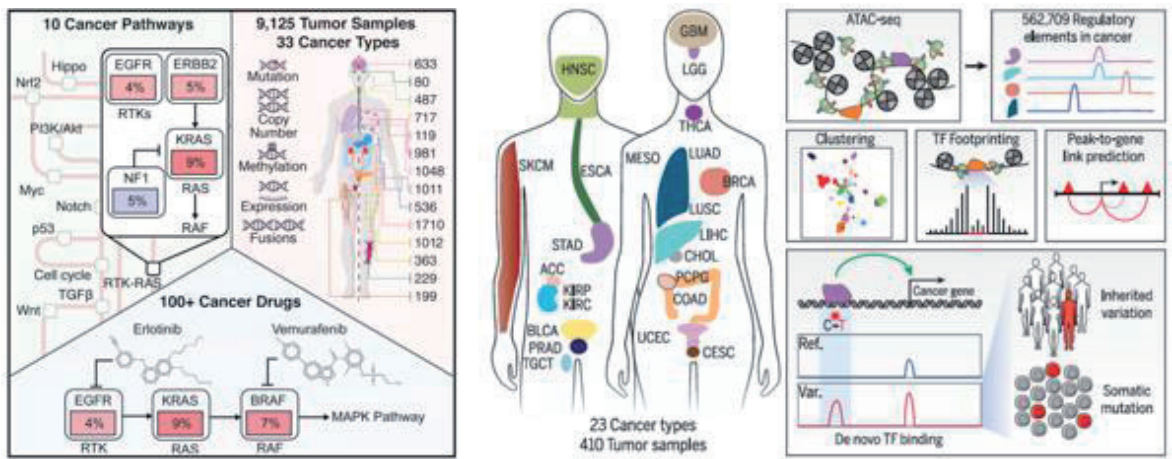
**Figura 7.** Rappresentazione semplificata delle varie fasi di sviluppo di farmaci. (1) inizio secolo 1900-1960: scoperte casuali. (2) 1970-2000: Selezioni chimiche in vivo ed in vitro. (3) dopo il 2000: uso di banche dati personalizzate e sequenziamento genomico e proteomico per selezionare l'azione di farmaci. (4) dopo il 2025: Medicina Personalizzata: .... Personalizzata, Preventiva, Predittiva, Partecipativa.

Esempi della prima fase: l'aspirina, la patente fu fatta a Berlino nel 1899, Hoffmann casualmente mette su una compagnia "la Bayer" e per caso scopre questo farmaco; così come successe per gli antibiotici come Fleming con la penicillina che scoprì anche lui per caso.

La seconda fase consiste nel fare analisi e selezioni chimiche in vivo e in vitro e di fare quindi sperimentazione nell'uomo. Così furono identificate le statine.

La terza fase, in cui siamo adesso, dopo il sequenziamento del genoma, in cui il futuro sarà quello di trovare farmaci intelligenti che possano funzionare soltanto su quel malato, soltanto in quella condizione, e quindi fare una medicina predittiva, personalizzata, preventiva e partecipativa in cui si può disegnare sul singolo malato il tipo di intervento che noi possiamo fare.

Un esempio è stato fatto su poche decine di migliaia di tumori, analizzando ed identificando le mutazioni, metilazioni e quindi aver potuto disegnare sul malato individuale i farmaci a seconda della mutazione e a seconda dello stato di avanzamento della malattia. A Tor Vergata stiamo effettuando una rete di sequenziamento a tappeto in cui sequenziamo tutto l'intero genoma di tessuti tumorali e cercare le mutazioni individuali o l'epigenoma, identificando geni che sono dei regolatori dell'espressione genica e quindi identificare esattamente quali sono le variazioni epigenetiche che possono essere predittive di una risposta clinica.



**Figura 8.** Pannelli in alto. L'analisi integrata di alterazioni genetiche in numerosi vie metaboliche biochimiche effettuate su quasi 10 mila tumori di diversa origine mostrano classiche alterazioni individuali e suggeriscono nuove opportunità per terapie individuali ed in combinazioni personalizzate. Pannelli in basso. L'analisi del profilo di espressione della cromatina in oltre 400 campioni da 23 tipi diversi di tumore, mostra oltre 562 mila elementi regolatori del DNA la cui attività identifica numerosi fattori di trascrizione o altri elementi regolatori che influenzano le risposte alla terapia e quindi le curve di sopravvivenza. Questi aprono la ricerca su nuovi possibili obiettivi farmacologici su base personalizzata. Da: *Oncogenic Signaling Pathways in The Cancer Genome Atlas*. Sanchez-Vega F, et al. *Cell*. 2018; 173(2): 321-337.e10. doi: 10.1016/j.cell.2018.03.035. Da: *The chromatin accessibility landscape of primary human cancers*. Corces MR, et al. *Science*. 2018; 362(6413). pii: eaav1898. doi: 10.1126/science.aav1898.

In conclusione: la medicina personalizzata è il futuro, ci permetterà di individuare i lavoratori a rischio, gli ambienti e cercare di interagire con l'aspetto professionale del lavoro.





## TAVOLA ROTONDA

### **L'innovazione scientifica nella riduzione del rischio ambientale e nella valutazione dello stato di salute.**

**Corrado Clini**

*Visiting professor presso la "School of Environment", Tsinghua University di Pechino*

La prima visione che mi viene in mente riguarda la complessità della valutazione del rapporto tra l'ambiente e la salute. Una complessità che richiede capacità di analisi più raffinata di quelle che abbiamo sviluppato fino ad ora, soprattutto in presenza di una variegata e spesso contraddittoria casistica di contenziosi in Italia che fanno riferimento ai possibili rapporti tra cause ambientali e danni alla salute. Io credo che questo tema debba essere approfondito partendo da alcuni dati ai quali non si può sfuggire, parliamo per esempio della gestione del riciclo dei rifiuti, di inceneritori o termovalorizzatori. E' un po' difficile contestualizzare il rapporto tra rischi ambientali e danni alla salute quando abbiamo situazioni nel nostro paese e a livello europeo che dimostrano sulla base di serie storiche consolidate che le problematiche relative ai rischi sono state in gran parte superate laddove i rischi venivano gestiti. Allora, è molto interessante e importante approfondire le tematiche relative, per esempio, al rapporto tra la variabilità individuale e i fattori di rischio, quindi le diverse risposte in relazione alle diverse caratteristiche individuali o di popolazione. Ma c'è un dato di fatto, abbiamo regioni in Italia, e in Europa, dove da trenta, quarant'anni vengono esercitate attività che, dal punto di vista dello stato di salute di una popolazione, soprattutto negli ultimi quindici anni, non hanno messo in evidenza significativi impatti. Chi conosce Copenaghen sa che l'impianto per la cogenerazione attraverso l'utilizzazione dei rifiuti è situato nel centro della città. Bisognerebbe riuscire a rendere pubblico questo dato e difenderlo tenendo conto che le fake news relative a questo settore sono molte. Bisogna continuare a lavorare per capire l'alimentazione e le problematiche relative all'inquinamento del traffico. Capire meglio, se in un qualche modo si può standardizzare, se è possibile il rapporto tra gruppi di popolazioni con le loro caratteristiche e le sorgenti di rischio proprio perché ci troviamo di fronte a popolazioni esposte allo stesso rischio, all'interno delle quali però ci sono risposte molto diverse da un punto di vista individuale. La seconda considerazione riguarda la possibilità di fare un salto, sostanzialmente andare oltre, dal punto di vista delle conoscenze, delle tecnologie e delle gestioni, rispetto a quello che oggi abbiamo come dato consolidato, ad esempio, nel settore dei rifiuti. Andare oltre è un obiettivo giusto, "rifiuti zero" è un obiettivo apprezzabile, ma questo obiettivo non può essere la scusa per tenere la situazione in un'altissima pericolosità e con altissimi costi, dal punto di vista della salute e dal punto di vista economico. Per esempio, la Sicilia, si trova in una situazione allucinante: la gestio-

ne dei rifiuti in Sicilia è ai livelli degli anni '50/'60, ci sono almeno 3 impianti per la produzione di cemento, con una norma che permette di usare un combustibile solido secondario nei cementifici. Ma in Sicilia non si riesce ad avere la possibilità di usarlo perché sarebbe pericoloso. L'alternativa che viene usata è il petcoke, che fra tutti i derivati dei combustibili fossili è il peggiore, dal punto di vista ambientale, con emissioni incredibili. L'opinione pubblica, supportata da "esperti", viene orientata a negare l'impiego di questo tipo di combustibile alternativo, perché convinti che provocherebbe danni importanti, tumori molto diffusi ecc. Pensate che in Germania il 65% dei cementifici è alimentato con combustibile solido secondario di qualità anche inferiori di quella normata in Italia, in Olanda l'85%, in Austria siamo attorno al 70%. Certamente abbiamo l'importante tema di "fare di più", però abbiamo in Italia un problema serissimo, fare un'operazione "verità" e farla avendo il coraggio di aprire un confronto con una capacità molto articolata e forte di creare false notizie e di farle diventare verità scientifiche.

## **Filippo Brandolini**

*Vice Presidente Vicario UTILITALIA*

Se parliamo di impianti per la gestione rifiuti che sono importanti per la collettività, dobbiamo utilizzare al meglio le conoscenze medico-scientifiche o tecnologiche. Se utopicamente la paura contrasta con la conoscenza, la realtà dice che le nostre scelte sono guidate dalla paura che nega la conoscenza, anche quando essa è scontata ed è evidente. Talvolta si nega l'evidenza dei fatti adombrando qualche dubbio sull'interlocutore, ci possono essere slogan del tipo "rifiuti zero" che sicuramente può essere una prospettiva interessante, ma che può anche essere declinata. Provo a fare qualche esempio concreto: nell'ambito della conoscenza, si può fruire del lavoro che ISPRA quotidianamente fa nel monitorare l'ambiente, nel conteggiare tutti i prodotti, che ci dà una conoscenza molto approfondita degli aspetti ambientali nel nostro paese. Uno di questi elementi di conoscenza sono le fonti di emissioni nell'atmosfera. ISPRA censisce quest'ultime in maniera molto analitica e le attribuisce alle fonti emmissive. Emerge che il contributo della gestione dei rifiuti delle fonti emmissive è vicino allo 0%, al di sotto dell'1%; l'altro elemento è che ovviamente le altre maggiori fonti di emissioni sono il traffico prodotto dal sistema di trasporti e il riscaldamento delle abitazioni. Andando a leggere questi dati si vede che la gestione dei rifiuti dal 1990 al 2018 è paurosamente e positivamente migliorata in termini di riduzioni delle emissioni, avvicinandosi allo zero, perché è intervenuto un cambiamento tecnologico e normativo. La costruzione di impianti di incenerimento riprese dopo la vicenda Seveso negli anni '90, a seguito di una delibera di un comitato interministeriale dell'84 applicativa della prima normativa nazionale sui rifiuti, il DPR 915/82, con la quale è stata prevista per gli impianti di incenerimento una serie di misure, come limiti di emissione più rigidi per quanto

riguarda i vari parametri di inquinamento, oltre ai livelli di temperatura nella camera di combustione molto più elevati con dei minimi importanti. Questo provvedimento, una volta applicato negli impianti che sono stati poi realizzati anche a seguito dell'innovazione tecnologica, ha consentito di ridurre le emissioni, portando a livelli marginali l'impatto ambientale di questi impianti. Sono poi seguite altre misure indotte dalle normative europee che hanno consentito ulteriormente di aumentare le performance ambientali di questi impianti e tra l'altro nuove misure europee usciranno nel corso di quest'anno. Questo è un esempio concreto di come l'innovazione tecnologica, accompagnata da provvedimenti normativi, possa intervenire nel miglioramento della gestione ambientale dei rifiuti di questi impianti che allo stato attuale sono indispensabili. "Rifiuti zero" è una prospettiva sicuramente interessante da perseguire, ma che attualmente non è fattibile e sostenibile. Un secondo esempio molto concreto: il 30-35% dei rifiuti urbani, quelli prodotti nelle nostre abitazioni, è rifiuto organico, un rifiuto che se non raccolto in maniera differenziata e, soprattutto, se smaltito in discariche è quello che produce maggiori impatti ambientali anche in termini di gas alteranti nelle discariche, produce il biogas che è un gas climalterante che, se non correttamente gestito, può produrre inquinamento della fauna. La raccolta differenziata quindi consente una gestione più efficace e sicura dal punto di vista ambientale. La raccolta può essere destinata ad impianti che trattano questi rifiuti per recuperare materia ed energia. Nel passato si è parlato di questo tema, ma c'è voluto del tempo per acquistare una consapevolezza. Adesso con il rifiuto organico si fa molto di più: si può estrarre da questo rifiuto del biogas, invece di farlo disperdere nell'ambiente lo si tratta recuperandolo in questi impianti di sintesi anaerobica e si può produrre biometano che sostituisce il metano fossile. Siamo in una situazione paradossale in questo paese: da un lato si spinge sul biometano, sulla circolarità perfetta di quel 30% di rifiuti delle nostre case, raccogliendoli producendo biometano che può essere utilizzato per il trasporto, ma mancano le norme che consentirebbero di realizzare questi impianti. Il rischio grosso è che l'economia circolare è qualcosa che tutti vediamo come una prospettiva miracolistica, ma non c'è nulla di miracolistico, ci contraddiciamo creando ostacoli "non tecnologici" che ci impediscono di perseguire questa prospettiva. Più in generale, l'innovazione sta nella prospettiva dell'economia circolare, nel cambio di paradigma, ma l'avvertenza è il non pensare che siamo noi gli attori dell'economia circolare, bisogna fare in modo che coloro che mettono qualsiasi tipo di merci, prodotti nel mercato facciano in modo che abbiano un tempo di vita più lungo possibile invertendo la tendenza del sistema consumistico; quando il prodotto non viene più utilizzato che sia facilmente riciclabile ovvero che ci sia un sistema di impianti che ne possano trattare in maniera efficace lo smaltimento, possibilmente non in discariche ma in impianti di recupero energetico per estrarre energia. Questa è una prospettiva che impone una grande assunzione di responsabilità da parte delle imprese: ci sono tanti esempi ma è evidente che servono impianti normativi che prevedano incentivi e penalizzazione molto efficaci sotto

questo profilo. E' interessante quello che ha deciso l'Unione Europea sulle plastiche monouso: dal 2025 le bottigliette di plastica per bevande devono contenere almeno il 25% di plastica proveniente da riciclo (nel 2035 il 35%). Il riciclo esce quindi da una logica volontaristica ed entra in una logica di mercato di tutela, di quella che è una commodity che si deve confrontare con le plastiche di origine fossili.

La tutela della salute e dell'ambiente passa attraverso anche il tema della legalità: la cosa che dispiace è, per esempio, che il famigerato inceneritore venga associato alle cause che producono inquinamento ambientale e danni alla salute dei cittadini, che invece deve essere ricondotta a quella che è la "terra dei fuochi", smaltimenti illegali della criminalità organizzata che illegalmente ha operato in tanti anni in quei luoghi. E' difficile contrastare questa mancanza di conoscenza: la Campania tuttora porta i suoi rifiuti in Olanda e in Portogallo; la città di Roma strutturalmente deve avvalersi di impianti che sono al di fuori del proprio territorio, al di fuori della regione. Non solo per i rifiuti da smaltire, ma anche per quelli della raccolta differenziata ed è ben noto che le realtà di Roma e Napoli conferiscono i propri rifiuti organici nel nord-est. Napoli spende 33 milioni di euro per conferire rifiuto organico e di raccolta differenziata in una provincia di Padova dove si produrranno compost e metano, ma per fare quei 600 chilometri si vanificano i benefici ambientali della raccolta differenziata. Per migliorare le condizioni di salute dobbiamo vincere le paure che costituiscono un freno, è evidente che un inceneritore non piace a nessuno, è un impianto che dal punto di vista umano produce un impatto ma è un impianto che consente di ridurre il danno rispetto alla non gestione dei rifiuti. Il problema dell'Italia è che si considera l'obiettivo "zero rifiuti" come qualcosa che è già successo, quando invece è ancora lontano, andando contro l'installazione di impianti.

### **Francesco Di Maria**

*Docente Macchine a Fluido, Dipartimento di Ingegneria, Università degli studi di Perugia*

Volevo fare tre considerazioni. Ho sentito parlare della conoscenza, che può aiutare a gestire il rischio ed avere un atteggiamento più consapevole. Ora, una cosa importante per la conoscenza è la comunicazione che richiede un grande supporto della ricerca, imparziale, chiara e un'autorevolezza, esprimendo quindi certi concetti e certi passaggi con processi che siano trasparenti da parte di associazioni o da enti accusati di imparzialità. Per quanto riguarda la mia esperienza personale, ho spesso portato a visitare impianti di trattamento dei rifiuti, di termovalorizzatori, di inceneritori ad esponenti politici o studenti e il loro atteggiamento era molto prevenuto. Far conoscere alla popolazione quello che succede in questi impianti è sicuramente un'operazione di trasparenza che aiuta certi percorsi, ciò per quanto riguarda la comunicazione. Poi, ho un dato di evidenza legato a ciò che è la normativa, quelli che sono i percorsi che portano alla costruzione di impianti o a certe

scelte tecnologiche o gestionali. La normativa è un aspetto importante che diventa riferimento per gli enti pubblici, i progettisti e tutti gli addetti ai lavori ai fini delle scelte o autorizzazione di certi impianti. Oggi la normativa ha uno strumento molto importante che certifica la rivoluzione tecnologica nel settore industriale e dei rifiuti, che è la direttiva emissioni industriali che, in tutti i settori della tecnologia, ciclicamente fa un percorso di revisione, di certificazione di quelle che sono le migliori tecniche disponibili, cioè fornisce le indicazioni, per i vari settori dell'industria, per il raggiungimento del migliore stato di efficienza in termini di riduzione delle emissioni.

## Principali riferimenti normativi

- **Parte IV del T.U.A. (D.Lgs. 152/06): Gestione integrata dei rifiuti, Obiettivi Generali e Specifici;**
- **Parte II del T.U.A. (D.Lgs. 152/06): Valutazione di Impatto Ambientale, effetti diretti ed indiretti di un progetto o opera sulla salute e sull'ambiente;**
- **Direttiva Emissioni Industriali (in via di approvazione): Indica le migliori tecniche disponibili (BAT) ed i livelli di emissione associati (AEL).**



La commissione europea fa questo percorso in maniera trasparente, avvalendosi di un pubblico molto ampio di esperti che provengono da diverse nazioni, da diversi settori dell'industria. Oggi, dal punto di vista delle tecnologie c'è stato un progresso molto forte: considerando la termovalorizzazione, grazie alla rivoluzione tecnologica è stato abbassato il limite di emissioni di questi impianti fino anche all'80%; ma se andiamo sul campo dei diversi impianti che operano a controllare le sostanze pericolose, ad esempio le diossine, si riscontra che sono anche 10 volte inferiori ai limiti della norma. Per cui certi problemi che sembrano più grandi, sono oggi marginali rispetto a quello che è il funzionamento di emissioni di questi impianti.

Oggi lo strumento più importante che esiste, di cui gli enti pubblici e i progettisti si avvalgono per modernizzare nuove tecnologie e gli impianti, è la valutazione di impatto ambientale che fa riferimento alla direttiva emissioni. Tra l'altro l'ultima direttiva comunitaria del 2014 ha introdotto obbligatoriamente delle linee guida in via di approvazione, che prevedono la valutazione di impatto sanitario contestuale alla valutazione di impatto ambientale per cui diventa uno strumento integrato.

Nel produrre ed organizzare informazione da studi sui fattori di rischio, di percorsi espositivi e sulle ricadute sanitarie di certe tecnologie e certi dati c'è ancora bisogno di supporto e trasparenza normativa e bisogna fare un grande lavoro per comunicare quello che è lo stato dell'arte. Terza e ultima considerazione che faccio: vogliamo applicare quest'idea di "rifiuti zero" che sicuramente nel futuro può essere fattibile, ma oggi esiste una lacuna normativa e di approccio di sistema, ossia mancano i supporti normativi che stabiliscono quando il materiale che io raccolgo come rifiuto, per poi trattarlo, cessa di essere rifiuto e quindi può rientrare nel circuito dei beni di consumo delle materie prime, passaggio fondamentale che però manca (come nella plastica e nel digestato) bloccando il sistema.

Altra cosa che deve far riflettere, è l'approccio rigido che ci impone l'Europa sulla gestione dei rifiuti: la gerarchia del riciclo e gli obiettivi che ci impongono, come ad esempio riciclare il 50% dei rifiuti entro il 2020, che arriverà al 65% entro il 2035 ecc...

## La gestione dei rifiuti: obiettivi generali, specifici e finalità (Parte IV, D.Lgs. 152/06 e ss.mm.ii.)

### GENERALI

Art. 179 (Criteri di priorità nella gestione di rifiuti)  
Implementazione della gerarchia (Materia)



### SPECIFICI

ART. 181 (riciclaggio e recupero dei rifiuti)  
Preparazione per il riuso e/o riciclo di almeno il:

- 50% entro il 2020;
- 55% entro il 2025;
- 60% entro il 2030;
- 65% entro il 2035.



ART. 178 (Finalità), Comma 2:  
I rifiuti devono essere recuperati o smaltiti **senza pericolo per la salute dell'uomo e senza usare procedimenti o metodi che potrebbero recare pregiudizio all'ambiente** e, in particolare: a) senza determinare rischi per l'acqua, l'aria, il suolo, nonché per la fauna e la flora; b) senza causare inconvenienti da rumori o odori; c) senza danneggiare il paesaggio e i siti di particolare interesse, tutelati in base alla normativa vigente.

E' importante che alla gerarchia e agli obiettivi imposti venga associata, allo stesso livello, il concetto della salute umana, quindi come obiettivo da perseguire non come una conseguenza di una gestione.

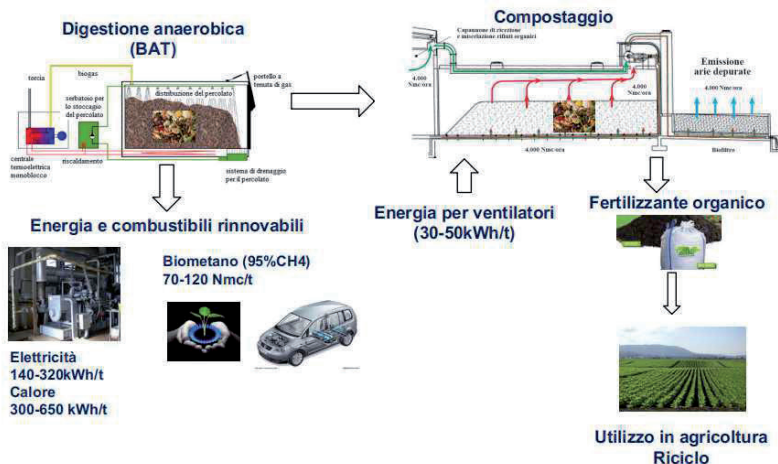
## La gestione dei rifiuti: principali metodi e tecnologie per il riuso ed il riciclo



La tecnologia dà una mano ed evolve, ma spesso sono i sistemi di gestione che possono ridurre le conseguenze sanitarie.

Oggi riciclare significa fare fertilizzante organico, però esistono anche altre opzioni potendo trattare questo materiale nei termovalorizzatori utilizzando il suo carbonio per produrre energia o come combustibile solido secondario.

## I trattamenti biologici per il riciclo della frazione organica: emissioni dirette, indirette ed evitate



## **Marzio Lasagni**

*Direttore Generale AISA Impianti S.p.A.- Arezzo*

L'idea nasce da lontano, in campo epidemiologico. Il professor Claudio Clini ci segue da molto tempo e in una discussione sui vaccini in cui feci l'osservazione che per un bambino di un anno sei vaccini erano troppi, la risposta del professor Clini mi spiazzò. Mi chiese su quali basi scientifiche io stessi facendo quella osservazione. Non ne avevo, era una mia sensazione, pensavo che sei vaccini tutti in una volta facessero male. Mi resi conto che anche io, che sono ingegnere, mi ero fatto trascinare da una sensazione generale per cui nel tam-tam mediatico e dei social, si dice che alcune cose fanno male; ma non avevo una base, non avendo mai studiato il problema, mi ero autoconvinto senza un ragionamento scientifico. Quindi è nata l'idea di fare qualcosa in più affidandoci al prof. Clini che fa questo di mestiere, studiando e cercando di capire se effettivamente può esserci una correlazione tra l'incenerimento e la salute umana, utilizzando le più recenti scoperte. Ed è venuta fuori l'idea di questo progetto, quello di fare un'indagine genomica con la collaborazione volontaria di esposti e non esposti. I risultati della ricerca saranno dati al comune residente (Arezzo), questo perché i risultati non possono appartenere ad un solo soggetto e perché il sindaco è il responsabile della salute pubblica della collettività. Quindi, non saremo noi i primi ad avere i dati. Quello che noi facciamo lo facciamo perché siamo estremamente convinti della necessità di trasparenza, dimostrando che la termovalorizzazione dei rifiuti non è un male e questo messaggio può passare solo grazie all'appoggio della scienza e della medicina. Abbiamo investito in campo comunicativo perché è l'unico modo per poter dimostrare l'assenza di rischio, mostrando alle persone che vengono che l'impianto non è un "bidone" ma uno strumento di tecnologia che conta più di 40 persone di cui il 75% è diplomato o laureato in campi inerenti, il livello scientifico è molto alto, perché per non creare danno all'ambiente deve esserci un progetto ad alta tecnologia.





Finito di stampare nel mese di ottobre 2019  
per i tipi della **Aldo Francisci Editore** - Abano Terme (Pd)  
Supplemento al n. 143 del periodico informAbano  
Iscritto al Registro Stampa del Tribunale di Padova al n. 733 del 1/6/1982  
*Copia omaggio*

