

S.S. Epidemiologia Molecolare

Tecniche avanzate di analisi statistica e bioinformatica applicate a studi su biomarcatori in popolazioni umane

Linea di ricerca: 1 – Prevenzione e Cancerogenesi

Programma: b - Biomarcatori biologici e molecolari di esposizione, di danno, di suscettibilità e di rischio di cancro

Responsabile scientifico: Marcello Ceppi

Altro personale della struttura partecipante al progetto: Cecilia Lando, Fabio Gallo

Anno di inizio: 2009

Durata: 36 mesi

Parole chiave: analisi statistica; bioinformatica; epidemiologia molecolare; genomica; high throughput

Altre strutture IST partecipanti: S.C. Epidemiologia, Biostatistica e Clinical Trials (D. Ugolini); S.C. Epidemiologia Clinica (P. Bruzzi); S.S. Oncologia Traslazionale Pediatrica (G.P. Tonini)

Altri Enti coinvolti: University of Maastricht, Holland (J. Kleinjans); CSIRO, Adelaide, Australia (M. Fenech); Università di Pisa (R. Barale); Laboratorio di Biologia Molecolare, Istituto G. Gaslini, Genova (L. Varesio); Dipartimento di Matematica, Università di Genova (F. Patrone), IRCCS S. Raffaele Pisana, Roma (S. Bonassi)

Tipologia progetto: clinico-epidemiologica osservazionale

Area di interesse: prevenzione primaria/secondaria

Soggetti cofinanziatori: Commissione Europea; Ministero della Salute; Associazione Italiana per la Ricerca sul Cancro

Background

L'epidemiologia molecolare è nata dalla necessità di integrare i recenti grandi sviluppi della biologia molecolare con disegni di studio più rigorosi e metodiche di indagine più complesse tipici dell'epidemiologia tradizionale. Tra i campi che maggiormente hanno beneficiato di questo interscambio vi è l'analisi statistica dei dati biologici individuali (biomarcatori). Nei primi lavori l'analisi dati era insufficiente ed imprecisa mentre attualmente si è passati alla definizione di protocolli di analisi da applicarsi specificamente a ciascun biomarcatore. In tempi più recenti l'evoluzione di nuove tecniche di biologia molecolare - definite di high throughput - ha creato la necessità di sviluppare nuovi approcci informatici e statistici adatti all'elaborazione di grandi matrici di dati. In particolare, la tecnologia dei microarray permette di analizzare simultaneamente l'espressione di decine di migliaia di geni con un singolo esperimento fotografando il "livello di attività" dei geni in un individuo. Esaminando il profilo di espressione dei geni di un individuo affetto da una malattia o esposto ad un agente genotossico e comparando l'espressione genica in individui malati rispetto ai sani o esposti rispetto ai controlli, è possibile individuare nuovi geni la cui espressione è associata a determinati stati patologici o di esposizione e predire l'appartenenza di un individuo, per esempio, ad una specifica classe di malattie genetiche, unicamente in base al suo profilo di espressione.

Obiettivo generale del progetto ed eventuali obiettivi secondari

Nei tre anni oggetto del progetto lo scopo principale è quello di applicare metodiche di analisi statistica allo specifico settore dei biomarcatori utilizzando le tecniche che meglio si adattano alla particolare natura degli end-points osservati. Scopo finale è la definizione di standard metodologici per la conduzione di analisi statistiche in studi di popolazione su biomarcatori. A tal fine verranno utilizzati dati prodotti all'interno dell'Istituto o nell'ambito di collaborazioni nazionali ed internazionali. L'altro obiettivo - che riguarda la genomica - è quello di identificare, mediante tecniche statistiche applicate ai dati di espressione genica, i meccanismi di interazione tra geni che regolano i complessi processi di attivazione alla base dell'insorgenza di malattie. Questo obiettivo è perseguito, in particolare, mediante lo sviluppo di modelli matematici in grado di incorporare al loro interno il livello di interazione associato a tutte le potenziali "coalizioni" di geni, ed il successivo sviluppo di indici di rilevanza volti alla selezione dei geni più influenti. Tali modelli, basati sulla disciplina matematica nota come Teoria dei Giochi, sono da noi applicati allo studio di dati di microarray provenienti da soggetti principalmente in età pediatrica, come bambini affetti da neuroblastoma o bambini sani esposti ad inquinamento ambientale.

Impatto assistenziale certo o potenziale

Soprattutto nel campo oncologico la genomica ha aperto nuove prospettive per la classificazione molecolare delle patologie, la comprensione dell'eziologia dei tumori e l'individuazione di nuovi bersagli terapeutici.

Risultati e prodotti 2009

Uno degli obiettivi della nostra unità è stato quello di applicare all'analisi di dati di espressione genica modelli matematici allo scopo di definire indici di rilevanza in grado di misurare l'influenza di ciascun gene nei processi alla base dell'insorgenza di malattie genetiche. In questo contesto, durante l'ultimo anno di attività, abbiamo focalizzato il nostro lavoro di ricerca sullo sviluppo di un nuovo metodo basato sull'analisi di reti di co-espressione utilizzando un nuovo approccio basato sulla Teoria dei Giochi. Le reti di co-espressione sono basate su rapporti di correlazione tra coppie di geni. I nodi della rete sono i geni e le connessioni sono definite in base alla co-espressione di coppie di geni.

Consuntivo 2009 - Programmazione 2010

In generale, la correlazione di Pearson viene utilizzata come misura di co-espressione tra due geni. L'interpretazione dell'interazione genica nelle reti di co-espressione genera la necessità di un metodo in grado di definire l'importanza dei nodi della rete (ovvero i geni). Tuttavia, le misure di importanza riportate in letteratura non tengono conto della forza di interazione all'interno di sottogruppi di geni, di qui la necessità di sviluppare modelli matematici ad hoc. Tali modelli sono stati applicati a due studi distinti, entrambi relativi all'espressione genica in età pediatrica: il primo studio riguarda dati di espressione genica provenienti da bambini affetti da neuroblastoma mentre il secondo studio riguarda dati di espressione genica proveniente da bambini sani esposti ad inquinamento ambientale. È stato fatto un ampio studio di validazione dei risultati mediante simulazioni numeriche e tecniche statistiche di analisi supervisionata, che ha anche condotto alla predisposizione di software dedicato all'analisi dell'espressione genica (implementato in R), che potrà essere facilmente applicato nei successivi studi di dati di espressione inerenti al progetto. Sempre nell'ambito dei metodi di analisi basati sulla Teoria dei Giochi, il valore Banzhaf, è stato utilizzato come un indice di rilevanza in alternativa al valore Shapley, ed è stato caratterizzato assiomaticamente sulla classe dei giochi definiti su dati di microarray. Un'altra applicazione di questi modelli è stata utilizzata per generare una signature prognostica nel contesto del tumore alla mammella. I nostri metodi di analisi sono stati inoltre applicati all'analisi di dati di microarray di microRNAs provenienti da tumori quarto stadio ad alto rischio, al confronto di trattamenti farmacologici mediante analisi di dati di espressione provenienti da pazienti affetti da artrite reumatoide ed all'analisi di correlazione tra espressione genica e micronucleo.

Pubblicazioni

Betti M.-Neri M.-Ferrante D.-Landi S.-Biava A.-Gemignani F.-Bertolotti M.-Mirabelli D.-Padoan M.-Ugolini D.-Botta M.-Bonassi S.-Magnani C.-Dianzani I.

Pooled analysis of NAT2 genotypes as risk factors for asbestos-related malignant mesothelioma.
Int. J. Hyg. Environ. Health 212(3):322-329, 2009

Gemignani F.-Neri M.-Bottari F.-Barale R.-Canessa P.-Canzian F.- Ceppi M.-Spitaleri I.-Cipollini M.-Ivaldi G.-Mencoboni M.-Scaruffi P.-Tonini G.P.-Ugolini D.-Mutti L.-Bonassi St.-Landi S.

Risk Of Malignant Pleural Mesothelioma And Polymorphisms In Genes Involved In The Genome Stability And Xenobiotics Metabolism.

Mutat. Res. Fund. Mol. M. 671:76/83, 2009

Moretti S.

Game Theory applied to gene expression analysis, 4OR.

Quart. J. Operat. Res. 7:195-198, 2009

Scaruffi P.-Stigliani S.-Moretti S.-Coco S.-De Vecchi C.-Valdora F.-Garaventa A.-Bonassi St.-Tonini G.P.

Transcribed ultra conserved region expression is associated with outcome in high risk neuroblastoma.

BMC cancer 9:441;1/441;9, 2009

Cutolo M.-Villaggio B.-Secchi M.A.-Paolino S.-Moretti S.-Gallo F.-Bonassi S.-Sulli A.-Seriolo B.- Montagna P.- Soldano S.

Inflammatory gene profile in early rheumatoid arthritis and modulation by leflunomide and prednisone treatment.

Ann. N.Y. Acad. Sci., accepted

Lucchetti R.- Moretti S.- Patrone F.- Raddrizzani P.

The Shapley and Banzhaf indices in microarray games.

Comput. Operat. Res., in press

Moretti S.

Statistical analysis of the Shapley value for microarray games.

Comput. Operat. Res., in press

Presentazioni a congressi

Gallo F.- Moretti S.- Fenech M.- Bonassi S.

Micronuclei frequency and gene expression in children exposed to air pollution.

Abstract of the 10th International Conference on Environmental Mutagens. Florence, 20 - 25 August 2009

Mirisola V.- Moretti S.- Bonassi S.- Gennari A.- Sormani M.P.- Pfeffer U.

An Application of Microarray Games to Gene Expression Analysis: The Microarray Genes Signature.

Proceedings of the sixth international meeting on computational intelligence methods for bioinformatics and biostatistics, Genova, 15-17 October 2009

Moretti S.- Fragnelli V.- Patrone F.- S. Bonassi

Coalitional Games to Measure the Centrality of Genes on Interaction Networks.

Proceedings of the sixth international meeting on computational intelligence methods for bioinformatics and biostatistics, Genova, 15-17 October 2009

Attività previste e risultati attesi nel 2010

Nel 2010 l'attività principale nell'ambito del progetto sarà quella di applicare metodiche di analisi statistica allo specifico settore dei biomarcatori e della genomica valutando le tecniche che meglio si adattano alla particolare natura degli end-points osservati. Scopo finale è la definizione di standard metodologici per la conduzione di analisi statistiche in studi di epidemiologia molecolare.

Consuntivo 2009 - Programmazione 2010

Si procederà infatti all'applicazione delle tecniche più innovative sui dati di microarray e su dati ad essi assimilabili. In un primo studio verrà analizzato il pattern genetico di pazienti lungo-sopravvissuti di cancro dello stomaco metastatico alla diagnosi rispetto a diversi gruppi di controllo; scopo dello studio è quello di evidenziare se qualche gene o gruppi di geni sono differenzialmente espressi tra i casi rispetto ai controlli. Si parte da un set iniziale di 96 geni a cui verranno applicate le procedure di normalizzazione e di controllo di qualità al fine di selezionare i dati più affidabili. Si procederà quindi a confrontare le medie di espressione nei vari gruppi. In un secondo studio, in corso di definizione, verranno confrontati diversi profili antigenici in pazienti positivi al virus HIV rispetto a popolazioni di controllo. Scopo dello studio sarà quello di valutare il livello di risposta anticorpale in soggetti immunodepressi.

Validazione di biomarcatori molecolari di esposizione, di effetto biologico precoce e di suscettibilità

Linea di ricerca: 1 – Prevenzione e Cancerogenesi

Programma: b - Biomarcatori biologici e molecolari di esposizione, di danno, di suscettibilità e di rischio di cancro

Responsabile scientifico: Marcello Ceppi

Altro personale della struttura partecipante al progetto: Cecilia Lando

Anno di inizio: 2009

Durata: 36 mesi

Parole chiave: epidemiologia molecolare; biomarker; polimorfismi genetici; danno al DNA; danno cromosomico

Altre strutture IST partecipanti: S.C. Epidemiologia, Biostatistica e Clinical Trials (D. Ugolini, R. Filiberti); S.S. Cancerogenesi Ambientale (C. Bolognesi)

Altri Enti coinvolti: IARC, Lione, Francia (P. Boffetta); FIOH, Helsinki (H. Norppa); CSIRO, Adelaide, Australia (M. Fenech); Berkeley University, Berkeley, USA (N. Holland); CSPO, Firenze (M. Peluso); Università di Pisa (L. Migliore); National Yang Ming University Medical School, Taipei, Taiwan (W. Peter Chang); Istituto Superiore di Sanità, Roma (A. Zijno); Vrije Universiteit, Bruxelles, Belgio (M. Kirsh-Volders); IRCCS S. Raffaele Pisana, Roma (S. Bonassi)

Tipologia progetto: clinico-epidemiologica osservazionale

Area di interesse: prevenzione primaria/secondaria

Soggetti cofinanziatori: Agenzia Spaziale Italiana; Associazione Italiana per la Ricerca sul Cancro; Ministero della Difesa; Regione Emilia-Romagna; Compagnia di San Paolo; Fondazione CARIGE

Background

L'Epidemiologia Molecolare in campo oncologico è finalizzata a: i) conduzione di studi di tipo descrittivo e analitico per valutare il ruolo dell'interazione fra geni e ambiente nell'eziologia, nella storia naturale e nella prognosi delle malattie, ii) prevenzione della patologia e del danno genetico valutando i rischi a livello individuale e di popolazione ed identificando sottogruppi di individui suscettibili tramite screening genetico. In questi ultimi anni le ricerche sull'interazione fra geni e ambiente e sulla validazione dei biomarcatori predittivi di malattia sono state fra le più produttive in letteratura rendendo questa linea di ricerca tra le più promettenti nell'intento di comprendere il contributo dei fattori genetici, dei fattori di rischio ambientali e della loro interazione nell'eziologia del cancro.

Obiettivo generale del progetto ed eventuali obiettivi secondari

Il progetto si articola su diversi endpoints:

- Individuazione di indicatori precoci di cancerogenesi. Scopo della ricerca è di individuare quei biomarcatori che misurati in individui sani permettano di valutare precocemente il rischio di cancro e quindi di consentire tempestivi interventi di prevenzione primaria e/o di chemoprevenzione in gruppi esposti a sostanze supposte cancerogene. Lo studio della associazione fra questi marcatori e lo sviluppo di tumori contribuisce inoltre alla migliore comprensione delle fasi iniziali della cancerogenesi.
- Studio dell'interazione gene-ambiente. Molti geni implicati nella cancerogenesi sono polimorfici e le diverse varianti possono avere differente attività enzimatica con conseguente alterazione della risposta cellulare agli agenti genotossici. I polimorfismi possono pertanto essere correlati ad un diverso rischio di cancro e l'individuazione di sottopopolazioni suscettibili permetterebbe di programmare interventi di prevenzione. L'espressione e l'attività di diversi enzimi variano con l'età e questo può influire sulla sensibilità ai cancerogeni nel corso della vita.
- Validazione di biomarcatori genetici in studi di popolazioni umane. Uno dei maggiori limiti nell'utilizzo dei biomarcatori è che non sempre è chiaro ciò che essi effettivamente misurano. Nell'ottica di un uso clinico di questi indicatori diventa quindi fondamentale la loro validazione in termini di predittori di rischio di cancro. Seguire nel tempo anche retrospettivamente popolazioni umane sottoposte a test genetici fornirà importanti indicazioni per chiarire il ruolo dei biomarcatori nell'ambito della cancerogenesi.

Impatto assistenziale certo o potenziale

Fra le principali aspettative dell'utilizzo di marcatori genetici in clinica vi è la possibilità di perfezionare le procedure diagnostiche ed abbreviarne i tempi, la personalizzazione delle terapie in funzione del background genetico individuale,

Consuntivo 2009 - Programmazione 2010

la comprensione anticipata dell'efficacia del trattamento e l'individuazione precoce (o la predizione) degli eventi avversi.

Risultati e prodotti 2009

Nel 2009 una delle attività principali dell'unità è stata quella di valutare l'utilizzo del conteggio di micronuclei in saggi condotti sulla mucosa esfoliata della bocca (MNmb), un test di facile attuazione in quanto meno invasivo del prelievo ematico, quale biomarcatore di esposizione e/o di effetto biologico precoce. Questo ha portato alla pubblicazione di alcuni lavori sia di tipo metodologico che di validazione del ruolo di biomarcatore di MNmb. Negli studi metodologici è stato fatto lo stato dell'arte del MNmb indicando le procedure di laboratorio più adatte al fine di standardizzare i risultati del test. Uno dei maggiori problemi nell'utilizzo dei biomarcatori è infatti la grande variabilità tra i dati provenienti da laboratori diversi dovuti all'applicazione di metodiche differenti. Lo scopo dei nostri lavori è stato quello di fornire delle guidelines per rendere più omogeneo il test. Analogo discorso è stato fatto per l'analisi statistica del MNmb. In un lavoro sono stati suggeriti i metodi di analisi statistica del MNmb più corretti ed innovativi, facendo anche accenno a modelli matematici mutuati da altre discipline ma non usati correntemente in studi di biomarcatori. Nello stesso studio è stato anche valutato l'impatto sul MNmb dello stile di vita, dell'esposizione occupazionale e dello stato di malattia tramite una metanalisi della recente letteratura. Alla fine dello scorso anno è terminata la raccolta dei dati individuali di MNmb provenienti da oltre 30 laboratori sparsi nel mondo (database HUMN-xl). Si tratta di circa 6000 test corredati di informazioni anagrafiche, sullo stato di salute, sull'esposizione a genotossici e sulle abitudini individuali (fumo, alcool, dieta) che una volta organizzati e controllati ci permetteranno di fare accurate valutazioni sull'attività di biomarcatore di esposizione e/o di danno del MNmb.

In uno studio sull'effetto dell'esposizione a pesticidi sul genoma è stato valutato il danno al DNA tramite comet-assay. Effettivamente l'analisi condotta dalla nostra unità ha evidenziato un eccesso di danno nei lavoratori esposti.

Altri studi in cui la struttura è stata coinvolta hanno riguardato l'analisi dei polimorfismi ovvero le varianti di uno stesso gene meno frequenti nella popolazione generale, al fine di valutare la loro eventuale implicazione nella cancerogenesi.

Pubblicazioni

Banelli B.-Bonassi S.-Casciano I.-Mazzocco K.-Di Vinci A.-Scaruffi P.-Brigati C.-Allemanni G.-Borzi' L.-Tonini G.P.-Romani M.

Outcome Prediction And Risk Assessment By Quantitative Pyrosequencing Methylation Analysis Of The Sfn Gene In Advanced Stage, High Risk, Neuroblastic Tumor Patients.

Int. J. Cancer Epub Jul 22, 2009

Bhalli Ja- Ali T- Asi Mr- Khalid Zm- Ceppi M- Khan Qm.

Dna Damage In Pakistani Agricultural Workers Exposed To Mixture Of Pesticides.

Environ Mol Mutagen. 2009 Jan;50(1):37-45.

Bonassi S.-Biasotti B.-Kirsch Volders M.-Knasmueller S.-Zeiger E.- Burgaz S.-Bolognesi C.-Holland N.-Thomas P.-Fenech M.-Humn(XI) Project Consortium.

State Of The Art Survey Of The Buccal Micronucleus Assay. A First Stage In The Humn(XI) Project Initiative.

Mutagenesis 24:295/302, 2009

Ceppi M.-Biasotti B.-Fenech M.-Bonassi S.

Human Population Studies With The Exfoliated Buccal Micronucleus Assay: Statistical And Epidemiological Issues.

Mutat. Res. Rev. Mutat. Epub Nov 19, 2009

Cesario A.-Dall'armi V.-Cusumano G.-Ferri L.-Margaritora S.- Cardaci V.-Cafarotti S.-Russo P.-Paleari L.-Sterzi S.-Pasqua F.- Bonassi S.-Granone P.

Post Operative Pulmonary Rehabilitation After Lung Resection For Nscl: A Follow Up Study.

Lung Cancer 66:268/269, 2009

Fenech M, Holland N, Knasmueller S, Burgaz S, Bonassi S.

Report on the buccal micronucleus assay workshop organized by the International Human Micronucleus (HUMN) project--Antalya, Turkey 2007.

Mutagenesis Mar;24(2):199-201, 2009

Rossi A.-Hansteen I.-Skjelbred C.-Ballardin M.-Maggini V.-Murgia E.-Tomei A.-Viarengo P.-Knudsen L.-Barale R.-Norppa H.-Bonassi S.

Association Between Frequency Of Chromosomal Aberrations And Cancer Risk Is Not Influenced By Genetic Polymorphisms In Gstm1 And Gstt1.

Environ. Health Perspect. 117:203/208, 2009

Thomas P.-Holland N.-Bolognesi C.-Kirsch Volders M.-Bonassi S.- Zeiger E.-Knasmueller S.-Fenech M.

Buccal Micronucleus Cytome Assay.

Nat. Protoc. 4:825/837, 2009

Attività previste e risultati attesi nel 2010

Nel 2010 verrà effettuata l'analisi del database HUMN-xl. Per compensare l'eccesso di variabilità del test dovuta a procedure di laboratorio non completamente standardizzate per i markers più innovativi, verranno attuate procedure di standardizzazione statistica e/o si farà ricorso a modelli statistici adatti all'analisi di dati organizzati in "cluster" (ad es. un laboratorio), che tengano conto della variabilità "dentro cluster" rispetto a quella "tra cluster". Le alterazioni del materiale genico verranno messe in relazione allo stato di malattia degli individui e/o all'esposizione a sostanze genotossiche rispetto ad individui sani/non esposti al fine di validare il biomarcatore in quanto indicatore di

Consuntivo 2009 - Programmazione 2010

danno/esposizione a cancerogeni. Validare il MNmb assay permetterà di disporre di un marcatore presente nell'organo bersaglio che nel caso di displasie del tratto oro-faringeo è atteso essere più sensibile ad un danno precoce rispetto ai markers valutati nei linfociti, con le conseguenti implicazioni cliniche.

Nel corso dell' anno sarà portata avanti l'analisi dei dati di uno studio volto a correlare le abitudini alimentari con l'insorgenza di micronuclei nei linfociti del sangue periferico.

E' in via di definizione uno studio atto a confrontare il danno alle basi del DNA in fumatori e non-fumatori.

E' appena iniziata l'analisi di un indicatore di funzionalità respiratoria misurato all'inizio ed alla fine di un programma riabilitativo applicato a soggetti portatori di patologie benigne del polmone.

E' in fase di pubblicazione uno studio sui polimorfismi degli enzimi di riparo del danno al DNA in relazione agli addotti al DNA condotto su tre gruppi di soggetti: pazienti di cancro polmonare, ammalati di patologie benigne del polmone e controlli sani.

Uno studio su dieta e malondialdehyde–deoxyguanosine adducts è in fase di pubblicazione.