

PROGETTO DI RICERCA – LILT TRENTO

Diagnosi precoce non invasiva di carcinoma colon-rettale: un approccio integrato basato su marcatori di RNA non codificante e di microbioma intestinale

Introduzione

Il carcinoma colon-rettale (CRC) è la terza neoplasia più frequente per incidenza nell'uomo e la seconda nella donna a livello mondiale. Uno screening efficace potrebbe condurre ad una riduzione sia dell'incidenza sia della mortalità. Ad oggi, l'esame che permette l'individuazione in modo accurato e affidabile del CRC è la colonscopia che può portare ad una riduzione dell'incidenza del tumore di circa il 60-90%. La colonscopia rappresenta però una metodica invasiva, costosa per il sistema sanitario e richiede uno staff preparato per l'esecuzione¹.

La ricerca scientifica nell'ambito della malattia colon-rettale negli ultimi anni si è orientata verso un miglioramento della prevenzione mirato all'identificazione di nuovi marcatori in grado di discriminare i soggetti con lesioni precancerose da quelli con CRC, soprattutto analizzando tessuti surrogati quali ad esempio plasma, siero o feci. Evidenze in letteratura hanno mostrato che alcune specifiche alterazioni genetiche/epigenetiche e determinate configurazioni della popolazione microbica intestinale (microbioma) sembrano essere dei candidati ideali sia per la diagnosi sia per la prognosi¹.

Gli *small non-coding RNA* (sncRNA) sono piccole molecole di RNA non codificante le cui funzioni vanno dal controllo trascrizionale alle modificazioni cromatiniche nel nucleo o di regolazione della traduzione nel citoplasma. Tra le varie peculiarità attribuite a queste molecole, c'è il loro potenziale ruolo rilevante nell'insorgenza di tumori, tra cui il CRC. Gli sncRNA sono presenti in tutti i tessuti e sono stati rilevati anche nei tessuti surrogati. Al momento sono stati ottenuti dei risultati molto promettenti per un insieme di microRNA (miRNA, una delle classi di sncRNA tra le più studiate) capace di discriminare CRC e lesioni precancerose considerando una coorte limitata di campioni di plasma²⁻⁴. Si tratta comunque di dati basati esclusivamente su coorti di pazienti poco numerose. Ad oggi non è ancora stato identificato un insieme di sncRNA che consenta una stratificazione dei pazienti in modo accurato rispetto alla lesione che presentano per cui la colonscopia è ancora l'unico strumento di diagnosi accurata per il CRC.

Un'altra evidenza riportata in letteratura è la dimostrazione di come la configurazione del microbioma intestinale analizzato da campioni fecali sia strettamente associato a varie patologie intestinali (per esempio la rettocolite ulcerosa o il morbo di Crohn) e in particolare al cancro colon-rettale. Studi preliminari⁵⁻⁸ a bassa risoluzione hanno inizialmente evidenziato questa associazione recentemente studiata anche con metodi di metagenomica di tipo *shotgun* ad alta risoluzione⁸.

Obiettivo principale del progetto

In questo progetto proponiamo **la definizione di un pannello di marcatori identificati tra sncRNA umani e marcatori microbici rilevabili in campioni di feci per la diagnosi precoce del CRC**. Il pannello comprenderà un insieme di marcatori da utilizzare **come efficace strumento di screening non invasivo ed altamente sensibile**, indipendentemente o in concomitanza di altre metodiche in uso e che consenta di stratificare i pazienti (con tumore, lesioni precancerose ad alto rischio, oppure sani con infiammazioni) analizzando esclusivamente campioni di feci.

Razionale

La ricerca di biomarcatori sia a livello di RNA umano, sia di DNA microbico per la diagnosi di CRC permetterà di ottenere uno strumento accurato, a basso impatto economico e con un'influenza limitata sul paziente che ha la potenzialità di migliorare gli attuali metodi impiegati per lo screening del CRC. La metodologia scientifica che proponiamo per ottenere questo insieme di biomarcatori è basata su un approccio integrato sperimentale e computazionale i cui dati sono ottenuti da tecnologie di sequenziamento di ultima generazione.

Questo progetto presenta quattro aspetti essenziali ed innovativi.

- **L' utilizzo di tessuti surrogati.** La scelta del tessuto che andremo ad analizzare avrà un basso impatto sia a livello di costi sia sul paziente, essendo le feci dei campioni biologici facilmente reperibili e le analisi non invasive. Inoltre, la nostra ipotesi sperimentale è che le feci, essendo a diretto contatto con il sito delle lesioni in esame, siano in grado di riflettere accuratamente e anticipatamente l'insieme di alterazioni fortemente correlate con l'insorgenza della patologia tumorale.
- **L'analisi integrata del microbioma intestinale e di RNA non codificanti umani.** Lo studio della diversità microbica presente nell'intestino sta rivoluzionando lo studio di molte malattie gastrointestinali ed è già stato associato a CRC. Tuttavia, nuovi campioni sequenziati con tecnologie di ultima generazione ed approcci computazionali ad alta risoluzione come quelli sviluppati dal nostro team di ricerca sono necessari per ottenere biomarcatori altamente predittivi. Inoltre, l'utilizzo combinato di biomarcatori microbici e di sncRNA non è mai stato provato ma ha grandissime potenzialità in quanto approcci complementari.
- **L'impiego di tecnologie di ultima generazione.** Le analisi sia a livello di sncRNA sia di microbioma verranno eseguite mediante l'utilizzo di tecnologie di sequenziamento di nuova generazione (*Next Generation Sequencing* - NGS). I proponenti del progetto hanno tutti una provata esperienza nell'utilizzo di tecnologie NGS che hanno permesso di effettuare una ricerca *genome-wide* e *metagenome-wide* non vincolata su dati ed informazioni già note ed effettuando un salto di qualità nella ricerca di biomarcatori *de novo*.
- **La creazione di un team di ricercatori multidisciplinari.** La riuscita di questo progetto presuppone conoscenze in diverse discipline, dalla biologia all'informatica e dalla clinica alla biologia computazionale. I ricercatori che propongono questo progetto sono specializzati in diverse discipline quindi costituiscono un gruppo eterogeneo capace di affrontare il problema proposto in maniera multidisciplinare; i ricercatori sono abituati a lavorare in gruppi di ricerca dove l'efficienza e la riuscita dei lavori presuppone una forte interazione tra persone di ambiti scientifici diversi.

Dati e risultati preliminari

Abbiamo standardizzato le condizioni sperimentali per l'isolamento di sncRNA, sia da feci che da plasma e tessuti e le analisi dei livelli di espressione dei miRNA presenti in plasma e feci mediante tecniche sperimentali convenzionali (PCR quantitativa)⁹. In un primo lavoro sono stati identificati specifici miRNA che permettevano in base ai loro profili di espressione di discriminare CRC e lesioni precancerose¹⁰.

Abbiamo anche effettuato analisi preliminari su campioni di feci e di plasma (esosomi) utilizzando la tecnica di sequenziamento NGS. I risultati ottenuti in diversi campioni biologici, confrontati con dati di sequenziamento da campioni di cellule (esfoliato cervicale) sono positivi. È interessante notare che il numero medio di sequenze ottenute dagli esperimenti di NGS (*read*) in feci e esosomi

plasmatici è paragonabile a quello ottenuto in tessuti solidi. Nelle analisi di questi dati il primo passo consiste nella eliminazione delle sequenze di adattatori, che porta ad una riduzione dell'insieme di dati di circa il 50%. Le sequenze vengono quindi mappate rispetto alle sequenze genomiche al fine di identificarne la provenienza genomica. Il numero medio di miRNA identificati sia nei campioni di feci sia in quelli plasmatici sono confrontabili con i dati in letteratura. E' interessante notare che per i campioni di feci è possibile l'utilizzo di più genomi di riferimento che permetta di aumentare la percentuale di sequenze mappate tra le *read* non specifiche per gli sncRNA.

Abbiamo anche svolto un'analisi preliminare dei dati metagenomici disponibili in relazione al CRC. In particolare, abbiamo rianalizzato i dati ad alta risoluzione ottenuti con la metagenomica da Zeller et al¹¹ utilizzando i nuovi metodi sviluppati dai membri della presente proposta per identificare organismi microbici a livello di ceppo¹²⁻¹⁴. Con questi nuovi metodi che possono trovare marcatori a livelli tassonomici e funzionali molto specifici, abbiamo ottenuto una predittività media superiore all'85% in particolare con l'utilizzo di marcatori specie-specifici. Nei nostri laboratori svolgiamo di routine analisi metagenomiche su campioni fecali associate ad altre patologie comprendenti tutti gli step sperimentali necessari (raccolta campioni, estrazione DNA, costruzione librerie di sequenziamento, sequenziamento su piattaforma Illumina HiSeq, analisi computazionale dei dati di sequenziamento) confermando quindi la fattibilità dell'approccio sperimentale di questa proposta. E' da sottolineare che noi ci focalizzeremo sullo studio del microbioma con *shotgun sequencing*, una tecnologia molto più ad alta risoluzione rispetto alla tecnica precedente basata sul sequenziamento del gene batterico 16S che ha tuttavia permesso per la prima volta di associare una disbiosi batterica intestinale con CRC⁵⁻⁸.

Metodi

Popolazione di studio

I pazienti arruolati in questo studio saranno soggetti con sintomatologia suggestiva per patologia del colon e del retto, che siano stati inviati dal Medico Curante per accertamenti specialistici e per l'esecuzione della colonscopia, ovvero pazienti con patologia del colon retto, presso la Clinica Santa Rita di Vercelli (Resp. dello studio Dott. Giuseppe Clerico).

Criteri di inclusione:

- Soggetti caucasici.
- Pazienti di entrambe i sessi, età > 30 anni.
- Anamnesi familiare negativa per tumore del colon retto.
- Anamnesi patologica negativa per neoplasie concomitanti o precedenti (compreso CRC).
- Pazienti che abbiano eseguito una colonscopia completa
- Pazienti che abbiano firmato il consenso informato per la partecipazione allo studio clinico
- Pazienti che forniscano unitamente al campione di materiale biologico , i questionari EPIC sulle abitudini alimentari e sullo stile di vita debitamente compilati.

In base a questi criteri di inclusione sono attesi i seguenti diversi gruppi di pazienti:

- 1) Soggetti sani
- 2) Soggetti con patologia colo rettale di tipo infiammatorio
- 3) Soggetti con lesioni adenomatose
- 4) Soggetti con CRC (nuova diagnosi)

Il reclutamento dei partecipanti in collaborazione con la Clinica S. Rita è attualmente in corso; il disegno dello studio è stato approvato dal Comitato Etico locale. Al momento sono stati raccolti oltre 70 campioni biologici, con relativi dati clinici e demografici dei pazienti. Per il presente progetto si prevede di raggiungere i 200 soggetti per arrivare ad avere un'equa distribuzione tra le categorie sopra elencate e avere un sufficiente potere statistico.

Informazioni dai questionari

Ai partecipanti verrà somministrato un questionario destinato a raccogliere l'anamnesi del paziente con dettagli riguardanti lo stile di vita e i dati storici personali, comprese le caratteristiche socio-demografiche, le esposizioni legate al lavoro, il fumo, le abitudini alimentari, la pratica di attività fisica, l'uso di contraccettivi e la storia riproduttiva, uso della terapia ormonale sostitutiva, la storia di malattie precedenti e/o in corso, qualsiasi trattamento medico/chirurgico e ricovero in ospedale.

Raccolta e trattamento dei campioni

I campioni di feci evacuati naturalmente saranno ottenuti da tutti i pazienti prima di iniziare la pulizia dell'intestino per la colonscopia. Le feci saranno raccolte in appositi tubi di raccolta contenenti una soluzione stabilizzante per DNA ed RNA (Norgen Biotek Corp) e restituite al momento della colonscopia nell'unità di endoscopia o al momento del prelievo di sangue. Aliquote di campioni di feci saranno conservati a -80°C fino al momento dell'estrazione dell'RNA. Cinque ml di sangue saranno raccolti in provette EDTA prima di eseguire la preparazione intestinale per la colonscopia. Il sangue sarà centrifugato a 1,200g per 10 minuti a 4°C. Il plasma verrà separato dal materiale cellulare, aliquotato e conservato a -80°C.

Identificazione di sncRNA

L'RNA dalle feci verrà estratto utilizzando lo Stool Total RNA Purification Kit (Norgen Biotek Corp). Gli esosomi saranno isolati dal plasma con la soluzione ExoQuick (System Biosciences) e l'RNA verrà estratto con il kit miRNeasy plasma/siero (Qiagen). La qualità dell'RNA sarà verificata secondo le linee guida MIQE (<http://mige.gene-quantification.info/>), mentre la concentrazione sarà quantificata mediante Qubit® microRNA Assay Kit (Invitrogen).

Lo studio può contare su un protocollo di lavoro già ottimizzato¹⁵ e su analisi di dati provenienti da uno studio in corso con risultati di NGS in plasma e feci di soggetti con CRC, lesioni precancerose o sani arruolati in collaborazione con la struttura endoscopica sopra citata. I dati raccolti aiuteranno a identificare un pannello di sncRNA legati alla presenza di lesioni cancerose. La complessità e la quantità di sncRNA saranno caratterizzate in feci ed in esosomi plasmatici mediante tecnologia NGS (Resp. Dott. Alessio G. Naccarati, Dott.ssa Barbara Pardini). La preparazione delle *library* verrà eseguita utilizzando il NEBNext Multiplex small RNA library Prep Set (protocollo E7330, New England BioLabs). Frammenti di 135-160 bp saranno utilizzati per l'analisi di RNA-Seq sulla piattaforma Illumina HiSeq2000 (Illumina). Una *flow cell* contenente 24 campioni indicizzati produrrà circa 4 milioni di *read* singole per campione di una lunghezza di circa 50 nucleotidi. Il sequenziamento dei campioni si svolgerà presso la *facility* di NGS del CIBIO (Trento).

I dati generati dal sequenziamento NGS saranno analizzati con un work flow ottimizzato per l'espressione quantitativa dei sncRNA (Resp. Prof.ssa Francesca Cordero). Le *read* saranno mappate con l'algoritmo SHRiMP rispetto ad un insieme di sequenze di riferimento (e.g miRNA, rRNA, tRNA, snRNAs e snoRNAs), i sncRNA differentemente espressi saranno validati mediante qRT-PCR.

Caratterizzazione del microbioma intestinale dai campioni fecali

Il sequenziamento metagenomico di tipo “shotgun” sarà eseguito presso la *facility* di NGS del CIBIO sulla piattaforma Illumina HiSeq utilizzando i materiali e protocolli Nextera XT per la preparazione delle librerie di sequenziamento (Resp. Prof. Nicola Segata). Vista la nostra esperienza nella raccolta, trasporto, e estrazione del DNA da campioni fecali, non prevediamo problemi nel processamento dei campioni nei nostri laboratori. Nello specifico, utilizziamo la metodologia che abbiamo validato durante il nostro lavoro passato e attuale con il progetto microbioma umano (Human Microbiome Project consortium) le cui specifiche sono disponibili al seguente link http://hmpdacc.org/doc/HMP_MOP_Version12_0_072910.pdf e utilizzate nei lavori ufficiali del consorzio^{16,17}. Genereremo un minimo di 5Gb per campione con sequenziamento *paired-end* con lunghezza delle *read* pari a 100nt per frammenti di ~250nt. Tale profondità di sequenziamento è sufficiente a coprire con 1X di coverage un genoma batterico di 5MB presente ad un’abbondanza relative di 0.1% (dopo la rimozione del DNA umano contaminante) e identificare organismi ad un’abbondanza di solo 0.01%. Analizzeremo i dati con i nostri approcci che rappresentano lo stato dell’arte nel campo della metagenomica¹⁸ e applicheremo le nostre pipeline aggiornate per le procedure di controllo qualità e normalizzazione/standardizzazione. Inoltre caratterizzeremo i campioni dal punto di vista funzionale, tassonomico, filogenetico^{12,19}. Stiamo anche ultimando nuovi metodi computazionali per migliorare la risoluzione della caratterizzazione tassonomica con approcci funzionali e tassonomici capaci di arrivare al livello di singoli ceppi. Integreremo poi i dati generati con quelli disponibili da altri studi su soggetti sani e da sequenziamenti disponibili per pazienti con CRC.

Analisi statistiche e potenza dello studio

Saranno costruite curve ROC e le conseguenti AUC calcolate per valutare la specificità e la sensibilità dei marcatori identificati. Un metodo basato su un classificatore *semi-supervised* sarà applicato per prevedere le relazioni tra i biomarcatori candidati e le diverse covariate (sesso, età, stadio del tumore, grado, ecc.) e applicheremo anche varianti degli strumenti che abbiamo sviluppato appositamente per la ricerca di biomarcatori²⁰. Le *signature* di sncRNA e di marcatori microbici identificate saranno analizzate integrando i dati provenienti da questionari sulle abitudini alimentari lo stile di vita.

Al momento non esistono linee guida per stimare le dimensioni necessarie di un campione negli esperimenti di RNA-seq per avere una sufficiente potenza statistica, in particolare per i *non-coding* RNA. A tal proposito abbiamo considerato però l’algoritmo proposto da Hart e collaboratori²¹. In base ai nostri studi pilota, abbiamo stimato un numero minimo di 50 *read* di sncRNA mappate ed un coefficiente di variazione interindividuale di 0.8. Considerando una differenza minima tra due gruppi a confronto (ad esempio CRC vs. controlli) di 2.0 nei livelli di espressione degli RNA e un numero di falsi positivi inferiori al 5%, siamo confidenti di avere una sufficiente potenza statistica (>0.80) con le dimensioni del campione proposto, anche tenuto conto di opportune correzioni nei livelli di significatività per confronti multipli (5-10 sncRNA). Dal punto di vista dell’analisi del microbioma, lo standard internazionale in termini di profondità di sequenziamento per campioni fecali è di 5GB per campione come descritto sopra, e la potenza statistica è analoga a quanto descritto per gli esperimenti di sncRNA considerando due livelli di differenza sull’abbondanza relativa di ogni feature metagenomica.

REFERENZE

- 1 Brenner, H., Kloor, M. & Pox, C. P. Colorectal cancer. *Lancet* **383**, 1490-1502, doi:10.1016/S0140-6736(13)61649-9 (2014).
- 2 Lujambio, A. & Lowe, S. W. The microcosmos of cancer. *Nature* **482**, 347-355, doi:10.1038/nature10888 (2012).
- 3 Xue, B. & He, L. An expanding universe of the non-coding genome in cancer biology. *Carcinogenesis* **35**, 1209-1216, doi:10.1093/carcin/bgu099 (2014).
- 4 Di Leva, G. & Croce, C. M. miRNA profiling of cancer. *Current opinion in genetics & development* **23**, 3-11, doi:10.1016/j.gde.2013.01.004 (2013).
- 5 Kostic, A. D. *et al.* Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe* **14**, 207-215, doi:10.1016/j.chom.2013.07.007 (2013).
- 6 Kostic, A. D. *et al.* Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. *Genome research* **22**, 292-298, doi:10.1101/gr.126573.111 (2012).
- 7 Marchesi, J. R. *et al.* Towards the human colorectal cancer microbiome. *PLoS One* **6**, e20447, doi:10.1371/journal.pone.0020447 (2011).
- 8 Zackular, J. P. *et al.* The gut microbiome modulates colon tumorigenesis. *MBio* **4**, e00692-00613, doi:10.1128/mBio.00692-13 (2013).
- 9 Tarallo, S. *et al.* MicroRNA expression in relation to different dietary habits: a comparison in stool and plasma samples. *Mutagenesis* **29**, 385-391, doi:10.1093/mutage/geu028 (2014).
- 10 Aherne, S. T. *et al.* Circulating miRNAs miR-34a and miR-150 associated with colorectal cancer progression. *BMC Cancer* **15**, 329, doi:10.1186/s12885-015-1327-5 (2015).
- 11 Zeller, G. *et al.* Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol Syst Biol* **10**, 766, doi:10.15252/msb.20145645 (2014).
- 12 Truong, D. T. *et al.* MetaPhlan2 for enhanced metagenomic taxonomic profiling. *Nature methods* **12**, 902-903, doi:10.1038/nmeth.3589 (2015).
- 13 Segata, N., Bornigen, D., Morgan, X. C. & Huttenhower, C. PhyloPhlan is a new method for improved phylogenetic and taxonomic placement of microbes. *Nature communications* **4**, 2304, doi:10.1038/ncomms3304 (2013).
- 14 Segata, N. *et al.* Metagenomic microbial community profiling using unique clade-specific marker genes. *Nature methods* **9**, 811-814, doi:10.1038/nmeth.2066 (2012).
- 15 Cordero, F., Beccuti, M., Arigoni, M., Donatelli, S. & Calogero, R. A. Optimizing a massive parallel sequencing workflow for quantitative miRNA expression analysis. *PLoS One* **7**, e31630, doi:10.1371/journal.pone.0031630 (2012).
- 16 Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207-214, doi:10.1038/nature11234 (2012).
- 17 Segata, N. *et al.* Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. *Genome biology* **13**, R42, doi:10.1186/gb-2012-13-6-r42 (2012).
- 18 Segata, N. *et al.* Computational meta'omics for microbial community studies. *Mol Syst Biol* **9**, 666, doi:10.1038/msb.2013.22 (2013).
- 19 Abubucker, S. *et al.* Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS computational biology* **8**, e1002358, doi:10.1371/journal.pcbi.1002358 (2012).
- 20 Segata, N. *et al.* Metagenomic biomarker discovery and explanation. *Genome biology* **12**, R60, doi:10.1186/gb-2011-12-6-r60 (2011).
- 21 Hart, S. N., Therneau, T. M., Zhang, Y., Poland, G. A. & Kocher, J. P. Calculating sample size estimates for RNA sequencing data. *J Comput Biol* **20**, 970-978, doi:10.1089/cmb.2012.0283 (2013).