

CURRICULUM VITAE

Dott. Sola Lorenzo

A. INFORMAZIONI PERSONALI

Data di nascita: 13 Dicembre 1996

Luogo di nascita: Pavia, Italia

Contatti: lorenzo.sola@unipv.it/lorenzo.sola01@universitadipavia.it

ORCID: <https://orcid.org/0000-0003-3149-2530>

GitHub: <https://github.com/LorenzoS96>

B. POSIZIONE ATTUALE

Assegnista di Ricerca presso il laboratorio di Genomica Computazionale, Dipartimento di Biologia e Biotecnologia “Lazzaro Spallanzani”, Università di Pavia (Supervisore: Prof. Francesco Lescai).

C. ISTRUZIONE

Dottorato di ricerca (Ph.D.) in Genetica, Biologia Molecolare e Cellulare, 27 Marzo 2024

Università di Pavia, Dipartimento di Biologia e Biotecnologie “Lazzaro Spallanzani”, Pavia, Italia

Tesi sperimentale: “*Telomeres, Interstitial Telomeric Sequences and their transcription*”

Relatore: Prof.ssa Elena Giulotto

Laurea Magistrale in “Molecular Biology and Genetics”, 110/110 *cum laude*, 20 Luglio 2020

Università di Pavia, Dipartimento di Biologia e Biotecnologie “Lazzaro Spallanzani”, Pavia, Italia

Tesi sperimentale: “*Telomerase RNA retrotranscription at DNA double-strand breaks throughout vertebrate genome evolution*”

Relatore: Prof.ssa Elena Giulotto; Correlatore: Prof. Solomon Nergadze

Laurea Triennale in Scienze Biologiche, 110/110, 27 Luglio 2018

Università del Piemonte Orientale “Amedeo Avogadro”, Alessandria, Italia

Tesi sperimentale: “*L’antibiotico resistenza negli Enterobatteri, in particolare quella alla colistina sembrava impossibile: metodi diagnostici della resistenza in E.coli uropatogeni a confronto*”

Relatore: Prof.ssa Elisa Gamalero

D. POSIZIONI

Ricerca ed esperienze professionali

Gennaio 2024 – Presente: attività di ricerca presso il laboratorio di Genomica Computazionale, Dipartimento di Biologia e Biotecnologia “Lazzaro Spallanzani”, Università di Pavia (Supervisore: Prof. Francesco Lescai)

Ottobre 2020 – Dicembre 2023: attività di ricerca durante il dottorato di ricerca in Genetica, Biologia Molecolare e Cellulare nel Laboratorio di Biologia Molecolare e Cellulare, Dipartimento di Biologia e Biotecnologie “Lazzaro Spallanzani”, Università di Pavia (Supervisori: Prof.ssa Elena Giulotto e Prof. Solomon Nergadze)

Ottobre 2018 – Luglio 2020: internato di ricerca per la preparazione ed il conseguimento della laurea magistrale nel Laboratorio di Biologia Molecolare e Cellulare, Dipartimento di Biologia e Biotecnologie “Lazzaro Spallanzani”, Università di Pavia (Supervisori: Prof.ssa Elena Giulotto e Prof. Solomon Nergadze)

Gennaio 2018 – Giugno 2018: internato per la preparazione ed il conseguimento della laurea triennale nel Laboratorio di Biologia Molecolare e Microbiologia, Ospedale Civile “Ss. Antonio e Biagio e Cesare Arrigo”, Alessandria (Supervisore: Dottor Andrea Rocchetti)

Altre esperienze, incluse attività di insegnamento:

Settembre 2020 – Giugno 2022: tutor per il laboratorio di “Biotecnologie Molecolari” (Supervisore: Prof. Solomon Nergadze)

Gennaio 2020 – Giugno 2020: studente part-time per l’assistenza di studenti stranieri per il corso di laurea magistrale “Molecular Biology and Genetics” (Supervisore: Prof.ssa Elena Giulotto)

Correlatore delle seguenti tesi di laurea:

- Tesi di Laurea Magistrale in Molecular Biology and Genetics di Reem Hijaz, Università di Pavia, Anno accademico 2020/2021, Titolo: *Investigation of the transcriptional status of satellite-less centromeres in the genus Equus*. Relatore: Prof.ssa Elena Giulotto, Correlatori: Dr.ssa Eleonora Cappelletti e Dr. Lorenzo Sola
- Tesi di Laurea Triennale in Biotecnologie di Irene Nicosia, Università di Pavia, Anno accademico 2022/2023, Titolo: *Valutazione della trascrizione di sequenze telomeriche interstiziali in fibroblasti primari e line cellulari tumorali*. Relatore: Prof. Solomon Nergadze, Correlatore: Dr. Lorenzo Sola
- Tesi di Laurea Triennale in Biotecnologie di Matteo Re, Università of Pavia, Anno accademico 2022/2023, Titolo: *Analisi del profilo di espressione di loci subtelomerici in linee cellulari umane*. Relatore: Prof.ssa Elena Giulotto, Correlatore: Dr. Lorenzo Sola

E. ATTIVITÀ DI RICERCA

1) TELOMERI, SEQUENZE TELOMERICHE INTERSTIZIALI E LORO TRASCRIZIONE

I telomeri sono complessi ribonucleoproteici che costituiscono le regioni terminali dei cromosomi lineari eucariotici. La componente nucleotidica, nei vertebrati, consiste nella ripetizione in tandem della sequenza TTAGGG. La parte proteica è costituita da un complesso, chiamato Shelterina, in grado di legare in modo specifico il DNA telomerico. In cellule germinali, staminali e nella maggior parte dei tumori (>90%) i telomeri vengono mantenuti dall’attività di un enzima specializzato chiamato telomerasi. La telomerasi è una ribonucleoproteina composta dalla trascrittasi inversa TERT (Telomerase Reverse Transcriptase) e da una molecola di RNA (TERC).

Sin dalla loro scoperta i telomeri sono stati considerati regioni trascrizionalmente silenti a causa della loro natura eterocromatica. Nel 2007 è stato dimostrato che le sequenze telomeriche sono trascritte in una famiglia di lunghi RNA non-codificati chiamata TERRA (Telomeric Repeat-containing RNA) (Azzalin et al., *Science*, 2007). La trascrizione dei telomeri è sostenuta a partire da una sequenza promotrice localizzata nella regione subtelomerica e caratterizzata da ripetizioni ricche in CpG (Nergadze et al., *RNA*, 2009).

Ripetizioni in tandem dell’esamero TTAGGG sono presenti non solo ai telomeri ma anche in siti cromosomici interstiziali (non telomerici e non centromerici) dove prendono il nome di sequenze telomeriche interstiziali (ITS). Queste sequenze sono state identificate in molte specie di vertebrati.

i. Trascrizione dei Telomeri e delle Sequenze Telomeriche Interstiziali

In seguito alla recente pubblicazione della prima versione completa del genoma umano (genoma T2T), durante il mio lavoro di tesi ho identificato oltre 500 siti telomerici interstiziali. Alcune di queste ITS sono caratterizzate dalla presenza di un promotore simile a quello già descritto nelle regioni subtelomeriche in grado di sostenere la trascrizione telomerica. Grazie alla nuova sequenza genomica ho disegnato delle coppie di primer in grado di amplificare TERRA derivante da regioni subtelomeriche specifiche e TERRA derivante da regioni interstiziali tramite la tecnica di qPCR. Queste sequenze sono difficili da studiare per la presenza di regioni estremamente ripetute a livello della sequenza promotrice. L'analisi di dati di RNA-seq ha evidenziato la difficoltà nell'identificazione dei trascritti di TERRA a livello bioinformatico per la presenza di regioni altamente ripetute.

In questo lavoro abbiamo dimostrato la trascrizione delle ITS in una nuova classe di lunghi RNA non-codificati precedentemente non caratterizzata e distinta dalla componente trascritta dai telomeri. Data la localizzazione abbiamo definito queste molecole con il termine di TERRA interstiziale (iTERRA). I risultati ottenuti durante il mio percorso di dottorato saranno oggetto di una pubblicazione attualmente in fase di stesura.

Lorenzo Sola tesi di dottorato in Genetica, Biologia Molecolare e Cellulare

Oltre a esperimenti di biologia molecolare ho svolto analisi bioinformatiche per lo studio di dati di RNA-seq disponibili online. Ho utilizzato diversi programmi di allineamento e di mappaggio finalizzati alla determinazione della migliore metodologia per la quantificazione delle *read* associate ai trascritti TERRA. In particolare, lo studio si è basato sull'utilizzo di programmi quali Bowtie2, STAR e Salmon con combinazioni di diversi parametri adatti ad analizzare sequenze ripetute. L'analisi ha previsto anche l'utilizzo di altre tipologie di software quali Samtools, deepTools, IGV, FastQC e TrimGalore!. Per l'analisi del genoma T2T sono stati usati programmi di analisi di sequenze di acidi nucleici (Multalin e RepeatMasker) oltre a diversi database e portali biologici (UCSC, NCBI, ENCODE). La quasi totalità delle analisi è stata svolta avvalendosi di un cluster HPC. In aggiunta, lo studio ha previsto l'utilizzo di R per l'analisi dei risultati ottenuti dal processamento dei dati di RNA-seq e per la creazione di grafici volti all'analisi dei dati di qPCR (pacchetto *tydiverse*). Grafici utili ai fini dello studio sono stati ottenuti grazie anche al linguaggio Python.

ii. Telomeric-like repeats flanked by sequences retrotranscribed from the telomerase RNA inserted at DNA double-strand break sites during vertebrate genome evolution

L'analisi di numerose ITS nel genoma umano e di topo e dei loro loci ortologhi in diverse specie di primati e roditori aveva precedentemente dimostrato come tali sequenze siano state inserite all'interno dei genomi in seguito a rotture a doppio filamento del DNA attraverso il meccanismo di non-homologous end joining (NHEJ). La presenza, nel genoma di topo e cavallo, di sequenze telomeriche interstiziali fiancheggiate da porzioni retrotrascritte dalla componente a RNA della telomerasi (differenti dallo stampo delle ripetizioni telomeriche) e definite come TERC-ITS ha evidenziato un diretto coinvolgimento della telomerasi nella riparazione delle rotture a doppio filamento del DNA.

Ho condotto un'estesa analisi evolutiva nel genoma di 30 specie di vertebrati, incluso l'uomo, la quale ha rivelato la presenza di loci TERC-ITS nella maggior parte delle specie analizzate. L'analisi ha dimostrato come diverse regioni della componente a RNA della telomerasi vengano coinvolte in specie diverse. Ho inoltre dimostrato che il coinvolgimento della telomerasi nel meccanismo di Non Homologous End Joining è conservato nei vertebrati.

La pubblicazione è basata sull'analisi bioinformatica di 30 genomi di vertebrati con l'utilizzo di database biologici (UCSC e NCBI), di algoritmi di allineamento (BLAST e BLAT) e di programmi per l'analisi di sequenze di acidi nucleici (Multalin e RepeatMasker).

2) STRUTTURA, FUNZIONE ED VOLUZIONE DEI CENTROMERI NEL MODELLO DEGLI EQUIDI

I centromeri sono specializzate strutture nucleoproteiche dei cromosomi eucariotici essenziali per una corretta segregazione dei cromosomi durante la divisione cellulare. Nonostante l'elevata conservazione della funzione centromerica, la cromatina centromerica è caratterizzata da sequenze di DNA altamente variabili. Questo paradosso è spiegato dal fatto che la funzione centromerica è specificata epigeneticamente da CENP-A (CENTromeric protein A), la variante centromero-specifica dell'istone H3.

Mentre i centromeri dei mammiferi sono tipicamente associati a DNA altamente ripetuto in tandem (DNA satellite), il genere *Equus* (cavalli, asini e zebre) è caratterizzato dalla presenza di numerosi centromeri privi di DNA satellite ma funzionali (Wade et al., *Science*, 2009; Piras et al., *PLOS Genetics*, 2010; Nergadze et al., *Genome Research*, 2018; Cappelletti et al., *Molecular Biology and Evolution*, 2022; Piras et al., *International Journal of Molecular Science*, 2023). La presenza di questi centromeri rende il genere *Equus* un modello unico per studiare il controllo epigenetico della funzione centromerica e il suo ruolo nell'evoluzione dei cariotipi.

i. Robertsonian fusion and centromere repositioning contributed to the formation of satellite-free centromeres during the evolution of zebras

Esperimenti di ChIP-seq con un anticorpo anti-CENP-A hanno dimostrato la presenza di un numero sorprendentemente alto di centromeri senza DNA satellite nelle zebre *Equus burchelli* (15/22) e *Equus grevyi* (13/23). Questi centromeri si sono originati tramite un meccanismo di riposizionamento del centromero (movimento della funzione centromerica in un nuovo sito privo di DNA satellite senza riarrangiamento nella struttura del cromosoma) o in seguito a fusione robertsoniana. In aggiunta, abbiamo migliorato il genoma di riferimento di *E. burchelli* con un assemblaggio *de novo*.

In questo lavoro, abbiamo caratterizzato un alto numero di centromeri senza DNA satellite nelle due specie di zebre e dimostrato per la prima volta che le fusioni robertsoniane sono una fonte importante nell'origine di centromeri senza satelliti durante l'evoluzione.

Cappelletti et al. Mol Biol Evol 2022

Il mio contributo al lavoro ha riguardato l'analisi di sequenze cromosomiche tramite il programma Chromeister e il mappaggio di *read* derivanti da esperimenti di ChIP-seq con il programma Bowtie2 con specifici settaggi per l'analisi di sequenze altamente ripetute.

ii. A satellite-free centromere in Equus przewalskii chromosome 10

Analisi di ChIP-seq hanno dimostrato che il cromosoma 10 di *Equus przewalskii* (EPR), ortologo del cromosoma 11 di *Equus caballus* (ECA), è caratterizzato dall'assenza di DNA satellite. L'analisi citogenetica tramite FISH di cromosomi metafisici ha poi evidenziato un buon grado di conservazione della localizzazione delle due maggiori famiglie di DNA satellite (37cen e 2PI) tra le due specie.

In questo lavoro, abbiamo dimostrato che il centromero del cromosoma 10 di *E. przewalskii* (EPR10) si è originato in seguito a un evento di riposizionamento del centromero avvenuto nell'antenato comune tra cavallo di Przewalski e cavallo domestico. L'alta similarità della distribuzione delle famiglie di DNA satellite 37cen e 2PI ha confermato la vicinanza evolutiva tra le due specie.

Piras et al. IJMS 2023

Il mio contributo ha riguardato il mappaggio di *read* di ChIP-seq con il programma Bowtie2.

iii. *The localization of centromere protein A is conserved among tissues*

Nel genere *Equus*, il dominio di legame di CENP-A può muoversi in una finestra genomica di circa 500 kb (“*centromere sliding*”) dando origine a epialleli per il legame di CENP-A. Questi alleli posizionali sono ereditati come caratteri mendeliani. Esperimenti di ChIP-seq condotti con un anticorpo anti-CENP-A in cinque tessuti di diversa origine embrionale derivanti da quattro individui hanno dimostrato che la posizione del dominio di legame di CENP-A è conservata nei vari tessuti. Questi dati suggeriscono che la posizione del centromero sia conservata durante lo sviluppo embrionale e che il fenomeno del *centromere sliding* avvenga durante la meiosi.

Cappelletti et al. *Commun Biol* 2023

Il mio contributo ha riguardato il mappaggio di *read* di ChIP-seq con il programma Bowtie2.

F. PUBBLICAZIONI

Lavori in extenso su riviste internazionali con revisori:

1. **Sola L**, Nergadze SG, Cappelletti E, Piras FM, Giulotto E, Santagostino M. Telomeric-like repeats flanked by sequences retrotranscribed from the telomerase RNA inserted at DNA double-strand break sites during vertebrate genome evolution. *International Journal Molecular Sciences*, 2021. doi: 10.3390/ijms222011048.
2. Cappelletti E, Piras FM, **Sola L**, Santagostino M, Abdelgadir WA, Raimondi E, Nergadze SG, Giulotto E. Robertsonian fusion and centromere repositioning contributed to the formation of satellite-free centromeres during the evolution of zebras. *Molecular Biology and Evolution*, 2022. doi: 10.1093/molbev/msac162.
3. Piras FM, Cappelletti E, Abdelgadir WA, Salamon G, Vignati S, Santagostino M, **Sola L**, Nergadze SG, Giulotto E. A satellite-free centromere in *Equus przewalskii* chromosome 10. *International Journal of Molecular Sciences*, 2023. doi: 10.3390/ijms24044134.
4. Cappelletti E, Piras FM, **Sola L**, Santagostino M, Petersen JL, Bellone RR, Finno CJ, Peng S, Kalbfleisch TS, Bailey E, Nergadze SG, Giulotto E. The localization of centromere protein A is conserved among tissues. *Communications biology*, 2023. doi: 10.1038/s42003-023-05335-7.

Partecipazione a congressi

Sola L, Nergadze SG, Cappelletti E, Piras FM, Giulotto E, Santagostino M. Insertion of reverse transcribed telomerase RNA and telomeric repeats at DNA double-strand break sites during evolution. Congresso dell'Associazione Genetica Italiana AGI2021 edizione online – 22-24 Settembre 2021. Presentazione poster.

Sola L, Nergadze SG, Cappelletti E, Piras FM, Santagostino M, Giulotto E. Insertion of Telomeric Repeats and TERC retrotranscripts at DNA double-strand break sites during evolution. CSHL 2021, Congresso Telomeres & Telomerase (Online) – 14-17 Dicembre 2021. Presentazione poster.

Contributo a lavori presentati ad altri congressi

Cappelletti E, Piras FM, Annalisa Roberti, Calleda FR, Marchi G, **Sola L**, Raimondi E, Nergadze SG, Giulotto E. Dynamics and evolution of mammalian centromeres. 3° Simposio Annuale del Dipartimento di Biologia e Biotecnologia, Medicina Molecolare e Istituto di Genetica Molecolare CNR – Pavia, Italia, 19-21 Febbraio 2020.

Cappelletti E, Piras FM, Hijaz R, **Sola L**, Petersen JL, Bellone RR, Finno CJ, Kalbfleisch TS, Bailey E, Nergadze SG, Giulotto E. Epigenetic characterization of horse centromeric domains in different tissues and individuals. ISAG 38° Conferenza online della Società Internazionale di Genetica Animale (2021) – 26-30 Luglio 2021.

Cappelletti E, Piras FM, **Sola L**, Ahmed WAA, Valdata A, Raimondi E, Nergadze SG, Giulotto E. Satellite-less centromere formation by centric fusion in zebras. Congresso dell'Associazione Genetica Italiana AGI2021 Edizione online – 22-24 Settembre 2021.

Cappelletti E, Piras FM, **Sola L**, Ahmed WAA, Petersen JL, Bellone RR, Finno CJ, Kalbfleisch TS, Nergadze SG, Giulotto E. Epigenetic characterization of horse centromeric domains in different tissues and individuals from the FAANG consortium. FAANG Workshop 2022 edizione online – 16 Febbraio 2022.

Cappelletti E, Piras FM, **Sola L**, Peng S, Barber A, Santagostino M, Petersen JL, Bellone RR, Finno CJ, Kalbfleisch TS, Ernest B, Nergadze SG, Giulotto E. The epigenetic landscape of the satellite-free centromere of horse chromosome 11. ISAG 39° Conferenza della Società Internazionale per la Genetica Animale, Cape Town, Sud Africa. 2-7 Luglio 2023.

Biundo M, Piras FM, Bruni E, **Sola L**, Nergadze SG, Giulotto E, Cappelletti E. Satellite-free centromere formation by centromere repositioning in Malayan tapir. Meeting unito AGI-SIMAG 2023, Cortona, Italia. 14-16 Settembre 2023.

G. COMPETENZE DI BIOINFORMATICHE

i. Ottima conoscenza di diversi programmi bioinformatici (STAR, Bowtie2, Salmon, Samtools, IGV, deepTools, MACS2) per l'analisi di dati di Next Generation Sequencing (ChIP-seq, RNA-seq e Nanopore). Ottima conoscenza di programmi per l'analisi di qualità di *read* NGS (FastQC, TrimGalore! e MultiQC) e per l'analisi di sequenze di DNA e RNA (ChromEster, MultiAlin, RepeatMasker) (Cappelletti et al., 2022; Cappelletti et al., 2023; Sola L., tesi di Ph.D.)

ii. Ottima conoscenza di banche dati (NCBI, UCSC e altri), portali (ENCODE, FAANG) e piattaforme biologiche (Galaxy). Buona conoscenza del sistema Nextflow. (Cappelletti et al., 2022; Cappelletti et al., 2023; Sola et al., 2021; Sola L., tesi di P.hD.)

iii. Ottima conoscenza del sistema operativo Linux, in particolare della distribuzione Ubuntu. Buona conoscenza di diversi linguaggi di programmazione inclusi Bash e R. Ottima conoscenza del cluster HPC (Slurm scheduler) e buona conoscenza del cloud computing (Google Cloud)

F. COMPETENZE DI BIOLOGIA CELLULARE E MOLECOLARE

i. Culture di fibroblasti umani e di cellule tumorali umane (Sola L., tesi di P.hD.)

ii. Manipolazione di colture batteriche; digestione con enzimi di restrizione; western blotting; immunoprecipitazione della cromatina (ChIP); ibridazione *in-situ* fluorescente (PNA FISH); estrazione, purificazione e quantificazione di DNA e RNA (Sola L., tesi di P.hD.); PCR e qRT-PCR (Sola L., tesi di P.hD.)

I. LINGUE

Italiano: madre lingua

Inglese: professionale

Data 30 Maggio 2024