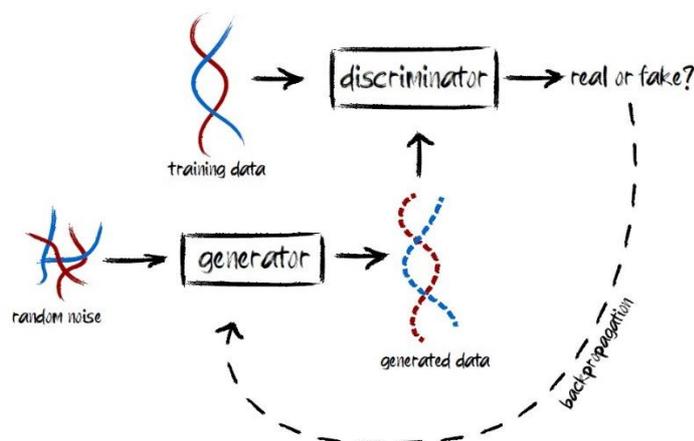


Padova, 5 febbraio 2021

## GENOMI SINTETICI PER UMANI IMMAGINARI

**Pubblicata su «PLOS Genetics» la [ricerca](#) in cui si spiega come possano essere creati al computer frammenti di genomi artificiali con caratteristiche umane partendo da bio-banche di patrimonio genetico. Non appartengono a nessun “donatore” reale, sono realistici e possono essere utilizzati come campioni virtuali di studio per tutti quei genomi reali non pubblicamente disponibili. Del team internazionale di ricercatori fa parte il Professor Luca Pagni docente di Computational Anthropology al Dipartimento di Biologia dell'Università di Padova.**

L'intelligenza artificiale, grazie ai nuovi algoritmi, è riuscita negli ultimi anni a replicare modelli complessi, mutuati dal mondo reale, e persino a generare “dati sintetici” di alta qualità come immagini realistiche di esseri umani immaginari (p.e. <https://thisxdoesnotexist.com/>). Se poi si potessero applicare le stesse tecniche dell'intelligenza artificiale anche alla biologia allora si potrebbe assegnare a questi esseri artificiali anche un patrimonio genetico. L'intento non è quello di costruire in laboratorio, in questo caso al computer, un superuomo, ma al contrario di poter far avanzare la ricerca biomedica fornendo una piattaforma di dati genomici attualmente non disponibile o accessibile.



*Figura 1 Una macchina generatrice (Generator) dà forma al rumore di fondo (random noise) mentre una macchina discriminatrice (Discriminator) valuta il prodotto, confrontandolo con i dati reali. A completamento del processo, l'algoritmo genererà dati artificiali che hanno le caratteristiche dei dati reali ma che sono in realtà completamente nuovi. Credits: Yelmen et al. 2021.*

Il team internazionale di ricercatori, che comprende le Università di Tartu, Parigi e Padova e di cui fa parte il **Professor Luca Pagni del Dipartimento di Biologia dell'Università di Padova**, è riuscita a generare al computer, partendo da un database genomico esistente, frammenti di genomi artificiali con caratteristiche reali.

«I database genomici esistenti - sostiene **Burak Yelmen**, primo autore dello [studio](#) e Junior Research Fellow di Modern Population Genetics all'Università estone di Tartu - sono una risorsa inestimabile per la ricerca biomedica, ma spesso non sono accessibili pubblicamente o sono protetti, anche giustamente, da estenuanti procedure

di accesso. Queste limitazioni creano una barriera e un disincentivo per il lavoro di ricerca. I genomi generati dall'intelligenza artificiale, o genomi artificiali come li chiamiamo, possono invece aiutarci a superare questi ostacoli al progresso della scienza all'interno di una cornice etica sicura».

I ricercatori hanno utilizzato due approcci principali per generare i genomi artificiali. Per prima cosa hanno addestrato con dati reali ricavati da un database genomico una rete generativa antagonista (GAN). Quest'ultima lavora in modo tale che, dato un set di addestramento, “impara” a generare nuovi dati con le stesse statistiche del set di addestramento. Un esempio corrente di uso di GAN è la “costruzione” di nuove fotografie che appaiono superficialmente autentiche agli osservatori umani perché incorporano molte caratteristiche (reali) del database di foto da cui prendono elementi. Poi hanno usato una macchina di Boltzmann ristretta (RBM), quest'ultima è un modello grafico probabilistico, comprensivo di un certo numero di parametri, che, se applicato ad una distribuzione di dati, è in grado di fornirne una

rappresentazione. Il team pluridisciplinare ha eseguito svariate analisi per confrontare le caratteristiche dei genomi artificiali con quelle dei genomi reali.

«Per quanto possa sembrare sorprendente - **dice Luca Pagani**, uno degli autori senior dello studio e docente all'Università di Padova - questi genomi artificiali, che emergono da pacchetti di dati inizialmente presi a caso e poi modellati, imitano le complessità che possiamo osservare all'interno di popolazioni umane reali e, per la maggior parte delle proprietà, non sono distinguibili dagli altri genomi della biobanca che abbiamo usato per addestrare il nostro algoritmo, tranne che per un dettaglio: non appartengono a nessun donatore umano».

Un problema non secondario che lo studio ha voluto verificare è quello inerente la tutela dei dati personali e sensibili contenuti nel database di partenza e in qualche modo assimilati artificialmente nel genoma artificiale prodotto: la somiglianza di genomi artificiali ha compromesso la privacy del soggetto

a cui apparteneva il genoma reale? La questione è particolarmente articolata e complessa anche perché non formalizzata nel dettaglio: la pubblicazione volontaria da parte di un singolo del suo genoma, lede la privacy di un fratello?

«Anche se rilevare problemi di privacy tra migliaia di genomi potrebbe sembrare la ricerca di un ago in un pagliaio - **conclude Flora Jay**, coordinatrice dello studio e ricercatrice del CNRS nel laboratorio interdisciplinare di informatica LRI / LISN dell'Université Paris-Saclay, Centro nazionale per la ricerca scientifica - la combinazione di più misure statistiche ci ha permesso di ovviare il più possibile a questo importante problema. Pensiamo che questo nostro sforzo possa portare miglioramenti nella valutazione e progettazione del modello generativo e alimenterà il campo dell'apprendimento automatico».

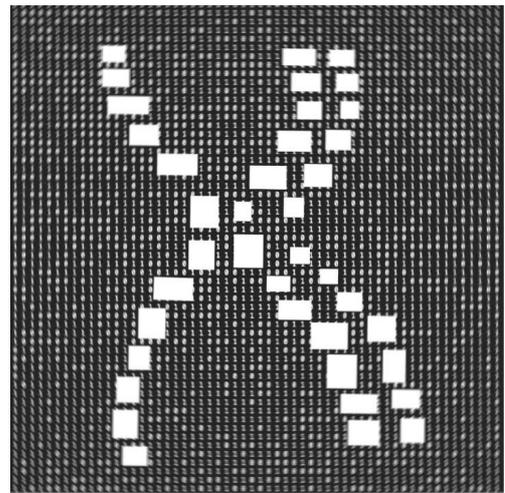


Figura 2 Rappresentazione di un cromosoma che emerge dal rumore di fondo digitale (Credits: Burak Yelmen)

Link all'articolo: <http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1009303>

Titolo: "Creating Artificial Human Genomes Using Generative Neural Networks" - «PLOS Genetics» - 2021 -

Autori: Burak Yelmen, Aurélien Decelle, Linda Ongaro, Davide Marnetto, Corentin Tallec, Francesco Montinaro, Cyril Furtlehner, Luca Pagani, Flora Jay