

## **Il mondo sorprendente del genoma di SARS-CoV-2**

Barbara Illi

Istituto di Biologia e Patologia Molecolari, Consiglio Nazionale delle Ricerche (IBPM-CNR), c/o Dipartimento di Biologia e Biotecnologie “Charles Darwin”, Sapienza Università di Roma.

*Note: evidenziate in corsivo sono riportate informazioni per “non biologi”.*

### **Introduzione**

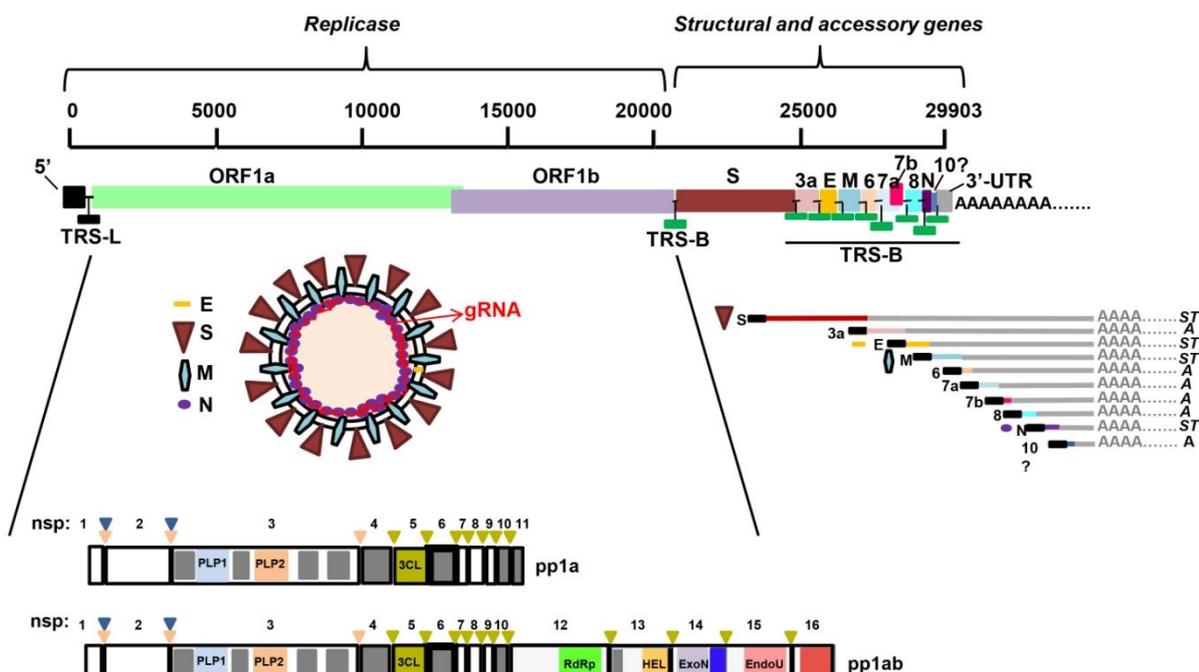
Il sopraggiungere dell'epidemia del nuovo Coronavirus (nominato prima “novel Coronavirus 2019 o 2019-nCoV, ora SARS-CoV-2, da Severe Acute Respiratory Syndrome Coronavirus 2) nel tardo Dicembre del 2019 rappresenta ancora una grande sfida per la comunità scientifica. Si sta compiendo uno sforzo enorme per raccogliere il maggior numero di informazioni nel minor tempo possibile ed, infatti, da Gennaio 2020 ad oggi, sono stati pubblicati circa 10000 articoli scientifici sull'argomento. Altre epidemie dovute a Coronavirus sono occorse negli ultimi 17 anni, l'epidemia di SARS e la “Middle East Respiratory Syndrome” (MERS), nel 2003 e nel 2012, rispettivamente. Tuttavia, queste due epidemie sono state così efficientemente controllate che i vaccini prodotti per limitarne la diffusione si sono rivelati inutili. SARS-CoV-2 (d'ora in poi semplicemente CoV-2) è per l'80% identico a suo fratello SARS-CoV. Questi virus condividono anche molte caratteristiche biologiche, incluso il meccanismo e le proteine che usano per entrare nella cellula ospite. Il meccanismo di replicazione e trascrizione del genoma di SARS-CoV-2 sembra comune a quello di altri Coronavirus. Allora, perché questo virus ha manifestazioni epidemiologiche e cliniche così diverse dagli altri Coronavirus? Per rispondere a questa domanda, dobbiamo fare un passo indietro al meccanismo di replicazione e trascrizione dei Coronavirus.

### **La sintesi dell'RNA nei Coronavirus**

I Coronavirus posseggono il genoma più grande - circa 30 kilobasi (KB) – di tutti i virus a RNA. *Come il DNA, l'RNA è costituito da “mattoni” chiamati nucleotidi (nt). Sia nel DNA che nell'RNA, l'informazione genetica è contenuta in una sequenza di “lettere”, chiamate basi, che sono uno dei componenti dei nucleotidi, e che per l'RNA sono: adenina (A), guanina (G), citosina (C) e uracile (U). Combinazioni differenti di queste 4 lettere, più o meno lunghe, costituiscono istruzioni diverse per la sintesi di tutti i prodotti proteici. Ciò è*

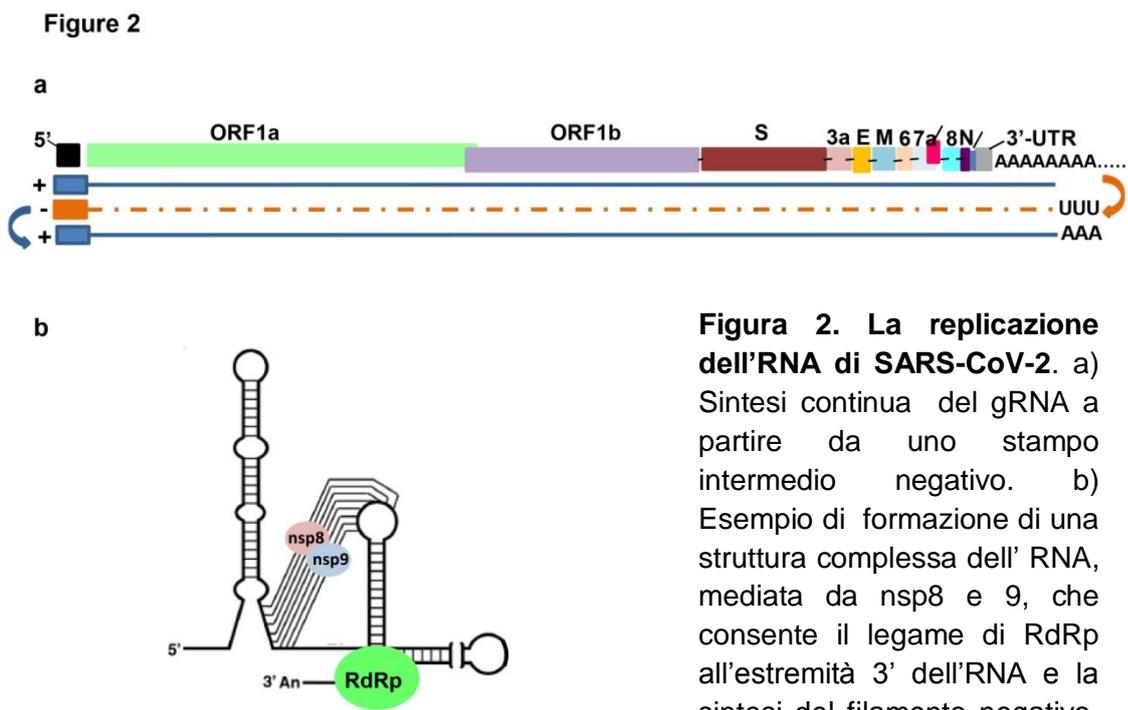
vero sia per gli organismi superiori, che per i microrganismi, come batteri e virus. Una volta che il Coronavirus è entrato nella cellula, il suo RNA viene immediatamente tradotto – dai ribosomi e da proteine specifiche della cellula ospite – in una poliproteina gigante, detta pp1ab, codificata dal gene Replicasi (Replicase), a partire da due siti chiamati “cornici aperte di lettura” (Open Reading Frames, ORFs), che sono regioni del genoma capaci di essere tradotte in proteine. Pp1ab è tagliato in 16 proteine più piccole non-strutturali (non structural proteins, nps). Rilevante è il fatto che 2 delle proteine prodotte precocemente dall’RNA virale – nps3 e 5 – servono al taglio di pp1ab. Tutte e 16 le nps servono alla replicazione dell’RNA genomico (gRNA) e alla produzione dei differenti RNA messaggeri (mRNA) sub-genomici virali (sgRNA). A valle del gene replicasi, è contenuta l’informazione per la produzione delle proteine strutturali – Spike, Envelope (proteina del rivestimento del virione), M (proteina di membrana), N (proteina del nucleocapside, che complessa l’RNA virale), che “impacchettano” il genoma virale e sono necessarie alla produzione di nuovi virioni - e di 6 proteine accessorie (3a, 6, 7a, 7b, 8, 10) il cui ruolo non è ancora ben noto<sup>1,2</sup> (figura 1). L’RNA dei Coronavirus è, di fatto, un lungo mRNA.

Figure 1



**Figure 1. Struttura del genoma di SARS-CoV-2.** La maggior parte del genoma è occupato dal gene Replicasi, che codifica per 16 nps, prodotte dal taglio della poliproteina pp1ab. A valle, è presente l’informazione per le proteine strutturali e accessorie. Abbreviazioni: L= leader; TRS-L=transcription regulatory sequence at the Leader (sequenza regolatoria della trascrizione al Leader); TRS-B=transcription regulatory sequence at the body (sequenza regolatoria della trascrizione all’interno del corpo del genoma); ORF=open reading frame (cornice di lettura aperta); S=spike; E=envelope (proteina del rivestimento virale); M=membrane (proteina della membrana); N=nucleocapsid (proteina del nucleocapside che complessa il gRNA). 3a, 6, 7a, 7b, 8, 10=geni accessorie. Nsp=non structural protein; PLP=papain-like protease (proteasi simile alla papaina); 3CL= chymotrypsin-like protease (proteasi simile alla chimo tripsina); RdRp=RNA-dependent RNA polymerase (RNA polimerasi RNA dipendente); HEL=helicase (elicasi); ExoN=exonuclease (esonucleasi); EndoU=endonuclease (endonucleasi); MTase=methyltransferase (metiltransferasi); UTR=untranslated region (regione non tradotta). ST= proteina strutturale; A=proteina accessoria. (Adattato da Kim et al., Cell, 2020 and Sola et al., Ann Rev Virol, 2015)

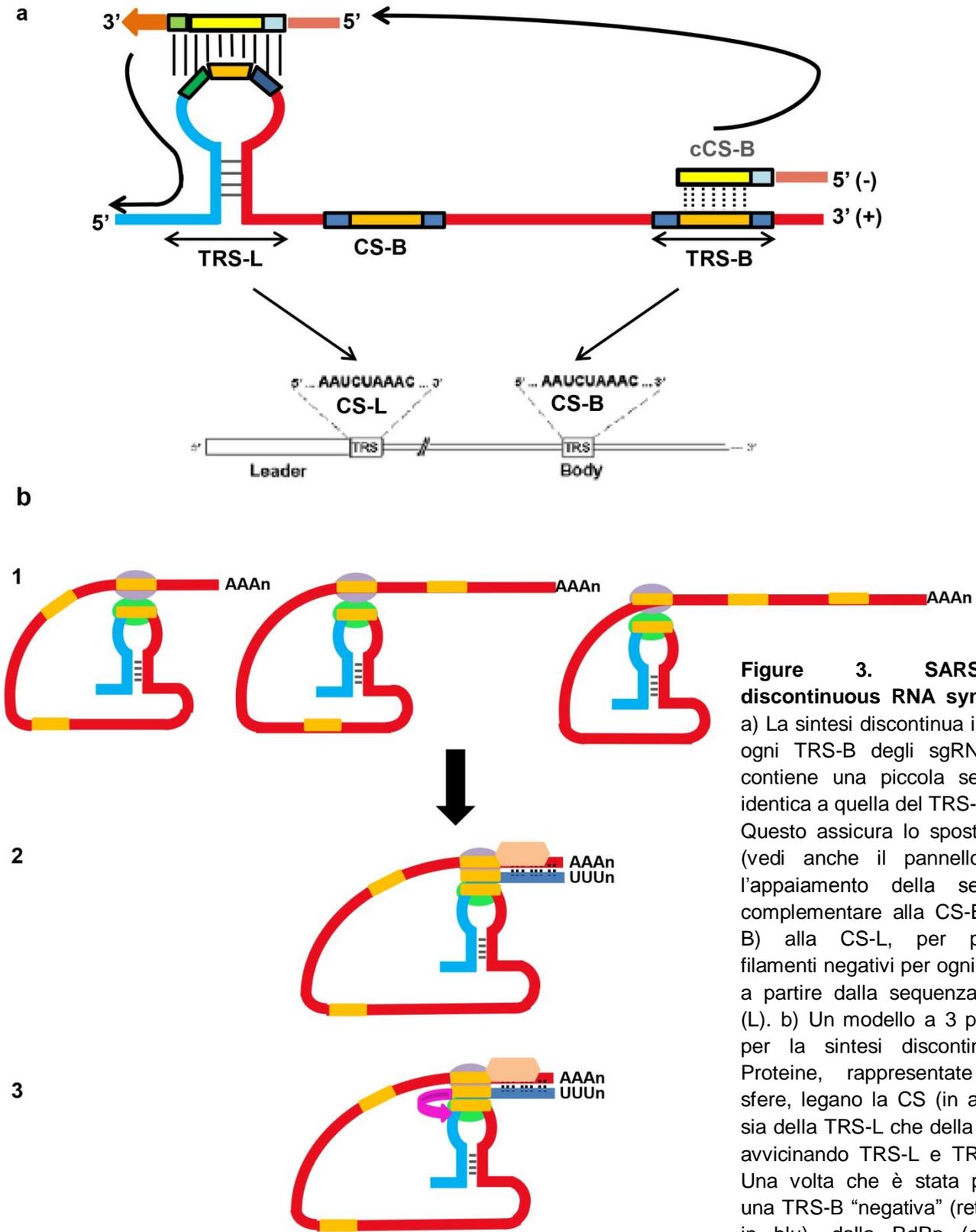
Perciò è un filamento singolo “positivo”, ossia segue una direzione cosiddetta “5’→3” (figura 2a). La replicazione dell’RNA, ossia la produzione di più copie del genoma virale, è un processo continuo. Il nuovo gRNA viene prodotto grazie alla sintesi di un filamento intermedio “negativo”, che serve da stampo per il nuovo gRNA positivo (figura 2a). Questo processo coinvolge principalmente nsp12, che ha attività RNA polimerasi dipendente dall’RNA (RdRp), e che sintetizza il primo filamento negativo legandosi all’estremità 3’ del gRNA, grazie anche alla formazione di strutture complesse che “flettono” l’RNA in modo da facilitarne la replicazione (figura 2b).



**Figura 2. La replicazione dell’RNA di SARS-CoV-2.** a) Sintesi continua del gRNA a partire da uno stampo intermedio negativo. b) Esempio di formazione di una struttura complessa dell’ RNA, mediata da nsp8 e 9, che consente il legame di RdRp all’estremità 3’ dell’RNA e la sintesi del filamento negativo. (Adattato da *Kim et al., Cell, 2020* and *Sola et al, RNA Biol, 2011*)

Al contrario, la sintesi degli sgRNA è un processo discontinuo. Ogni sgRNA ha, all’estremità 5’, una sequenza di 70 nt chiamata Leader (L), che è presente una volta sola all’estremità 5’ dell’intero gRNA. La sintesi discontinua degli sgRNA dipende da sequenze, chiamate sequenze regolatrici della trascrizione (TRS), presenti a valle del Leader (TRS-L) e che precedono ogni sgRNA (TRS-B) (figura 1). Le TRS sono caratterizzate da sequenze conservate (CS, conserved sequences) di 6-7 nt, identiche per il TRS-L e ogni TRS-B e che permettono l’appaiamento della sequenza CS-L, con la sequenza complementare CS-B (cCS-B, complimentary CS-B) dell’ RNA intermedio nascente negativo, quando l’RNA si

Figure 3



**Figure 3. SARS-CoV-2 discontinuous RNA synthesis.**

a) La sintesi discontinua inizia ad ogni TRS-B degli sgRNA, che contiene una piccola sequenza identica a quella del TRS-L. (CS). Questo assicura lo spostamento (vedi anche il pannello b) e l'appaiamento della sequenza complementare alla CS-B (cCS-B) alla CS-L, per produrre filamenti negativi per ogni sgRNA a partire dalla sequenza leader (L). b) Un modello a 3 passaggi per la sintesi discontinua. 1. Proteine, rappresentate come sfere, legano la CS (in arancio), sia della TRS-L che della TRS-B, avvicinando TRS-L e TRS-B. 2. Una volta che è stata prodotta una TRS-B "negativa" (rettangolo in blu), dalla RdRp (esagono rosa), questa si sposta alla TRS-L, aggiungendo una copia "negativa" della stessa TRS-L e completando il filamento negativo dell' sgRNA, che serve da stampo per quello "positivo", codificante. (Adattato da Sola et al., *Ann Rev Virol*, 2015)

piega per allineare la TRS-L con ogni TRS-B<sup>1,2</sup> (figura 3a). Questo processo assicura la produzione di ciascun sgRNA, che è, sostanzialmente, una fusione tra la sequenza Leader all'estremità 5' dell'RNA virale e la sequenza dell'sgRNA preceduto dalla TRS-B e che richiede interazioni a distanza RNA-RNA, mediate da complessi RNA-proteina. Un esempio di questo meccanismo è mostrato in figura 3b.

### **L'RNA polimerasi dipendente dall'RNA: Jumping Jack Flash**

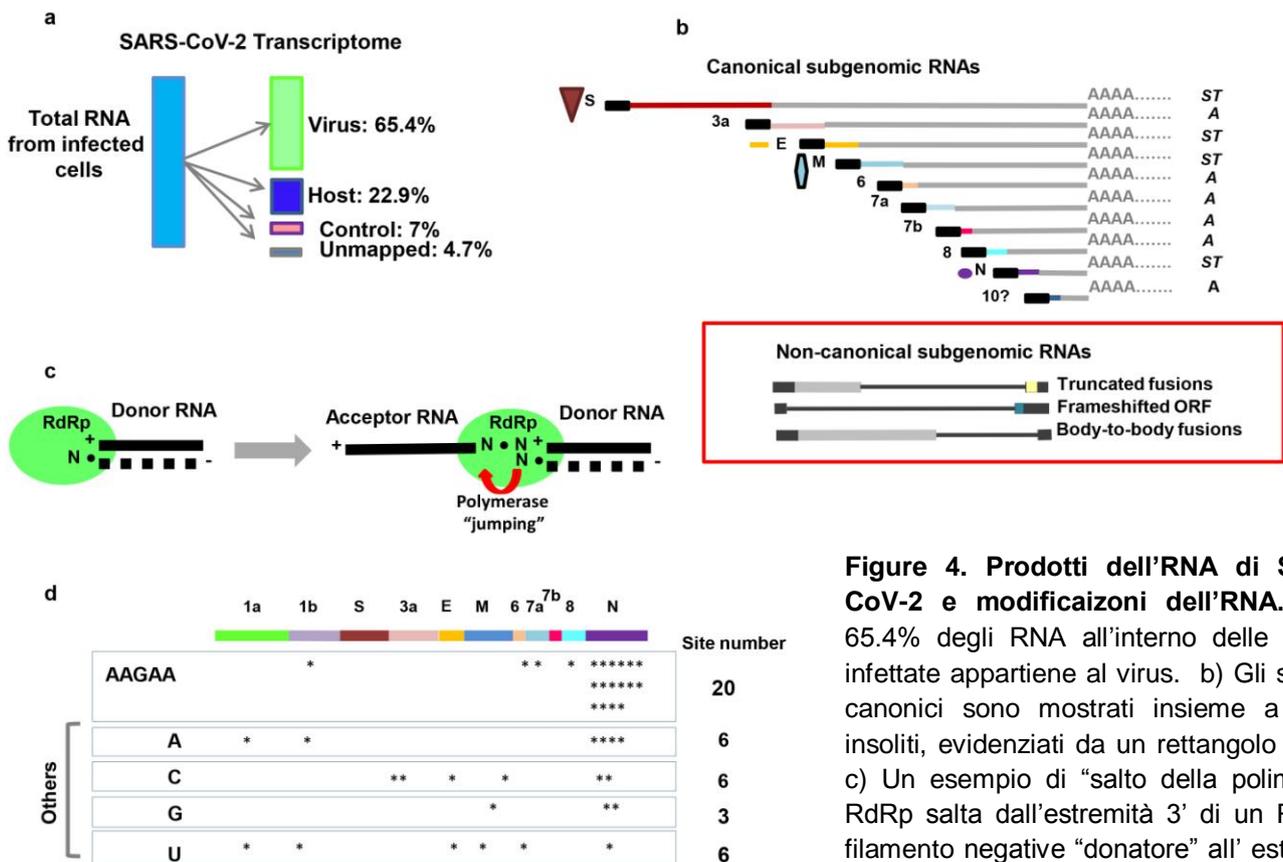
La sintesi dell'RNA di CoV2 è, sostanzialmente, identica a quella degli altri CoV umani. Tuttavia, le caratteristiche di alcuni dei suoi sgRNA suggeriscono meccanismi di sintesi alternativi. L'intero pannello degli mRNA di CoV-2 (il cosiddetto "trascrittoma") è espresso più del patrimonio degli mRNA della cellula ospite, come se, in qualche modo, l'espressione dei geni cellulari venisse spenta (figura 4a)<sup>4</sup>. Il 92% degli sgRNA di CoV-2 è prodotto secondo il meccanismo descritto nel paragrafo precedente e questi sono principalmente rappresentati dagli mRNA per N, S, M, E, 3a, 6, 7a, 7b e 8 (figura 4b), ma un certo numero di fusioni hanno struttura insolita (figura 4b, rettangolo rosso). La presenza di alcune sequenze di 3-4 nt, identiche all'estremità 5' e 3' di queste fusioni, potrebbe far ipotizzare un meccanismo di "salto dell'RNA polimerasi" (polymerase jumping), già descritto in altri sistemi (figura 4c). Questo produce un certo numero di sgRNA a funzione ignota nel ciclo vitale del virus, ma con la capacità potenziale di essere tradotti in proteine, agli stessi livelli delle proteine accessorie. Perciò, sarà importante verificare se queste proteine effettivamente esistano e che ruolo possano avere.

### **Ma le cose sono più complicate di così: le modificazioni dell'RNA.**

*L'Epigenetica è quella branca della Scienza che studia le risposte del genoma agli stimoli intracellulari, ambientali, meccanici ecc. Le modificazioni epigenetiche sono i "marchi" di queste risposte e, in sostanza, non sono altro che modifiche chimiche che possono essere apportate ai genomi e anche alle proteine. Le cosiddette "macchine epigenetiche" sono responsabili di queste modificazioni e sono complessi multiproteici attivati dagli stimoli sopra menzionati. Le modificazioni epigenetiche cambiano le caratteristiche di un genoma, ad esempio dando istruzioni su quali geni devono essere accesi o spenti, in modo che la cellula possa rapidamente rispondere a segnali interni o esterni. Modificazioni epigenetiche ben note sono quelle a carico del DNA e degli istoni associati, che nell'insieme formano la cromatina, quindi, i cromosomi, all'interno della cellula. Anche l'RNA può essere modificato ed, in effetti, modificazioni dell'RNA virale sono state già*

osservate in altri virus. Queste possono essere rappresentate dall'aggiunta di gruppi metilici (su A o C), dalla perdita di gruppi amminici dalle basi, o aggiunta di nucleotidi (ad esempio di uracile). Sono state identificate modificazioni dell'RNA in SARS-CoV-2 (figura 4d), ma la loro natura non è stata ancora identificata, né tantomeno il loro ruolo. Tuttavia, si pensa possano avere un ruolo nel controllo della stabilità dei vari sgRNA o potrebbero contribuire ad eludere il sistema immunitario dell'ospite.

Figure 4



**Figure 4. Prodotti dell'RNA di SARS-CoV-2 e modificazioni dell'RNA.** a) Il 65.4% degli RNA all'interno delle cellule infettate appartiene al virus. b) Gli sgRNA canonici sono mostrati insieme a quelli insoliti, evidenziati da un rettangolo rosso. c) Un esempio di "salto della polimerasi". RdRp salta dall'estremità 3' di un RNA a filamento negativo "donatore" all'estremità 3' di un RNA accettore, appaiando piccole sequenze complementari. d) Gli asterischi corrispondono alle modificazioni dell'RNA. Il numero più alto si verifica a carico di sequenze AAGAA. Il numero di modificazioni è mostrato a destra. (Adattato da Kim et al., Cell, 2020).

## Conclusioni

Le caratteristiche biologiche di SARS-CoV-2 sono, in larga parte, sovrapponibili a quelle degli altri CoV. Nonostante ciò, il meccanismo di sintesi del suo RNA è molto complesso e, in alcuni casi, sfugge ai meccanismi canonici. Specificamente, l'alta frequenza di fusioni

insolite potrebbe produrre varianti virali, un punto che merita particolare attenzione se pensiamo ai meccanismi di farmaco-resistenza ed elusione del sistema immunitario che CoV-2 potrebbe adottare per sopravvivere e propagarsi. Le modificazioni dell'RNA potrebbero, anch'esse, svolgere un ruolo in questi ambiti. Queste caratteristiche andrebbero studiate in tessuti animali, dove è attiva una risposta immunitaria. Inoltre, sarebbe interessante valutare se simili modificazioni dell'RNA si riscontrassero in altri CoV, per comprendere meglio le proprietà di SARS-CoV-2.

## Referenze

1. Snijder EJ, Decroly E, Ziebuhr J. The nonstructural proteins directing Coronavirus RNA synthesis and processing. *Adv Virus Res.* 2016;96:59-126. doi:10.1016/bs.avir.2016.08.008.
2. Sola I, Almazán F, Zúñiga S, Enjuanes L. Continuous and Discontinuous RNA Synthesis in Coronaviruses *Annu Rev Virol.* 2015;2: 265-88. doi: 10.1146/annurev-virology-100114-055218.
3. Masters PS. The molecular biology of Coronaviruses. *Adv Virus Res.* 2006;66:193-292.
4. Kim D, Lee JY, Yang JS, Kim JW, Kim VN, Chang H. The Architecture of SARS-CoV-2 transcriptome. *Cell.* 2020 Apr 18. pii: S0092-8674(20)30406-2. doi: 10.1016/j.cell.2020.04.011.