A blockchain model of cancer evolution

Alfred C. Chin
Weill Cornell/Rockefeller/Sloan Kettering Tri-Institutional MD-PhD Program, New York, NY
Correspondence: acc4002@med.cornell.edu

Blockchain technology has taken the world by storm in recent years. Originally developed by Satoshi Nakamoto for the cryptocurrency Bitcoin, blockchain has been adapted for diverse applications such as enhancing food traceability, securing electronic health records, and other data management problems. However, blockchain has never been applied to biological mechanisms. Presented here is a conceptual framework for probing cancer evolution using blockchain principles, the first example of "blockchain biology."

Blockchains are decentralized, append-only ledgers. Instead of a centralized entity, for example a bank, controlling an entire ledger, multiple parties (nodes) form a network to maintain a synchronized, distributed, and identical record. Decentralization safeguards the integrity of the ledger when individual nodes are lost. The ledger is comprised of blocks that store data, such as the details of a financial transaction, and are linked chronologically to create a metaphorical chain of blocks. The append-only design of blockchain guarantees a complete, traceable, and virtually tamper-proof ledger.

Clonal evolution in cancer exhibits strikingly similar features as blockchain (Figure 1). Accruing genetic and epigenetic alterations in a stepwise, sequential manner, cancer cell clones are subject to Darwinian natural selection throughout their growth. Clonal architectures involve a founder mutation, for example *ETV6-RUNX1* fusion in acute lymphoblastic leukemia, that drives clonal expansion and subsequent diversification (1). Defining the ledger as the complete history of a cancer, this critical origin can be represented by the genesis block of a blockchain. The dataset in each block harbors a snapshot of the cancer state in time, ideally the entire single-cell omics signature. Accordingly, appending a new block to the ledger corresponds to adding an updated snapshot of the cancer state to the cancer history. Appending new blocks is critical because cancer cells are constantly subjected to dynamic evolutionary pressures, including resource competition, microenvironmental constraint, and therapeutic intervention (2). Decentralization can be achieved by treating every cell as an individual node and connections as intercellular relationships. Such connections are realistic given the significant computational advances in characterizing cell-cell communication (3). Reconstructing the ledger necessitates integration of the intrinsic omics of a single cell and all its intercellular relationships.

What guarantees that a newly appended block is an accurate updated snapshot of the cancer state? The cryptographic hash and proof-of-work mechanisms of a blockchain can guarantee that the evolutionary trajectory is faithfully documented. Cryptographic hash functions are one-way functions (inputs can only be determined by trial-and-error, not rationally, from outputs) that map an arbitrary dataset to a fixed value such as a string of binary digits. Each block contains the hash of the previous block and its own unique hash that is a function of both its intrinsic data and the previous hash, enabling an append-only chain. A cryptographic hash function can map a single-cell omics signature to a dimension-reduced fingerprint of the cancer. Such processing is realistic given the substantial progress made in computational methods for multimodal integration of single-cell omics data (4). The linear organization of blocks ensures that changes during the inter-block timeframe in any arbitrary feature of the cancer, say flux through a signaling pathway in a specific cell, can be determined by comparing the contents of block '*n*+1' and block '*n*'. Proof-of-work dictates that hashes need to meet certain conditions, thus requiring brute force computations as a prerequisite for adding new blocks due to the one-way nature of cryptographic hash functions. Because adjusting the hash conditions modulates the difficulty of adding new blocks, proof-of-work establishes the inter-block timeframe and tunes the temporal resolution of the cancer history.

Reported here is an elementary framework of "blockchain biology" applied to cancer evolution. Considerable development is needed to advance this model to a functional computational tool. The most immediate application is a computational approach to retrospective lineage tracing (5). Other fields beyond cancer dynamics and developmental biology that may be amenable to interrogation include cell signaling networks and synthetic programmable circuits, particularly if "blockchain biology" is integrated with experimental approaches. Lastly, as evidenced by the diverse forms of cryptocurrencies, significant methods development is crucial because there are numerous possibilities for blockchain infrastructure.

References:

1. Greaves M, Maley CC. Clonal evolution in cancer. *Nature*. 481, 306–313 (2012).
2. Ferrando AA, López-Otín C. Clonal evolution in leukemia. *Nat Med*. 23, 1135–1145 (2017).
3. Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat Protoc*. 15, 1484–1506 (2020).
4. Efremova M, Teichmann SA. Computational methods for single-cell omics across modalities. *Nat Methods*. 17, 14–17 (2020).
5. Baron CS, van Oudenaarden A. Unravelling cellular relationships during development and regeneration using genetic lineage tracing. *Nat Rev Mol Cell Biol*. 20, 753–765 (2019).

Competing Interests: A.C.C. trades cryptocurrencies, primarily Bitcoin and Ethereum.

Funding: A.C.C. is supported by NIH Medical Scientist Training Program Training Grant T32GM007739.

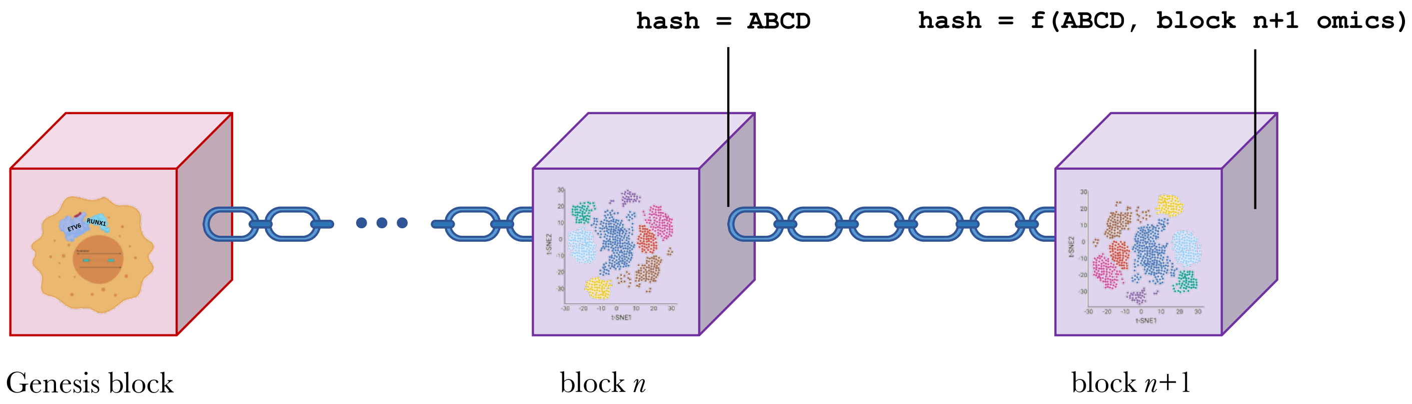

Figure 1. Schematic for blockchain model of cancer evolution. Founder mutation (*ETV6-RUNX1* fusion for ALL as an example) initiates genesis block. Contents of every block encompass the complete single-cell omics of the cancer at a certain time point. Each block is marked by a unique hash that is a function of the hash of the previous block and its own contents. Proof-of-work determines the timeframe elapsed between each block. The entire ledger is decentralized across every cancer cell.