
Neural network and kinetic modelling of human genome replication reveal replication origin locations and strengths

Jean-Michel Arbona*¹, Benjamin Audit², Olivier Hyrien³, Hadi Kabalane⁴, and Arach Goldar^{5,6}

¹Laboratoire de Biologie et Modélisation de la Cellule – Université de Lyon, ENS de Lyon, Centre National de la Recherche Scientifique - CNRS – France

²Laboratoire de Physique – Université de Lyon, ENS de Lyon, CNRS – France

³Institut de Biologie de l'École normale supérieure (IBENS), Département de Biologie, Ecole Normale Supérieure, CNRS, Inserm, PSL Research University, F – 75005 Paris – France

⁴Univ Lyon, ENS de Lyon, Univ Claude Bernard Lyon 1, CNRS, Laboratoire de Physique, F – 69342, Lyon – France

⁵Service de Biologie Intégrative et Génétique Moléculaire (SBIGeM) – CEA – CEA Saclay 91191 Gif sur Yvette cedex, France

⁶Institut de Biologie et de Technologies de Saclay (IBITECS) – CEA – CEA Saclay bat 532 91191 Gif sur Yvette cedex, France

Résumé

In human and other metazoans, the determinants of replication origin location and strength are still elusive. Origins are licensed in G1 phase and fired in S phase of the cell cycle, respectively. It is debated which of these two temporally separate steps determines origin efficiency. Experiments can independently profile mean replication timing (MRT) and replication fork directionality (RFD) genome-wide. Such profiles contain information on multiple origins' properties and on fork speed. Due to possible origin inactivation by passive replication, however, observed and intrinsic origin efficiencies can markedly differ. Thus, there is a need for methods to infer intrinsic from observed origin efficiency, which is context-dependent. Here, we show that MRT and RFD data are highly consistent with each other but contain information at different spatial scales. Using neural networks, we infer an origin licensing landscape that, when inserted in an appropriate simulation framework, jointly predicts MRT and RFD data with unprecedented precision and underlies the importance of dispersive origin firing. We furthermore uncover an analytical formula that predicts intrinsic from observed origin efficiency combined with MRT data. Comparison of inferred intrinsic origin efficiencies with experimental profiles of licensed origins (ORC, MCM) and actual initiation events (Bubble-seq, SNS-seq, OK-seq, ORM) show that intrinsic origin efficiency is not solely determined by licensing efficiency. Thus, human replication origin efficiency is set at both the origin licensing and firing steps.

Mots-Clés: DNA replication, stochastic modelling

*Intervenant