# **Introduction :**

In most organism, it is the deoxyribonucleic acid (DNA) that defines the genetic information. In order to ensure the survival of species, cells has to accurately duplicate and transmit these double-stranded molecules of DNA to progeny through the process of DNA replication. However, DNA replication is not the only challenge that a cell encounters when trying to preserve hereditary information. In fact, DNA is continuously damaged either by exogenous stressors or by endogenous processes (Ref). Despite the fact that it is important to permit a certain rate of genomic changes so the species could evolve in a long term, it is extremely important that the genome is effectively maintained and repaired for the sake of the cell’s survival in the short term. Therefore, cells do accumulate mutations but at an exceptionally low rate. In spite of that, certain mutations are not tolerated and will result in cell death while other mutations could confer a growth advantage to the cell and results in an uncontrollable growth leading eventually to cancer development (Wilson and Hunt, 2015). This type of mutation is known as ‘drivers’ while the remainders that do not confer any growth advantages are known as ‘passengers’. There are only a few ‘driver mutations’ in a cancer genome while most of the mutations are ‘passengers’. The ‘passengers’ could be regarded as the imprints of all the mutational processes operative during the lifetime of a cancer patient whereas the ‘drivers’ confer clonal expansion allowing for a better resolution to study these mutational processes (Figure 1) (Stratton et al., 2009).

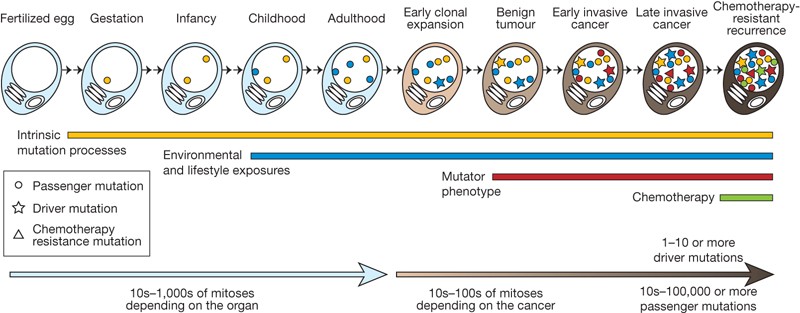


Figure 1. xxxxxx (from Stratton et al., 2019)

Owing to large scale sequencing initiatives such as The Cancer Genome Atlas -TCGA- and Pan-Cancer Analysis of Whole Genome -PCAWG- (ICGC/TCGA, 2020), thousands of cancer mutational catalogs were obtained spanning most the cancer types. Thus, offering an unique opportunity to extract the mutational patterns left by the aforementioned processes. In 2013, Alexandrov and colleagues published the first mathematical approach for deciphering these mutational patterns operating in each cancer type, this method is currently known as SigProfiler (Alexandrov et al., 2013a) and is the one used to extract the reference signatures (Alexandrov et al., 2013b, Alexandrov et al., 2020) that are available at the Catalog Of Somatic Mutations In Cancer -COSMIC- (Tate et al., 2019). Since then, several other tools have been developed for the same purpose (See Omichessan et al., 2019 for a comprehensive list of tools). The mutational patterns described here are usually referred to as mutational signatures and are represented as multinomial distributions for context-based mutation types. Basically, six mutation types are used [C>A, C>G, C>T, T>A, T>C, T>G] frequently added to them the flanking 5’ and 3’ bases, which yields 96 mutation types, also termed as channels (Alexandrov et al., 2013a). The largest analysis performed so far was also carried out by Alexandrov’s group in 2020, where they extracted 49 single-base substitution signatures (SBS), 11 doublet base substitution signatures (DBS), and 17 insertion-deletion signatures (IDS). Some of these signatures are of known etiology while many remain ambiguous (Alexandrov et al., 2020).

As previously discussed, DNA polymerases replicate the DNA faithfully under normal conditions, with about one mutation per 10¹⁰ nucleotides per cell division is estimated to occur in human cells (Wilson and Hunt, 2015). Any abnormalities could deregulate the replication program and rise mutations. In fact, a defective exonuclease or proofreading domains were shown to be associated with the characteristic mutational signatures SBS10a, SBS10b, SBS10c, and SBS10d (Alexandrov et al., 2020). Furthermore, replication stress is a phenomenon that can challenge genome stability through interference with replication fork progression. It can be originated from different events of endogenous or exogenous nature. For instance, DNA lesions such as thymine dimers that are induced by UV-light, or unusual DNA structures like hairpins, or even conflicts between replication and transcription machinery usually results in replication stress (Mazouzi et al., 2014). The replication stress response involves many pathways themselves including several components. Depletion or damaging of one of these constituents will lead to unresolved replication stress which could produce different deleterious events (Zeman and Cimprich, 2014). Such perturbations would impact the mutational catalog suggesting that components operative during replication or replication stress are potential candidates to explain other signatures.

Previous studies have identified an asymmetrical distribution of mutations along the strands. This asymmetry can be divided into transcriptional and replicative asymmetries. The former could be due to the transcription-coupled repair (TCR) mechanism that is activated when the RNA-polymerase gets stalled when facing a DNA lesion. Consequently, the non-transcribed template is more prone to accumulate mutations. This kind of asymmetry was observed to associate mainly with mutations originating from UV lights and smoking. Whereas the replicative asymmetry was observed with POLE, APOBEC, and MSI-associated mutations. As a matter of fact, the lagging strand remains single-stranded for longer periods making it more fragile, to that adds the different activities of polymerases α, δ, and ε as well as their different proofreading proprieties (Nicholas et al., 2015). In collaboration with other team at I. Curie, the host team has recently revealed that how a better understanding of APOBEC mutation signatures and the locations of APOBEC mutation hotspots can help to distinguish the passenger from driver mutations, and how APOBEC mutations induce driver mutations leading to Bladder cancer development (Refs). Furthermore, recent studies in the team have shown that R-loops (RNA:DNA hybrids) are preferentially enriched at the transcription termination sites (TTS) of highly expressed, convergent genes. Replication stress markers were also determined to be enriched at these regions suggesting a mechanism of replication-transcription conflicts (TRC) that is resolved by topoisomerases (TOP) in normal cells considering that double strand breaks form in TOP1-depleted cells (Promonet et al., 2020). Accordingly, Reijns and colleagues showed that the mutational signature ID4 referenced in COSMIC database is similar to the signature they extracted which is associated with the defective activity of TOP1 at sites where ribonucleotides (rNTPs) were mis-incorporated (Reijns et al., 2022). This falls in accordance with the fact that rNTPs embedded in the DNA sequence cause replication stress (Zeman and Cimprich, 2014).

In conclusion, there is mounting evidence suggesting that the replication program and the replication stress play a significant role in shaping the mutational landscape of a cancer genome. Hence, we hypothesize that there might be mutational signatures specific to different replication stresses. Furthermore, there could be an association between the rates at which certain mutational processes operates and replication stress occurrence. To that end, the goal of this project is to search for mutational signatures associated with replication stress. More specifically, if there are different mutational processes operating at transcription starting sites (TSS) and TTS of a given set of genes and if there is a mutational asymmetry between the two regions, and what are their association with transcription-associated replication stress and/or R-loop formation. This will provide novel insights to better understand the mechanisms associated with mutation signatures and their roles in genome evolution and cancer development.

Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J., & Stratton, M. R. (2013a). Deciphering signatures of mutational processes operative in human cancer. *Cell reports*, *3*(1), 246–259.<https://doi.org/10.1016/j.celrep.2012.12.008>

Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A., Behjati, S., Biankin, A. V., Bignell, G. R., Bolli, N., Borg, A., Børresen-Dale, A. L., Boyault, S., Burkhardt, B., Butler, A. P., Caldas, C., Davies, H. R., Desmedt, C., Eils, R., Eyfjörd, J. E., Foekens, J. A., Greaves, M., … Stratton, M. R. (2013b). Signatures of mutational processes in human cancer. *Nature*, *500*(7463), 415–421. <https://doi.org/10.1038/nature12477>

Alexandrov, L. B., Kim, J., Haradhvala, N. J., Huang, M. N., Tian Ng, A. W., Wu, Y., Boot, A., Covington, K. R., Gordenin, D. A., Bergstrom, E. N., Islam, S., Lopez-Bigas, N., Klimczak, L. J., McPherson, J. R., Morganella, S., Sabarinathan, R., Wheeler, D. A., Mustonen, V., PCAWG Mutational Signatures Working Group, Getz, G., … PCAWG Consortium (2020). The repertoire of mutational signatures in human cancer. *Nature*, *578*(7793), 94–101.<https://doi.org/10.1038/s41586-020-1943-3>

Cancer Genome Atlas Res. Netw. 2022. Genomic data commons data portal. Data Portal, Natl. Cancer Inst., Rockville, MD. <https://portal.gdc.cancer.gov/>

Haradhvala, N. J., Polak, P., Stojanov, P., Covington, K. R., Shinbrot, E., Hess, J. M., Rheinbay, E., Kim, J., Maruvka, Y. E., Braunstein, L. Z., Kamburov, A., Hanawalt, P. C., Wheeler, D. A., Koren, A., Lawrence, M. S., & Getz, G. (2016). Mutational Strand Asymmetries in Cancer Genomes Reveal Mechanisms of DNA Damage and Repair. *Cell*, *164*(3), 538–549.<https://doi.org/10.1016/j.cell.2015.12.050>

ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium (2020). Pan-cancer analysis of whole genomes. *Nature*, *578*(7793), 82–93.<https://doi.org/10.1038/s41586-020-1969-6>

Mazouzi, A., Velimezi, G., & Loizou, J. I. (2014). DNA replication stress: causes, resolution and disease. *Experimental cell research*, *329*(1), 85–93. <https://doi.org/10.1016/j.yexcr.2014.09.030>

Omichessan, H., Severi, G., & Perduca, V. (2019). Computational tools to detect signatures of mutational processes in DNA from tumours: A review and empirical comparison of performance. *PloS one*, *14*(9), e0221235.<https://doi.org/10.1371/journal.pone.0221235>

Promonet, A., Padioleau, I., Liu, Y., Sanz, L., Biernacka, A., Schmitz, A. L., Skrzypczak, M., Sarrazin, A., Mettling, C., Rowicka, M., Ginalski, K., Chedin, F., Chen, C. L., Lin, Y. L., & Pasero, P. (2020). Topoisomerase 1 prevents replication stress at R-loop-enriched transcription termination sites. *Nature communications*, *11*(1), 3940.<https://doi.org/10.1038/s41467-020-17858-2>

Reijns, M., Parry, D. A., Williams, T. C., Nadeu, F., Hindshaw, R. L., Rios Szwed, D. O., Nicholson, M. D., Carroll, P., Boyle, S., Royo, R., Cornish, A. J., Xiang, H., Ridout, K., Genomics England Research Consortium, Colorectal Cancer Domain UK 100,000 Genomes Project, Schuh, A., Aden, K., Palles, C., Campo, E., Stankovic, T., … Jackson, A. P. (2022). Publisher Correction: Signatures of TOP1 transcription-associated mutagenesis in cancer and germline. *Nature*, *605*(7910), E7. <https://doi.org/10.1038/s41586-022-04812-z>

Stratton, M. R., Campbell, P. J., & Futreal, P. A. (2009). The cancer genome. *Nature*, *458*(7239), 719–724.<https://doi.org/10.1038/nature07943>

Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., Bindal, N., Boutselakis, H., Cole, C. G., Creatore, C., Dawson, E., Fish, P., Harsha, B., Hathaway, C., Jupe, S. C., Kok, C. Y., Noble, K., Ponting, L., Ramshaw, C. C., Rye, C. E., Speedy, H. E., … Forbes, S. A. (2019). COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic acids research*, *47*(D1), D941–D947.<https://doi.org/10.1093/nar/gky1015>

Wilson, J. and Hunt, T., 2015. *Molecular biology of the cell*. New York, NY: Garland Science, pp.3, 237-287.

Zeman, M. K., & Cimprich, K. A. (2014). Causes and consequences of replication stress. *Nature cell biology*, *16*(1), 2–9.<https://doi.org/10.1038/ncb2897>