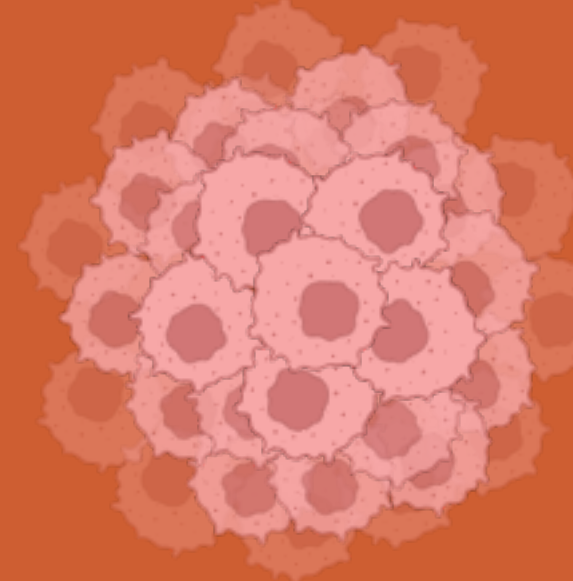


# Building Mutational Signatures in Cancer using Deep Bayesian Neural Networks



## Introduction

Cancer is characterised by uncontrolled growth of cells. To sustain this level of growth, the cancer cells have acquired several capabilities such as evading cell death and the construction of vascular infrastructure. Most of these capabilities are, primarily, acquired by mutation of the genome. These genetic variations result in unique “mutational signatures” within the DNA sequence.

## Methods

Variant Call Format files were processed to identify Single Base Substitutions within a broader nucleotide context. Non-negative Matrix Factorization was employed to decompose mutational signatures, which were then verified against the COSMIC database. Optimization of NMF parameters was crucial for accurate signature extraction. Comparative analysis of the signatures utilized cosine similarity and Jensen Shannon Distance metrics for a robust quantitative assessment

## Contact

Jan Alfonso Busker  
j.a.busker@st.hanze.nl

## RESULTS

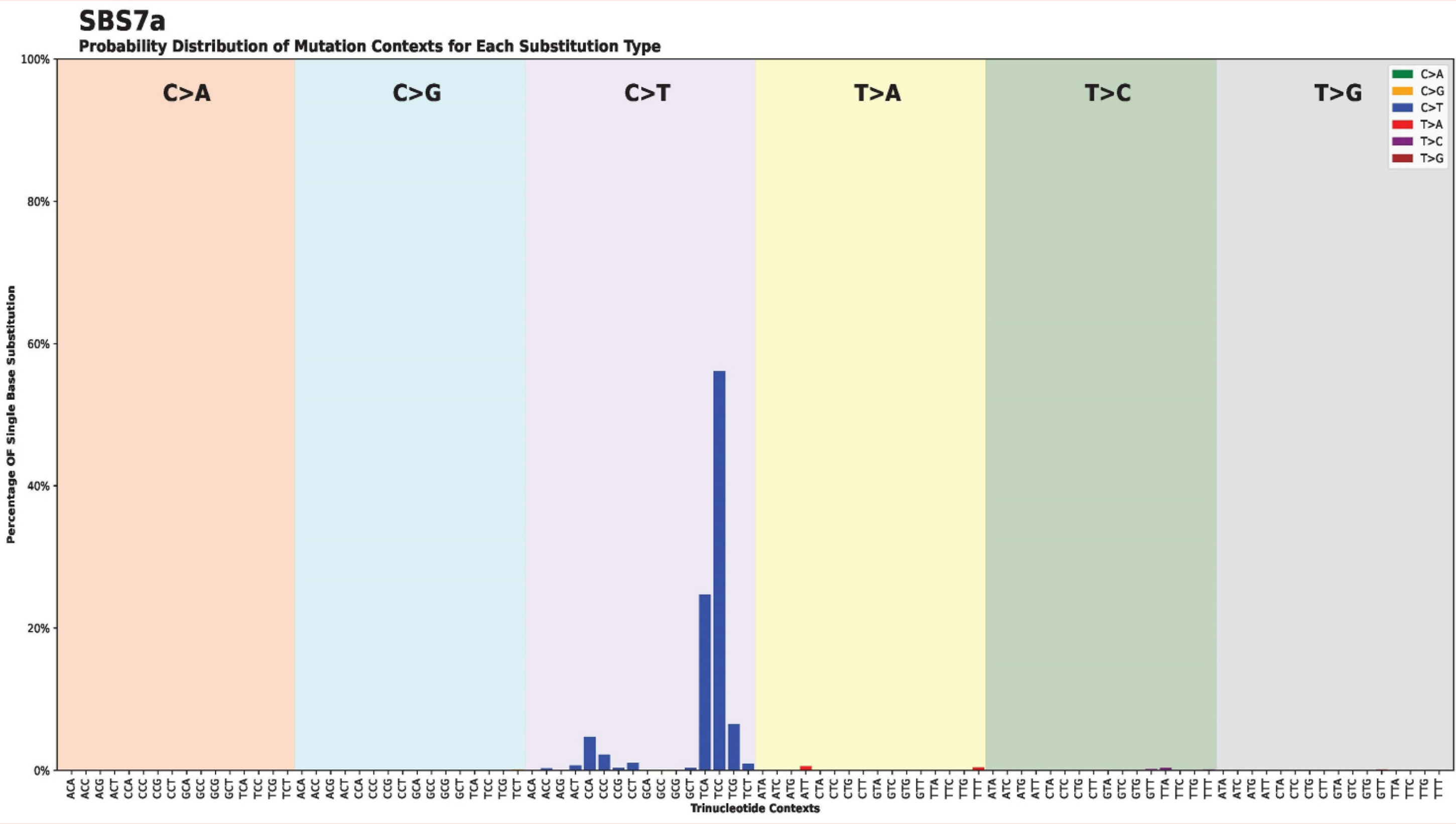


Figure 1: ‘96’ context signature plot of the whole mutation spectrum of SBS7a. This ‘barplot’ shows percentage of single base substitution for all the pyrimidine mutations.

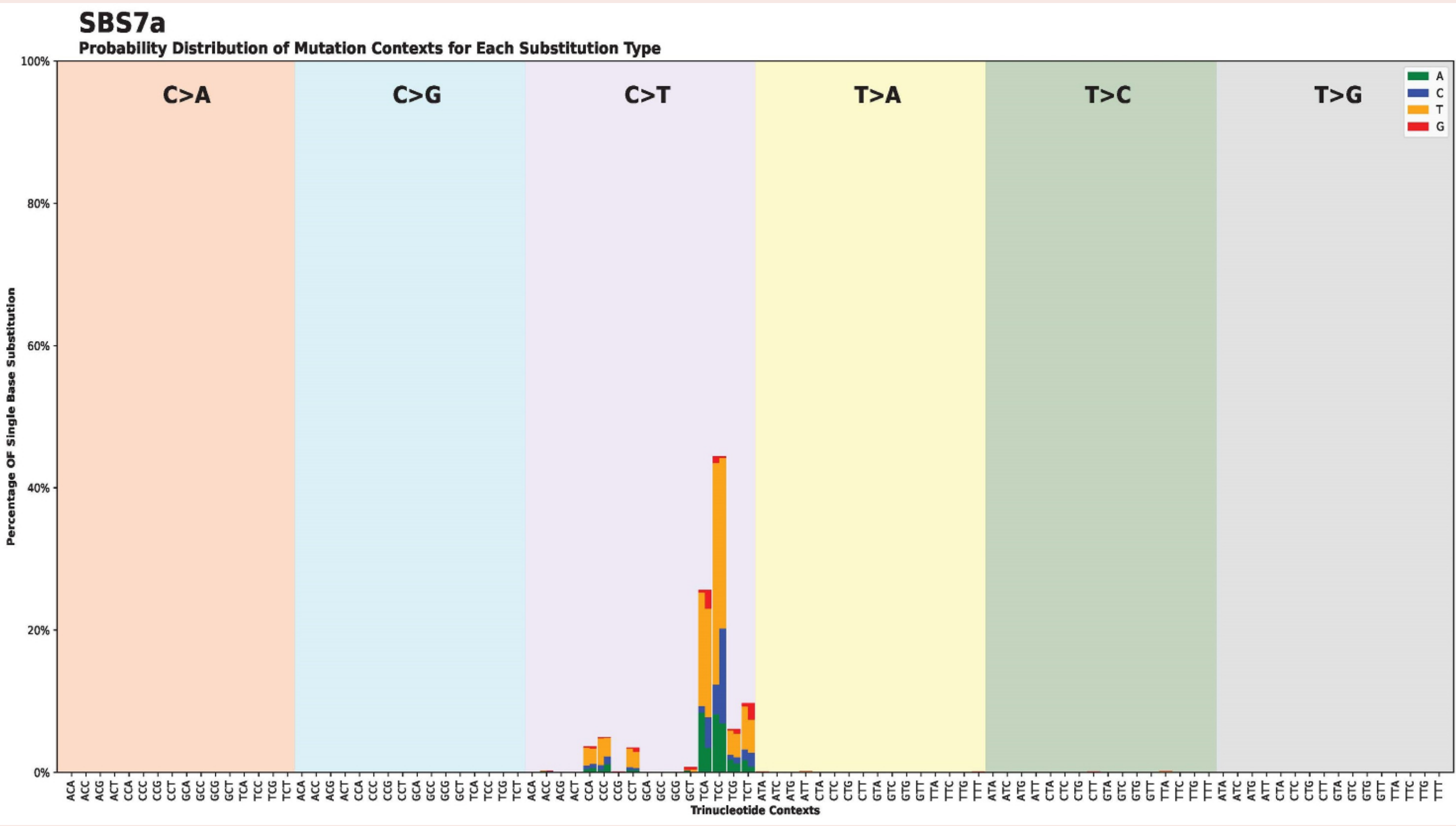


Figure 2: ‘1536’ context, on each side 1 nucleotide, signature plot of the whole mutation spectrum of SBS7a. This ‘barplot’ shows percentage of single base substitution for all the pyrimidine mutations. And the distribution of each extra nucleotide on each side.

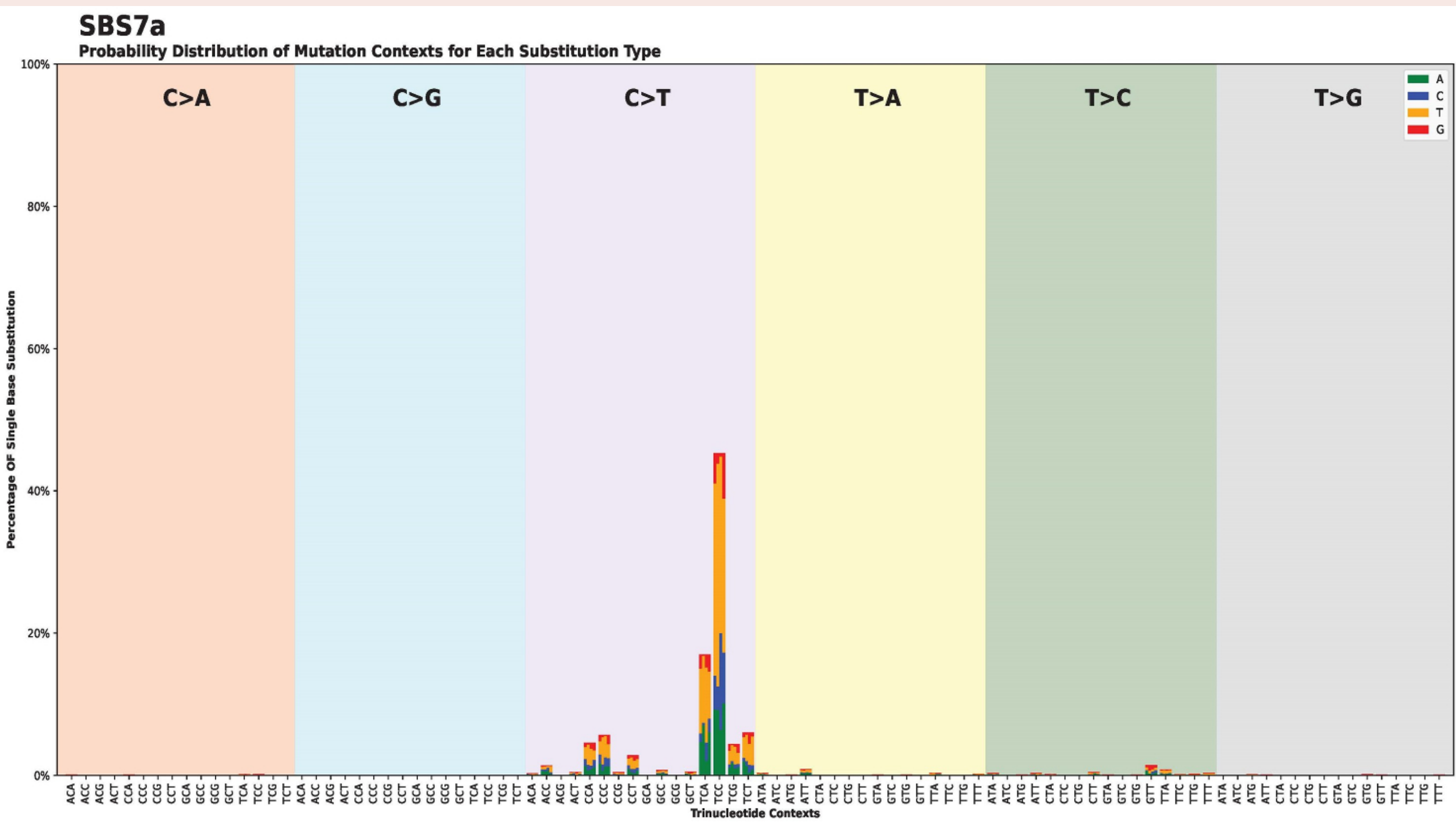


Figure 3: ‘24576’ context, on each side 2 nucleotide context, signature plot of the whole mutation spectrum of SBS7a. This ‘barplot’ shows percentage of single base substitution for all the pyrimidine mutations. And the distribution of each extra nucleotide on each side.

## Discussion

In the context of mutational signature analysis, it is evident that not all signatures uniformly benefit from a larger nucleotide context. This study, however, has demonstrated that specific signatures, particularly SBS7a and SBS10a, show a marked improvement in interpretability and clarity with the inclusion of a larger context.

## Future perspective

Looking ahead, there is potential for future research to extend this context even further, possibly exploring broader genomic regions or delving into aspects like replication timing.

## Acknowledgements



Supervisors:  
Prof. G.A. Lunter – g.a.lunter@umcg.nl  
Dr. H.C. Donker – h.c.donker@umcg.nl



Supervisor:  
M.A. Noback – m.a.noback@pl.hanze.nl