# Rates of somatic hypermutation: through time and along the genome

Joint PhD position between Institut Imagine (Paris) and CIML (Marseille)

Natanael Spisak, Institut Imagine, Inserm
Thomas Dupic, Centre d'Immunologie Marseille–Luminy, Inserm

Mutation is the original source of genetic variation: it fuels evolution, and while most mutations are neutral, a subset causes cancer and genetic disease. Yet how mutations arise remains poorly understood, in part because mutagenesis reflects the net effect of several intertwined processes, including DNA damage, replication, and multiple repair pathways. This project focuses on somatic hypermutation in B cells, a setting in which disentangling the sources of mutation is more tractable.

B cells are unique: during affinity maturation they deliberately hypermutate parts of their genome to improve their recognition of pathogens. Hypermutation targets the immunoglobulin loci, but stray mutations occasionally reach other genes[1], where they can contribute to lymphomas and autoimmunity[2]. Because somatic hypermutation is active in every healthy individual, runs at a high rate, and has a reasonably well-characterised molecular machinery, it is a natural model system for the kinetics of DNA damage and repair, and is central to B-cell biology in its own right.

Existing data on somatic hypermutation covers only short fragments of immunoglobulin genes. Long-read sequencing will let us capture much wider windows and probe off-target sites genome-wide. We will combine this targeted sequencing approach with mechanistic and statistical modelling to address several questions that have so far remained open.

**What is the mutation rate per unit of time?** The mutation rate sets the pace of affinity maturation, so its estimate is central to interpreting repertoire data through quantitative models. The most commonly used estimates express the mutation rate per cell division[3], whereas recent experiments suggest that mutations track time and exposure to damage rather than the number of cell divisions[4]. We set out to infer new estimates from longitudinal data in human and mouse.

**How localised is hypermutation around the target loci?** We will measure how sharply the mutation rate drops off away from the immunoglobulin loci, aiming to identify the mechanism(s) that constrain it to a relatively narrow window. Comparing mutation patterns inside and outside the hypermutating region, alongside ATAC-seq and RNA-seq readouts, will let us probe the effects of chromatin accessibility, transcriptional activity, and proximity to enzyme-targeted motifs.

**Can we disentangle the signatures of the different actors?** Hypermutation in B cells results from the interplay of several enzymes, most importantly the activation-induced cytidine deaminase (AID) and the error-prone polymerase η. We hypothesize that each leaves a distinct mutational signature, and will separate them with non-negative matrix factorisation and related methods, following approaches used for cancer mutational signatures[5,6].

**What goes wrong in pathological hypermutation?** Lymphomas and some inborn errors of immunity show aberrant somatic hypermutation[7], and off-target activity has recently been implicated in

autoimmunity[2]. We will combine the rate estimates and signature decomposition developed above to characterise how the mutational mechanisms may be altered in disease.

**Approach**

The student will develop statistical and mechanistic models of the mutation process and infer their parameters from data using a range of statistical methods, including machine learning techniques for simulation-based inference. They may also contribute to or lead data generation using single-molecule sequencing (PacBio HiFi) on human and mouse samples. The balance between wet-lab and computational work will be adjusted to the candidate's interests and background.

**Profile**

We are looking for a candidate with a strong quantitative background (computational biology, physics, mathematics, or computer science) and a genuine interest in biology. While prior wet-lab experience is not required, candidates interested in both the computational and experimental sides of the project would be an ideal fit.

**Practical details**

- Duration: 3 years, starting fall 2026
- Funding: fully funded PhD contract (ANR)
- Location: Paris (primary) and Marseille
- To apply: send a CV, a short motivation letter, and contact details of one or two referees to natanael.spisak@gmail.com and thomas.dupic@inserm.fr by June 1st.
  Informal enquiries are welcome.

**References**


[1] Machado et al. *Diverse mutational landscapes in human lymphocytes*. Nature 608, 724–732, 2022.
[2] Nicola et al. *Polyclonal selection of immune checkpoint mutations in thyroid autoimmunity*. Nature, 2026.
[3] Kleinstein, Louzoun and Shlomchik. *Estimating hypermutation rates from clonal tree data*. J. Immunol. 171, 4639–4649, 2003.
[4] Pae et al. *Transient silencing of hypermutation preserves B cell affinity during clonal bursting*, Nature 641, 486–494, 2025.
[5] Supek and Lehner. *Clustered mutations in yeast and in human cancers point to mechanisms of mutagenesis in actively transcribed genes*. Cell 170, 534–547, 2017.
[6] Alexandrov et al. *The repertoire of mutational signatures in human cancer*. Nature 578, 94–101, 2020.
[7] Durandy, Kracker and Fischer. *Primary antibody deficiencies*. Nat Rev Immunol. 13, 519–533, 2013.