**Systematic evaluation of somatic *cis*-regulatory mutations in follicular lymphoma**

Firat Uyulur1, Findlay Bewicke-Copley1, Chinedu Anthony Anene1, Matthias Schlesner2, ICGC MMML-Seq Project, Reiner Siebert3, Jessica Okosun4, Jude Fitzgibbon1, Jun Wang1

1Centre for Cancer Genomics and Computational Biology, Barts Cancer Institute, Queen Mary University of London, London, United Kingdom; 2DKFZ Heidelberg, Germany; 3 Institute of Human Genetics, Ulm University and Ulm University Medical Center, Ulm, Germany; 4Centre for Haemato-Oncology, Barts Cancer Institute, Queen Mary University of London, London, United Kingdom.

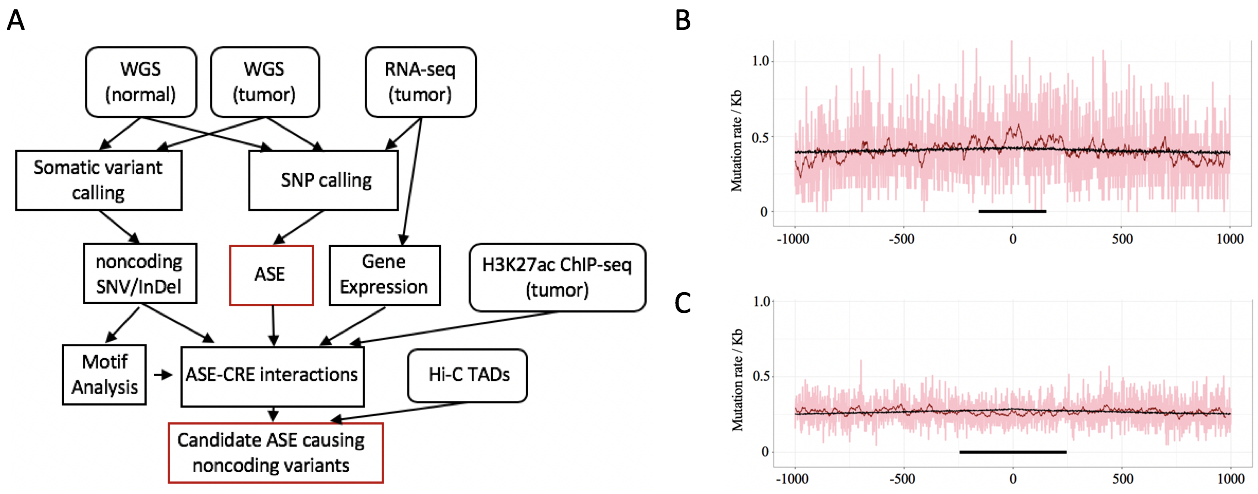
**Background**: Follicular lymphoma (FL) is an incurable indolent B cell malignancy characterized in the majority of cases by the t(14;18) translocation. While the mutational landscape of the coding genome is nearing completion, less is known about the characteristics of its non-coding genome. Our expectation is that the distribution of non-coding mutations will be non-random, determined by epigenomics features, such as chromatin modification and accessibility. Our strategy is therefore to integrate whole-genome epigenomic marks and mutations in order to enrich for variants with regulatory potential and identify unique mutational processes specific to these *cis*-regulatory elements (CREs). Allele-specific expression (ASE) patterns allow further refinement to resolve functional CREs and bona fide mutations associated with changes in gene expression.

**Methods**: The H3K27Ac consensus CREs were determined using the ChIP-Seq data of 9 FL patient primary cells (Koues et al., *Immunity* 2015). DNase I-hypersensitive sites (DHSs) and Hi-C of the B-lymphocyte cell line GM12878 was downloaded from ENCODE. Whole genome sequencing and RNA-Seq of 70 FL patients with relative high tumour cellularity (≥30% for DNA tumour purity and ≥25% for RNA B-cell content) were obtained from the International Cancer Genome Consortium project ICGC MMML-Seq. The mutation rate within CREs and DHSs in comparison to their flanking sequences was estimated using a previous pipeline (Sabarinathan et al., *Nature* 2016). We further developed *cis*-ASE, an integrated analytic pipeline for the identification of recurrent ASE genes, and significantly associated CREs and mutations (Fig1.A).

**Results**: In total 1.04 million non-coding mutations, corresponding to 14.8K mutations per sample with a median of 9,991 non-coding mutations were identified in our series of 70 FL samples. 62K (6.0%) mutations were located within H3K27Ac bound CREs and there was an elevated mutation rate in H3K27Ac CREs compared to the corresponding flanking regions of 1kb up and downstream (χ² test *p* = 1.95e-19, Fig1.B). For DHSs, we observed the opposite pattern, with a lower mutation rate in DHSs than in the flanking regions (*p* = 0.04, Fig1.C), most likely reflecting the higher accessibility to global genome repair machinery in DHSs in relation to flanking sequences. Dividing mutations into high (≥0.3) and low (<0.3) adjusted variant allele frequency (VAF) groups (accounting for tumour purity), we observed significantly higher mutation rate in CREs than in the flanking sequences, that was specific for mutations with higher VAFs (*p* = 6.95e−38), as the difference was much weaker for low VAF mutations (*p* = 0.02). Mutation signatures 6 and 20, linked to defective DNA mismatch repair, were highly enriched for mutations within H3K27Ac CREs (44%) versus other regions outside (27%) (*p* < 0.001).

ASE was assessed using 45.6k informative SNPs (1.7%) per sample wherein matched genotype data was available from WGS and RNA-Seq profiles. *Cis*-ASE identified on average 480 ASE genes per sample (binomial test, adjusted *p*<0.05), corresponding to 1,943 recurrent ASE genes with a minimum threshold of ≥5 samples. These ASE events were not significantly influenced by local copy number changes or promoter methylation. KEGG pathway analysis of recurrent ASE genes identified adherens junction, B cell receptor signaling pathway and Fc gamma R-mediated phagocytosis as the most overrepresented pathways. *Cis*-ASE further identified 18 ASE-CRE interactions where CRE mutations were significantly correlated with an imbalance in elevated alternative allele ratios. These ASE genes included recognized lymphoma related genes including *BCL2*, *STAT6*, *MAF*, and additional novel targets. The pattern of aberrant somatic hypermutation (aSHM) was assessed for these 18 CREs, and we further narrowed down to 15 significant ASE-CRE interactions for further functional validation, consisting of 92 bona fide mutation candidates present in 37 FL samples of our cohort (52.9%).

**Conclusion**: Our study identified unique mutation processes operating in H3K27Ac CREs. Using an integrated genomic approach of whole-genome mutations, chromatin marks and RNA-Seq, we explored ASE-CRE interactions, and identified 15 H3K27Ac bound CREs enriched for *cis*-regulatory mutations significantly associated with ASE of targeted genes, deserving for further exploration to enrich our understanding of FL non-coding genome.



**Figure 1. Cis-regulatory mutations in follicular lymphoma.** (A) An overview of *cis*-ASE pipeline. Mutation rate with CREs compared to the flanking sequences for (B) H3K27Ac CREs and (C) DHSs.