

A Quantitative Evolutionary Approach Utilising High Resolution Chromosomal Copy Number Analysis Accurately Stratifies Patients with Ulcerative Colitis and Low Grade Dysplasia by Future Colorectal Cancer Risk

Al Bakir, I.¹; Curtius, K.¹; Nasreddin, N.²; Clarke, T.S.O.¹; Moorghen, M.³; Rodriguez-Justo, M.⁴; Jansen, M.⁴; Wright, N.A.¹; Leedham, S.J.²; Hart, A.L.³; Graham, T.A.¹

¹Barts Cancer Institute, London, United Kingdom; ²Wellcome Centre for Human Genetics, Oxford, United Kingdom; ³St. Mark's Hospital, Harrow, United Kingdom; ⁴University College Hospital, London, United Kingdom

Introduction: Low grade dysplasia (LGD) in ulcerative colitis (UC) demonstrates a variable risk of progression to colorectal cancer (CRC). Chromosomal copy number alterations (CNAs) are known to occur in UC epithelium. The correlation between LGD CNA burden and future CRC risk is unknown. Shallow whole genome sequencing is a novel, cost-effective technique for high resolution CNA analysis in formalin-fixed, paraffin-embedded tissue.

Methods: We analysed 34 LGD lesions from 22 'progressor' patients who subsequently developed HGD/CRC a median 427 days later (IQR 218-907), and 49 LGD lesions from 45 matched 'non-progressor' patients who remained HGD/CRC-free for >5 years. Histological grading was confirmed by two blinded pathologists.

Results: Both maximal total CNA burden and number of CNA events are greater in LGD of progressor patients than in LGD of non-progressors ($p < 0.001$). Specific CNA events occur at much higher frequencies in progressor LGD, including 4q loss, 5p gain, 17p loss and 17q loss ($OR > 20$, $p_{adj} < 0.01$). Multivariate analysis combining genetic, clinical and endoscopic data demonstrates CNA burden as the only significant risk factor for future CRC risk ($p < 0.001$). Survival analysis of the combined 67 progressor and non-progressor patients demonstrates that those patients bearing LGD with the 25% greatest number CNA events and/or a CNA event on chromosome 17 are much more likely to develop CRC/HGD than the remaining patients ($HR 14.8$, $p < 0.001$). ROC analysis combining clinical and genomic data allows for highly accurate CRC risk prediction, with an AUC of 0.92. Temporospatial phylogenetic analysis in 10 progressor patients with metachronous and/or synchronous neoplasia demonstrates evidence of both clonal expansion (multiple shared CNA events between lesions) and mosaicism (no shared events between lesions).

Conclusion: LGD demonstrates a surprising diversity in CNA burden; some LGD CNA profiles are indistinguishable from the CNA profile of the CRC which subsequently arises in that patient. Shallow whole-genome sequencing output can be used to accurately predict the future CRC risk of LGD.