# Identification of new causative genes in inherited colorectal cancer

Alexandre Xavier MA (Rennes)

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in Medical Genetics

November 2019

**Statement of originality**

I hereby certify that the work embodied in the thesis is my own work, conducted under normal supervision. The thesis contains no material which has been accepted, or is being examined, for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made. I give consent to the final version of my thesis being made available worldwide when deposited in the University's Digital Repository, subject to the provisions of the Copyright Act 1968 and any approved embargo.

16/10/2019

…….. ……..

Date Alexandre Xavier

**Statement of Collaboration:**
I hereby certify that the work embodied in this thesis has been done in collaboration with other researchers. I have included as part of the thesis a statement clearly outlining the extent of collaboration, with whom and under what auspices.

**Statement of Authorship:**

I hereby certify that the work embodied in this thesis contains a published papers/scholarly work of which I am a joint author. I have included as part of the thesis a written statement, endorsed by my supervisor, attesting to my contribution to the joint publications/scholarly work.

**Thesis by Publication:**

I hereby certify that this thesis is in the form of a series of published papers of which I am a co-author. I have included as part of the thesis a written statement from each co- author, endorsed by the Faculty Assistant Dean (Research Training), attesting to contribution to the joint publications.

16/10/2019

…….. ……..

Date Alexandre Xavier

**Acknowledgements:**

I would like, first and foremost, to thank Laureate Professor Rodney J. Scott and Dr Bente Talseth-Palmer for giving me the opportunity to fulfil my PhD project with them. I will always be thankful for Rodney's approachability as his doors remained open even at the busiest of time. Rodney also taught me how to distinguish important things from others. I am also infinitely grateful for Bente's unyielding support during my project as well as her encouragement when I wanted to explore new fields.

I would also give special thanks to Dr Kelly Kiejda. Kelly, I consider you as my 3rd unofficial supervisor and I will always be grateful that you took me under your wing when I was alone and with no one else to turn to.

I want to express my gratitude to Dr Katherine Bolton. You were the one who welcomed me to this lab and you took it upon yourself to teach me everything. Nothing I have done would have been possible without your help.

I would also like to thank Xiajie Zhang, my wonderful stable partner for her support and love through nearly three years of PhD together. You are the reason I have been able to go through all the hardships that PhD life has thrown at me and I would like to share all my achievements with you, as they are all yours.

I want to express all my love to my parents. Thank you, it looks like you did not do a bad job. Thanks you for everything.

The Hunter Cancer Research Alliance played a key role during my PhD. On top of partially funding my PhD scholarship and funding my project through the RHD BTT award, they provided support for every aspect of my project. They were instrumental in the completion of this thesis. In addition to being an amazing structure, they are amazing human beings and I would like to personally thank Dr Stephen Ackland, Ann Thomas, Gemma Hulsing and Sarah Nielsen for all the help they provided.

Many thanks to Joshua Atkins, the Intersect people and Carlos Riveros. Thanks to you, I discovered genomic analysis, coding and bioinformatics.

My gratitude also goes to every single individuals in the IBM lab (and more generally level 3 west at HMRI). Thanks you for being so helpful and for all the smiles in the corridors, they help everyone to go through their day.

I also thank the University of Newcastle for funding a large portion of my PhD scholarship. UoN and HMRI are one of the best working environment I have ever seen.

I would also like to thank, in no particular order and for various reasons:

- Alexandre Gangnant and Paul Noël, you are the “ME” and “TE” to my “OR”.
- William Laporte. Before you, I was a pack of one wolf and now we are a pack of two wolfs.
- Chloe Sanini. For not giving up when I told her to.
- Ophélie Laporte and Romain Nugier. You are at the same time the chillest and most hardcore people I know.
- Ornella Tilly and Jérôme Xavier. For actually reading this thesis.
- Sofia Sanini. For being herself.
- Élodie Gourreau. For simultaneously making the best hummus and having the coolest hair.
- Marie-Caroline Grosset. For all the art.
- Debbie and John Amas for always being there for us.
- Sean Burnard. For still trying to keep me fit and being so wholesome.
- Kumar Uddipto. For Cookies!
- Mamta Pariyar. For being a little bundle of joy.
- Brianna Morten. For being the most helpful person I have ever met.
- Shankar.
- David Mossman and Michael Hipwell for teaching me the same things again and again

ABSTRACT

Colorectal cancer (CRC) remains a heavy burden for all national health systems. It is the third most frequently diagnosed cancer and the second leading cause of death in Australia and worldwide. Around 80% of CRC diagnosed each year are sporadic and somewhere between 7% and 8% have a clearly identified genetic predisposition (inherited CRC cancer; 5% for Lynch Syndrome (LS), 1% for Familial Adenomatous Polyposis (FAP) and 1-2% inclusive for various syndromes with very low incidences), with the remaining ~ 12%-13% being described as "familial". For many patients with a clinical diagnosis of LS and FAP, no causative mutation has been identified in *MSH6*, *MLH1*, *MSH2* or *PMS2* (for LS patients) and in *APC* or *MUTYH (*for FAP patients) as a result of genetic testing.

For those patients and their families, it is critical to identify the genetic cause underlying their increased CRC risk to offer early detection, tightened monitoring and, if required, suitable surgical management.

Establishing an exhaustive list of known genetic risk factors for inherited CRC is essential for families burdened with a high incidence of CRC. Patients with a strong family history of CRC will usually undergo a tighten monitoring. Removing this psychological burden in individuals proven to be non-carriers of pathogenic germline variants is critical.

Initial investigations focused on the Mismatch Repair (MMR) pathway in patients with LS and those with Lynch-Like Syndromes (LLS). 274 DNA samples from LLS patients were sequenced for the 22 genes involved in the MMR pathway to determine the presence of pathogenic variants. The results confirmed that LLS patients harbour pathogenic variants in genes that are not part of routine clinical screening: *POLD1*, *EXO1*, *MLH3*, *RFC1* and *RPA1*. The results indicate that additional MMR genes are involved in the increased risk of CRC in LLS patients.

As the technology evolved and became more cost-effective, whole exome sequencing (WES) was employed. Forty-eight patients with a clinical diagnosis of FAP were recruited based on their family history of CRC, their polyp status and their negative mutational status of *APC* and/or *MUTYH*. WES was used to interrogate all coding regions of the genome. Analysis of pathogenic variants showed that genes involved in DNA repair were frequently associated with a pathogenic variant. In addition, CNV analysis revealed the deletion of large portions of *CFHR3,* known to cause Atypical Haemolytic Uremic Syndrome, leading to ulcerative colitis, a known risk factor in CRC. Analysing the Polygenic Risk Score (PRS) for CRC risk-factors show an enrichment in inflammatory bowel syndrome-related markers.

During the WES analysis of FAP-like patients, an absence of a precise and automated method to predict pathogenicity in cohorts sharing the same phenotype was apparent. To overcome this, we

developed TAPES, a bioinformatics tool that can predict pathogenicity more precisely that can also calculate variant enrichment using only publicly available control sequences. TAPES also integrate powerful variant filtering and can generate useful reports (such as pathway analysis or calculating the total gene burden in a cohort).

In conclusion, the research presented herein helps strengthen the knowledge of familial CRC. The involvement of novel MMR genes in LLS was also revealed thereby expanding the known number of genes associated with this disorder. DNA-repair related genes as well as those involved in inflammation were shown to play an important role in FPS. Finally, a refined analytical pipeline for WES sequencing interpretation was developed providing new bioinformatics tools for the rapid delivery of results.

**List of publications included as part of this thesis:**

1) Xavier A, Olsen MF, Lavik LA, Johansen J, Singh AK, Sjursen W, et al. Comprehensive mismatch repair gene panel identifies variants in patients with Lynch-like syndrome. Mol Genet Genomic Med. 2019;7(8):e850.
2) Xavier A, Scott RJ, Talseth-Palmer BA. TAPES: A tool for assessment and prioritisation in exome studies. PLOS Computational Biology. 2019;15(10):e1007453.
3) Xavier A, Scott RJ, Talseth-Palmer BA. Exome sequencing of unexplained familial polyposis identifies both known and novel causative genes *To be submitted to Clinical Genetics (August 2020)*
4) Xavier A, Scott RJ, Talseth-Palmer BA. IBD-related markers associate with the age of onset for unexplained familial polyposis patients *To be submitted to Clinical Genetics (August 2020)*

**List of additional publications:**

1) Hansen MF, Johansen J, Sylvander AE, Bjornevoll I, Talseth-Palmer BA, Lavik LA, Xavier A, Engebretsen L. F, Scott R. J, Drablos F, Sjursen W. Use of multigene-panel identifies pathogenic variants in several CRC-predisposing genes in patients previously tested for Lynch Syndrome. Clin Genet. 2017. In Appendices section

**List of oral/poster conference presentations:**

1) Xavier A, Scott RJ, Talseth-Palmer BA. TAPES: a Tool for Assessment and Prioritisation in Exome Studies. *Australian Society for Medical Research satellite conference 2019*, Newcastle, NSW, Australia. Poster
2) Alexandre Xavier, Maren Fridtjofsen Hansen, Liss A. Lavik, Ashish Kumar Singh, Rodney J. Scott1,4, Wenche Sjursen and BenteA. Talseth-Palmer: New causative genes in inherited colorectal cancer: A new landscape of mutation for Hereditary Non-Polyposis Colorectal Cancer, *Australian Society for Medical Research conference* 2017*,* Sydney, Australia. Poster
3) Xavier A, Scott RJ, Talseth-Palmer BA. Colorectal Polyposis syndromes. Friday Seminar Series, 19 October 2018, Newcastle, NSW, Australia. Oral

4) Xavier A, Scott RJ, Talseth-Palmer BA. New causative genes in inherited colorectal cancer: Colorectal Polyposis syndromes. *HEaPS seminar 29 March 2017,* Newcastle, NSW, Australia. Oral

*5)* Hansen MF, Johansen J, Sylvander AE, Bjornevoll I, Talseth-Palmer BA, Lavik LA, Xavier A, Engebretsen L. F, Scott R. J, Drablos F, Sjursen W. A new landscape of mutation for Hereditary Non-Polyposis Colorectal Cancer. *HCRA conference 2018, Rapid Fire Session.* Newcastle, NSW, Australia. Oral

*6)* Xavier A, Scott RJ, Talseth-Palmer BA.: Genetics in Inherited Colorectal Cancer, HCRA community showcase, 2017, Newcastle NSW, Australia

**List of Awards:**

1) *HCRA RHD* Award from the Biomarkers and Targeted Therapies Flagship, Hunter Cancer Research Alliance, 2017, AUD $5000 for research purpose
2) *Highly commended Rapid Fire* for A new landscape of mutation for Hereditary Non-Polyposis Colorectal Cancer. *HCRA conference 2018, Rapid Fire Session.*
3) HCRA Future Leaders Group publication award for Xavier A, Scott RJ, Talseth-Palmer BA. TAPES: A tool for assessment and prioritisation in exome studies. PLOS Computational Biology. 2019;15(10):e1007453. AUD $2500 toward open access publication fees

**List of Abbreviations**

| Abbreviation | Expanded term |
|---|---|
| **AC** | Amsterdam Criteria |
| **ACMG** | American College of Medical Genetics |
| **AMP** | Association for Molecular Pathology |
| **BER** | Base Excision Repair (pathway) |
| **BG** | Bethesda Guidelines |
| **bp** | Base Pair |
| **CRC** | Colorectal Cancer |
| **CS** | Cowden Syndrome |
| **DNA** | DeoxyriboNucleic Acid |
| **FAP** | Familial Adenomatous Polyposis |
| **FCCTX** | Familial Colorectal Cancer Type X |
| **FPS** | Familial Polyposis Syndrome |
| **GO** | Gene Ontology |
| **HNPCC** | Hereditary Non-Polyposis Colorectal Cancer |
| **IBD** | Inflammatory Bowel Disease |
| **IHC** | Immunohistochemistry |
| **Indel** | Insertion/Deletion |
| **JPS** | Juvenile Polyposis Syndrome |
| **LLS** | Lynch-Like Syndrome |
| **LS** | Lynch Syndrome |
| **MAP** | MUTYH-Associated Polyposis |
| **MLPA** | Multiplex Ligation-dependent Probe Amplification |
| **MMR** | MisMatch Repair (pathway) |
| **MSI** | Micro-Sattelite Instability |
| **NAP** | NTHL1-Associated Polyposis |
| **NGS** | Next-Generation Sequencing |
| **NSAID** | Non Steroidal Anti-Inflammatory Drug |
| **OMIM** | Online Mendelian Inheritance in Man |
| **PHTS** | PTEN Hamartoma Tumour Syndrome |
| **PJS** | Peutz–Jeghers Syndrome |
| **PPAP** | Polymerase Proofreading-Associated Polyposis |
| **PRS** | Polygenic Risk Score |
| **SAM / BAM / CRAM** | Sequence Alignment Map / Binary / Compressed |
| **SNP** | Single Nucleotide Polymorphism |
| **SNV** | Single Nucleotide Variant |
| **SPS** | Serrated Polyposis Syndromes |
| **VCF** | Variant Calling Format (file) |
| **VUS** | Variant of Unknown Significance |
| **WES** | Whole Exome Sequencing |
| **WGS** | Whole Genome Sequencing |

**List of Tables**



**List of Figures**

# TABLE OF CONTENTS: