Characterising Novel Somatic Variants in Inherited Cylindromas

Author: Sammy Shaya

Supervisors: Dr Neil Rajan and Professor Joris Veltman

Word count: 975

Introduction

Malignant tumours arise from the clonal expansion of a single cell which has acquired a combination of genetic defects instigating uncontrollable cell division which overrides inbuilt mechanisms to inhibit proliferation (1). Inherited tumour predisposition syndromes arise from the acquisition of germline mutations in tumour suppressor genes (2–6). *CYLD* Cutaneous Syndrome (CCS) is such an inherited syndrome leading to the cylindroma, spiradenoma and trichoepitheliomas tumours which each have metaplastic and malignant transformative potential (7) (**Figure 1**).

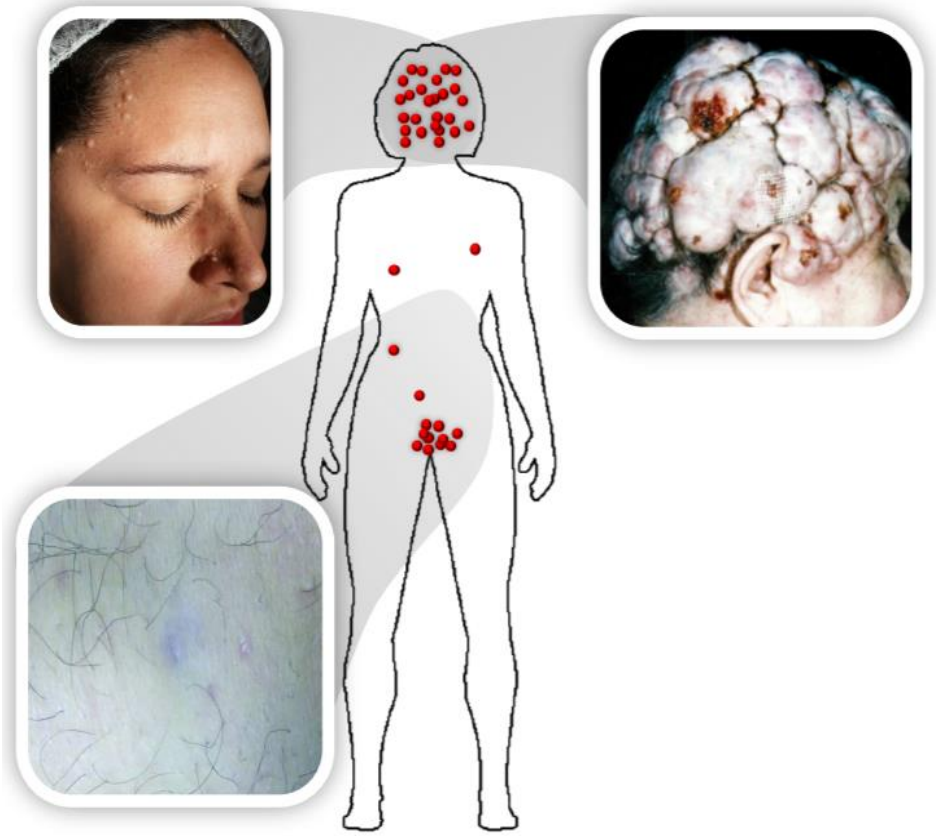

Figure 1. Photographs of the three cardinal symptoms of CCS (21–23)
Extensive scalp cylindromas historically classed as “turban tumour” (upper right), spiradenoma (lower left), and trichoepithelioma (upper right).

Only germline *CYLD* gene mutations undergoing loss of heterozygosity (LOH) are a validated cause for tumourigenesis. Recent discovery of recurrent somatic mutations in epigenetic modifier genes (EMGs) implicate aberrant epigenetic modifications as potentially influential on tumour heterogeneity. Epigenetic modifications are dynamic and alter gene expression variably in response to extracellular stimuli and are frequently dysregulated in carcinogenesis (8–14).

EMGs frequently undergo somatic mutations in cancer which are predicted to alter gene expression and are frequently observed; with limited understanding of resultant effects on tumour proliferation (15–18). Deciphering epigenetic modifications and somatic mutations in EMGs influence epigenetic control, can shed light on the instigation and development of tumours, the regulatory pathways which control their progression, and provide novel therapeutic targets for treatment.

Aim

Hypothesis: *Somatic mutations in EMGs are predicted to occur in hereditary cylindroma tumours in subclonal proportions and may influence pathology.*

Aims: Identify somatic mutations in protein-coding regions of EMGs, for cylindroma and CLM tumours. Model the subclone proportions of somatic mutations within tumours and interpret their significance regarding pathology.

Specific objectives:

- Obtain cylindroma tumour whole-exome sequence dataset externally.
- Filter statistically significant, deleterious somatic variants specific to tumours.
- Prioritise candidate variants from most recurrently altered genes.
- Conduct a subclone analysis on candidate EMG variants and estimate variant subclone proportions in tumour cells within samples.
- Interpret the most significant subclone EMG variants.

Methods

Whole exome data from 31 cylindromas and patient-matched blood controls was obtained externally from patients diagnosed with CCS, which had been subjected to a bioinformatic pipeline to produce variant call files (VCFs). The EpiFactors database (the largest EMG database available) was used as a reference to narrow the variants to those found in 723 EMGs and the *CYLD* gene (19).

The cylindroma EMG variant data was subjected to subtractive calling which removed germline variants (**Figure 2**). The data was then subjected to a bespoke variant filtering pipeline to prioritise somatic variants which alter protein coding DNA sequences and frequently recur within the samples.

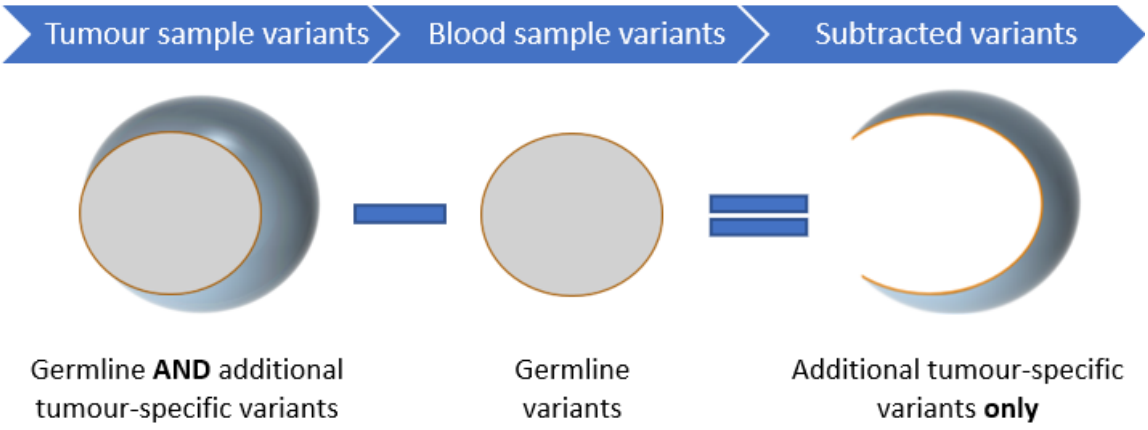

Figure 2. Schematic of subtractive calling process.
Through comparison of VCF data between tumours and their corresponding patient's blood sample, variants common to both blood and tumour in similar proportions were removed. This excludes variants which are thought to be originating from germline tissue, leaving de novo variants or germline variants with elevated variant allele frequencies in subtracted VCFs – thought to be tumour-specific.

Genes harbouring the most frequently altered protein-coding sequence were considered as candidate genes and their variants were assessed for functional impact on tumours. This was followed by a subclone analysis of candidate variant to allow variant subclone modelling within tumours where segments of tumour cells containing candidate variants were quantified by their proportions.

Results

The raw data set contained 159,572 canonical subtracted variants evenly distributed across all cylindromas. 401 variants remained after variant filtering which were considered definitive tumour-specific somatic variants (**Figure 3**). The 401 filtered variants were present in 130 of 723 EMGs. 63 EMGs contained recurrent variants (in >1 cylindroma sample) and the 7 genes that most frequently harboured variants across samples were considered EMGs of interest (Table 1). These represented 92 of 401 filtered variants and CYLD represented 23 of these. Tumours contained somatic variants belonging to between zero and 5 EMGs of interest.

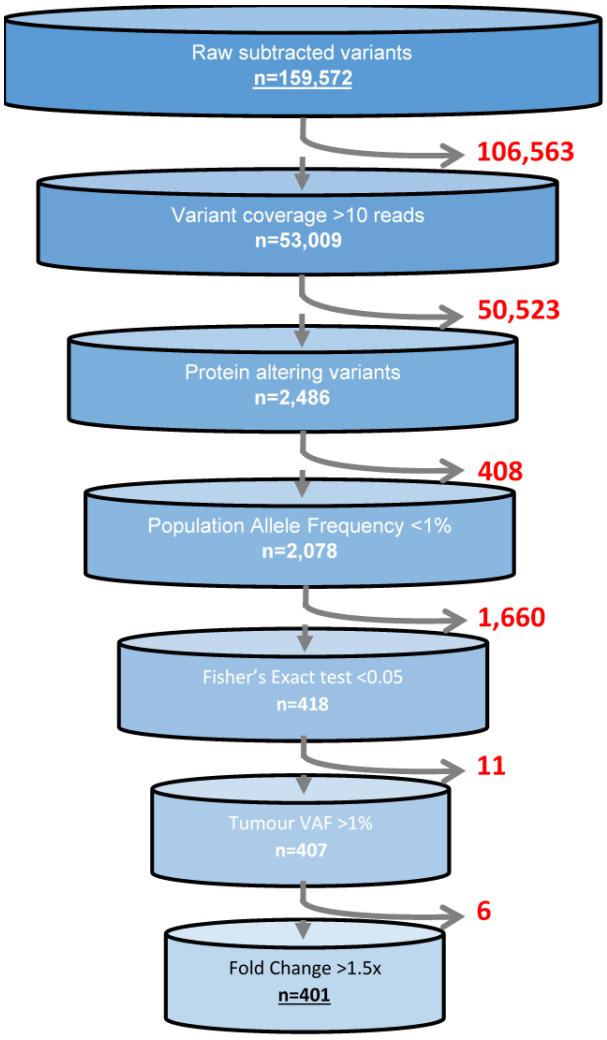

Figure 3. Schematic of variant filtering pipeline.
Flow diagram showing each filtering criteria and the number of subtracted variants which was excluded at each step.

Table 1. *Seven genes most frequently harbouring filtered variants across all cylindroma samples.*

Gene	Cytogenic location	Number of tumours harbouring filtered variants in gene	No. of filtered variants	No. of unique filtered variants
GeneA	16q12.1	20	23	10
GeneB	22q11.1	10	19	4
GeneC	12q23.3	9	12	3
GeneD	6p21.33	8	9	8
GeneE	7q36.1	8	12	9
GeneF	1q42.13	8	9	2
GeneG	5p14.2	7	8	6

7 of 31 samples underwent subclone analysis of novel candidate genes' variants (**Figure 4**).

		GeneB	GeneC	GeneD	GeneE	GeneF	GeneG
2T1	Variant	c.943G>C	c.964_966-1del682			c.757C>G	
	Tumour content (%)	40.3	111.6			38.0	
	P-value	0.00346	0.000000837			0.0425	
2T2	Variant	c.943G>C	c.964_966-1del682		c.2,468T>C		
	Tumour content (%)	34.2	70.6		13.9		
	P-value	0.00198	0.000151		0.0302		
6T1	Variant				c.944G>A		
	Tumour content (%)				14.5		
	P-value				0.0123		
6T2	Variant				c.2,464_2,465insG		
	Tumour content (%)				3.8		
	P-value				0.0358		
7T2	Variant		c.1,090+1_1,091-1del1,219	c.4,202G>T		c.758C>G	
	Tumour content (%)		32.7	21.1		40.7	
	P-value		0.0436	0.0342		0.00111	
10T1	Variant				c.4,219C>G		
	Tumour content (%)				25.9		
	P-value				0.0491		
12T1	Variant				c.1,173C>A		c.2,126C>G
	Tumour content (%)				42.0		74.2
	P-value				0.00922		0.0000061

Figure 4. Results of subclone analysis.
Tumours had subclones modelled based on calculated tumour cell proportions and are displayed in each tumour on the left column. Each subclone variant has its cDNA variant, percentage in the tumour, and p-value for the observed difference in read counts between germline and blood read data; by a Fisher's Exact Test.

Subclones in the *GeneE* and *GeneG* EMGs were the only subclone variants localising to functional domains of their proteins (**Figure 5**). Both proteins carry out methyltransferase action, and alteration of the SET domain sequence is shown to eliminate this function. The nonsense and insertion subclone variants in *GeneE* (c.1,173C>A and c.2,464_2,465insG, respectively) will prevent the SET domain mRNA sequence from being translated and truncates the putative protein – it is therefore predicted to lack methyltransferase action. The *GeneG* variant was predicted to not eliminate protein function.

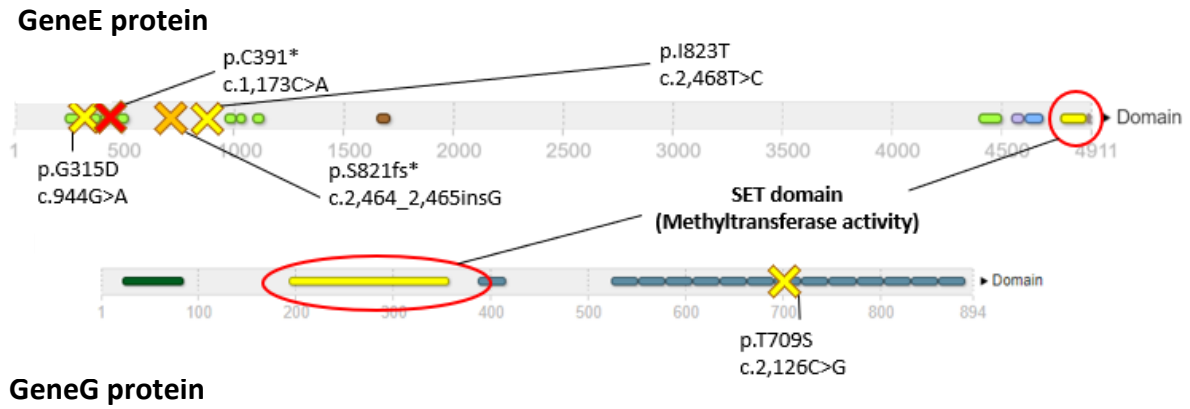

Figure 15. GeneE protein and GeneG protein with corresponding subclone variants superimposed.
The GeneE and GeneG protein sequences are displayed with addition of their functional domains. The SET domain, circled in red, is mandatory for methyltransferase action. The four *GeneE* subclone variants are placed on the sequence in the corresponding position they affect. The nonsense (red) and frameshift (orange) variants will cause protein truncation. The one *GeneG* subclone variant is also shown. All missense variants are shown in yellow.

Discussion

Cylindromas in CCS have frequent heterogeneous somatic mutations in EMGs. Loss of function (LOF) variants in *GeneE* were the most significant in this cohort and represented 4-42% of tumour cells within 4 of 7 samples in the subclone analysis.

GeneE protein adds methyl groups to Histone-3 Lysine-4 amino acid in the octameric histone cores of chromatin which upregulates local gene transcription (20). The effect of the truncating variants observed in these *GeneE* subclones is largely speculative. In the literature *GeneE* is shown to be frequently mutated in malignancy, and murine models have proven that inactivation of the GeneE protein causes embryonic lethality and malignancy compared to functional GeneE protein controls mice. Furthermore, GeneE protein's methyltransferase activity is shown to help upregulate homologous recombination DNA repair pathways; which are shown to be downregulated in other cancers. These lines of evidence suggest that GeneE protein dysregulation may play a fundamental role in cancer pathophysiology, and that the *GeneE* variants observed in cylindromas may inactivate the protein. Further study is required to determine the effect of these variants on protein function, and whether this has implications to tumourigenesis.

Limitations in this study largely surround the subclone analysis as the assumptions underlying this are unvalidated. EMG variant proportion in tumour cells is measured relative to tumour cell purity of the sample, however the zygosity status of the variant is indeterminable from this dataset. Adding to this, the mutational mechanism of *CYLD* LOH events is unknown where different mechanisms would lead to different estimations of purity, and estimations of relative EMG variant subclone proportion.

Conclusion

It is apparent that LOF mutations in EMGs are present in hereditary cylindromas and these could have potential to dysregulate gene expression which could influence known CCS tumourigenic pathways. This suggests that Chromatin Immunoprecipitation Sequencing (ChIP-Seq) studies would help determine whether epigenetic dysregulation occurs in hereditary cylindromas; and performing it in these samples as a follow up would link somatic changes to changes in epigenetic modification profile

1. Hanahan D, Weinberg RA. Hallmarks of Cancer: The Next Generation. Cell [Internet]. 2011 Mar 4 [cited 2019 Jul 18];144(5):646–74. Available from: https://www.sciencedirect.com/science/article/pii/S0092867411001279?via%3Dihub
2. Ryland GL, Doyle MA, Goode D, Boyle SE, Choong DYH, Rowley SM, et al. Loss of heterozygosity: What is it good for? BMC Med Genomics [Internet]. 2015 Dec 1 [cited 2019 Jul 18];8(1):45. Available from: http://bmcmedgenomics.biomedcentral.com/articles/10.1186/s12920-015-0123-z

3. Payne SR, Kemp CJ. Tumor suppressor genetics. Carcinogenesis [Internet]. 2005 Dec 1 [cited 2019 Jul 18];26(12):2031–45. Available from: http://academic.oup.com/carcin/article/26/12/2031/2390810/Tumor-suppressor-genetics

4. Knudson AG. Hereditary Cancer, Oncogenes, and Antioncogenes1 [Internet]. 1985 [cited 2019 Jul 18]. Available from: https://pdfs.semanticscholar.org/0a4d/2fb55efc3fd1a0958c9e231802faa568c4f5.pdf

5. Green AR, Wyke JA. ANTI-ONCOGENES: A Subset of Regulatory Genes Involved in Carcinogenesis? Lancet [Internet]. 1985 Aug 31 [cited 2019 Jul 18];326(8453):475–7. Available from: https://www.sciencedirect.com/science/article/pii/S0140673685904040

6. Klein G, Klein E. Evolution of tumours and the impact of molecular oncology. Nature [Internet]. 1985 May 1 [cited 2019 Jul 18];315(6016):190–5. Available from: http://www.nature.com/articles/315190a0

7. Young A, Kellermayer R, Szigeti R, Tészás A, Azmi S, Celebi J. CYLD mutations underlie Brooke-Spiegler, familial cylindromatosis, and multiple familial trichoepithelioma syndromes. Clin Genet [Internet]. 2006 Aug 10 [cited 2019 Jun 20];70(3):246–9. Available from: http://doi.wiley.com/10.1111/j.1399-0004.2006.00667.x

8. Sharma S, Kelly TK, Jones PA. Epigenetics in cancer. Carcinogenesis [Internet]. 2010 [cited 2019 Aug 2];31(1):27. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2802667/

9. Jones PA, Baylin SB. The Epigenomics of Cancer. Cell [Internet]. 2007 Feb 23 [cited 2019 Aug 2];128(4):683–92. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17320506

10. Wong JJL, Hawkins NJ, Ward RL. Colorectal cancer: a model for epigenetic tumorigenesis. Gut [Internet]. 2007 Jan [cited 2019 Jul 26];56(1):140–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16840508

11. Egger G, Liang G, Aparicio A, Jones PA. Epigenetics in human disease and prospects for epigenetic therapy. Nature [Internet]. 2004 May 27 [cited 2019 Aug 2];429(6990):457–63. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15164071

12. Gillette TG, Hill JA. Readers, writers, and erasers: chromatin as the whiteboard of heart disease. Circ Res [Internet]. 2015 Mar 27 [cited 2019 Aug 2];116(7):1245–53. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25814685

13. Maruyama R, Choudhury S, Kowalczyk A, Bessarabova M, Beresford-Smith B, Conway T, et al. Epigenetic Regulation of Cell Type–Specific Expression Patterns in the Human Mammary Epithelium. Schübeler D, editor. PLoS Genet [Internet]. 2011 Apr 21 [cited 2019 Aug 2];7(4):e1001369. Available from: http://dx.plos.org/10.1371/journal.pgen.1001369

14. Torres IO, Fujimori DG. Functional coupling between writers, erasers and readers of histone and DNA methylation. Curr Opin Struct Biol [Internet]. 2015 Dec [cited 2019

Aug 2];35:68–75. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26496625

15. Rao RC, Dou Y. Hijacked in cancer: the KMT2 (MLL) family of methyltransferases. Nat Rev Cancer [Internet]. 2015 Jun [cited 2019 Jul 26];15(6):334–46. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25998713

16. Baylin SB, Jones PA. A decade of exploring the cancer epigenome — biological and translational implications. Nat Rev Cancer [Internet]. 2011 Oct 23 [cited 2019 Aug 2];11(10):726–34. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21941284

17. Shen H, Laird PW. Interplay between the Cancer Genome and Epigenome. Cell [Internet]. 2013 Mar 28 [cited 2019 Aug 2];153(1):38–55. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23540689

18. You JS, Jones PA. Cancer Genetics and Epigenetics: Two Sides of the Same Coin? Cancer Cell [Internet]. 2012 Jul 10 [cited 2019 Aug 2];22(1):9–20. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22789535

19. Medvedeva YA, Lennartsson A, Ehsani R, Kulakovskiy I V., Vorontsov IE, Panahandeh P, et al. EpiFactors: a comprehensive database of human epigenetic factors and complexes. Database [Internet]. 2015 Jul 7 [cited 2019 Jun 26];2015:bav067. Available from: https://academic.oup.com/database/article-lookup/doi/10.1093/database/bav067

20. Ruault M, Brun ME, Ventura M, Roizès G, De Sario A. MLL3, a new human member of the TRX/MLL gene family, maps to 7q36, a chromosome region frequently deleted in myeloid leukaemia. Gene [Internet]. 2002 Feb 6 [cited 2019 Jul 27];284(1–2):73–81. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11891048

21. Friedrich RE. Dermal cylindroma of the scalp (turban tumour) and subjacent calvarian defects. Anticancer Res [Internet]. 2010 May 1 [cited 2019 Jun 21];30(5):1793–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20592381

22. Zheng Y, Tian Q, Wang J, Dong X, Jing H, Wang X, et al. Differential diagnosis of eccrine spiradenoma: A case report. Exp Ther Med [Internet]. 2014 Oct [cited 2019 Jun 21];8(4):1097–101. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25187803

23. Navarrete-Dechent C, Bajaj S, Marghoob A, Gonzalez S, Munoz D. Multiple familial trichoepithelioma: confirmation via dermoscop. Dermatol Pract Concept [Internet]. 2016 Jul 31 [cited 2019 Jun 21];6(3):51–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27648384