Open

npg

ORIGINAL ARTICLE

Different *APC* genotypes in proximal and distal sporadic colorectal cancers suggest distinct WNT/β-catenin signalling thresholds for tumourigenesis

M Christie[1,2], RN Jorissen[1], D Mouradov[1], A Sakthianandeswaren[1], S Li[1], F Day[1,2], C Tsui[1], L Lipton[1,2,3], J Desai[1,3], IT Jones[4], S McLaughlin[5], RL Ward[6], NJ Hawkins[7], AR Ruszkiewicz[8], J Moore[9], AW Burgess[10], D Busam[11], Q Zhao[11], RL Strausberg[12,13], AJ Simpson[12,13], IPM Tomlinson[14], P Gibbs[1,2,3] and OM Sieber[1,2]

Biallelic protein-truncating mutations in the adenomatous polyposis coli (*APC*) gene are prevalent in sporadic colorectal cancer (CRC). Mutations may not be fully inactivating, instead producing WNT/β-catenin signalling levels 'just-right' for tumourigenesis. However, the spectrum of optimal *APC* genotypes accounting for both hits, and the influence of clinicopathological features on genotype selection remain undefined. We analysed 630 sporadic CRCs for *APC* mutations and loss of heterozygosity (LOH) using sequencing and single-nucleotide polymorphism microarrays, respectively. Truncating *APC* mutations and/or LOH were detected in 75% of CRCs. Most truncating mutations occurred within a mutation cluster region (MCR; codons 1282–1581) leaving 1–3 intact 20 amino-acid repeats (20AARs) and abolishing all Ser-Ala-Met-Pro (SAMP) repeats. Cancers commonly had one MCR mutation plus either LOH or another mutation 5' to the MCR. LOH was associated with mutations leaving 1 intact 20AAR. MCR mutations leaving 1 vs 2–3 intact 20AARs were associated with 5' mutations disrupting or leaving intact the armadillo-repeat domain, respectively. Cancers with three hits had an over-representation of mutations upstream of codon 184, in the alternatively spliced region of exon 9, and 3' to the MCR. Microsatellite unstable cancers showed hyper-mutation at MCR mono- and di-nucleotide repeats, leaving 2–3 intact 20AARs. Proximal and distal cancers exhibited different preferred *APC* genotypes, leaving a total of 2 or 3 and 0 to 2 intact 20AARs, respectively. In conclusion, *APC* genotypes in sporadic CRCs demonstrate 'fine-tuned' interdependence of hits by type and location, consistent with selection for particular residual levels of WNT/β-catenin signalling, with different 'optimal' thresholds for proximal and distal cancers.

INTRODUCTION

Somatic mutations in the adenomatous polyposis coli (*APC*) gene are detected in ~70% of sporadic colorectal cancers (CRCs), with biallelic hits thought to initiate tumourigenesis.[1,2] Germline *APC* mutations underlie familial adenomatous polyposis (FAP) characterized by hundreds to thousands of adenomas in the colorectum and a 100% lifetime-risk of CRC if untreated.[3]
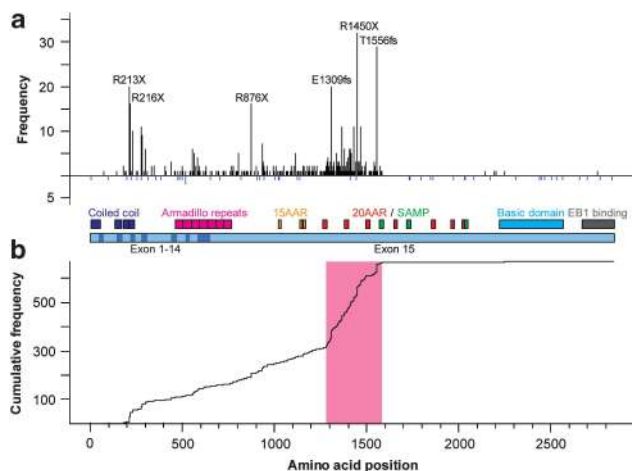
APC has several cellular functions, but negative regulation of WNT/β-catenin signalling is probably its most important for colorectal tumourigenesis.[4] In the absence of WNT ligand, APC forms a complex with GSK3B, CSNK1A1 and AXIN that recruits and phosphorylates β-catenin, targeting β-catenin for proteasomal degradation.[5] The main APC domains involved in β-catenin regulation include a series of seven armadillo repeats (ARM, codons 453–767) that interact with multiple proteins including

the WNT/β-catenin pathway regulatory protein PP2A,[6,7] three 15 amino-acid repeats (15AARs, between codons 1021 and 1170) and seven 20 amino-acid repeats (20AARs, between codons 1265 and 2035) that bind β-catenin, and three interspersed Ser-Ala-Met-Pro (SAMP) repeats that bind AXIN (Figure 1).[4]

Although CRCs usually acquire biallelic protein-truncating mutations in *APC*,[1,2,8,9] several lines of evidence indicate that *APC* is not a typical tumour-suppressor gene. In FAP patients, the severity of polyposis usually correlates with the location of the germline *APC* mutation: mutations in codons 1250–1464 are associated with the highest polyp numbers, whereas mutations in the 5' and 3' regions (codons 1–157 and 1595–2843) and the alternatively spliced region of exon 9 (codons 312–412) are associated with an attenuated phenotype (<100 adenomas, AFAP). However, exceptions to these genotype–phenotype

Figure 1. Distribution of somatic APC mutations by amino acid for 630 sporadic CRCs. (**a**) Frequencies of protein-truncating and missense mutations are shown above and below the x axis, respectively. (**b**) Cumulative frequency of truncating mutations; the somatic mutation cluster region (codons 1282–1581) is highlighted. APC protein domains and exon structure are indicated.

associations occur with both inter- and intra-familial variation in polyp number.[10]

Germline APC mutations are scattered in the 5′ half of the gene with the exception of two hotspot-codons, 1061 and 1309.[11] In contrast, sporadic cancers show a broader clustering of somatic APC mutations in codons 1281–1556, the so-called mutation cluster region (MCR).[1,12,13] The MCR is contained within regions of APC involved in β-catenin downregulation.

In addition, studies on FAP adenomas have revealed interdependence between germline and somatic APC hits.[14–16] Patients with germline mutations around codon 1300, which retain 1 intact 20AAR, acquire loss of heterozygosity (LOH), whereas patients with germline mutations after codon 1398, which retain 2–3 intact 20AARs, acquire truncating mutations 5′ to the MCR. Conversely, patients with germline mutations 5′ to the MCR tend to have somatic mutations in the MCR. This suggests that second hits in APC are selected to produce a 'just-right' level of WNT/β-catenin signalling optimal for tumour development, with the combined hits (or 'just-right' genotypes) resulting in only partial loss of β-catenin regulation.[14–16] Studies of adenomas from AFAP patients have further revealed third hits in APC targeting the germline mutant allele to achieve an optimal genotype.[17]

Data on two-hit associations for APC in sporadic CRC are limited. In CRC cell lines, APC mutations in codons 1194–1392, most of which retain 1 20AAR, were found to be associated with LOH.[13] Analysis of mutation data for cell lines and primary sporadic CRCs further suggested an overrepresentation of tumours with one hit leaving 2 20AARs and the other hit removing all 20AARs.[16] Hypermethylation of APC promoter 1A, but not promoter 1B,[18] has been described in sporadic CRC with associated partial reduction in transcript levels. However, no consistent alteration in WNT/β-catenin signalling was detected, and promoter methylation did not appear to substitute for truncating mutation.[19]

A recent study has suggested that CRCs in the proximal and distal large intestine may have different APC mutation spectra[20] consistent with these representing different genetic categories of disease.[21] Pooling data from studies of sporadic and Lynch syndrome-associated tumours, Albuquerque et al.[20] reported that proximal microsatellite stable (MSS) cancers were more likely to have APC mutations leaving 2–3 20AARs, whereas distal MSS cancers were more likely to have mutations leaving 0–1 20AARs. For microsatellite unstable (MSI) cancers, they further found an overrepresentation of frameshift mutations at nucleotide-repeat

sequences in the MCR producing truncated proteins leaving 2–3 20AARs.[22] However, data were insufficient to evaluate overall APC genotypes.

The major limitation of previous analyses of the somatic APC mutation spectrum in sporadic CRC is that gene screening has generally been incomplete with analyses restricted to the MCR or the 5′ region of the gene. As a result, a complete description of APC genotypes in sporadic CRC and unbiased analysis of the interdependence between APC hits is lacking. Here, we have performed a comprehensive survey of somatic APC mutations in 630 sporadic CRCs, including mutation screening for the entire coding region of the gene and LOH analysis.

## RESULTS

Prevalence and types of somatic APC aberrations

Sequencing of the entire APC coding region detected 621 putative protein-truncating mutations in 437 of 630 (69.4%) sporadic CRCs. Of these, 56.8% (n = 353) were nonsense mutations, 39.0% (n = 242) were out-of-frame insertions or deletions and 4.2% (n = 26) were splice-site mutations. In addition, we detected 50 missense mutations in 43 of 630 (6.8%) cases, and 12 synonymous mutations in 12 of 630 (1.9%) cases (Supplementary Table 1). LOH at APC was detected in 32.1% (202/630) of tumours, occurring by chromosomal deletion (Del) in 79.2% (160/202) and copy-neutral (CN) events in 20.8% (42/202) of cases (Supplementary Figure 1). LOH was associated with the presence of truncating APC mutation, with 83.2% (168/202) of LOH cases also harbouring a truncating change (P < 0.001). For cancers without LOH, there was an overrepresentation of cases with two truncating mutations as compared with a Poisson distribution (P < 0.001). In contrast, presence of a missense APC mutation was not associated with LOH (P = 0.400) or the presence of a truncating mutation (P = 0.391), suggesting that most missense changes were non-pathogenic bystanders. Missense and synonymous mutations were excluded from further analyses.

Overall, 74.8% (471/630) of cancers showed at least one truncating mutation or LOH, 21.6% (136/630) had one detected hit (1 mutation/LOH − or 0 mutations/LOH +), 50.5% (318/630) had two hits (2 mutations/LOH − or 1 mutation/LOH +), 2.7% (17/630) had three hits (3 mutations/LOH − or 2 mutations/LOH +) and 25.2% (159/630) exhibited no APC hit.

Distribution of somatic APC mutations

The distribution of protein-truncating mutations in APC is shown in Figure 1. Nearly all truncating mutations occurred 5′ to codon 1582 (99.2%, 616/621) with pronounced clustering in codons 1282–1581, and we considered this latter region to define the somatic MCR. Overall, 51.9% (322/621) of truncating mutations occurred within the MCR, which represents only 10.6% of the APC coding sequence (P < 0.001). The MCR as identified here corresponded very closely to codons 1285–1584 in which truncating mutations produce APC proteins with 1–3 intact 20AARs, but no intact SAMP repeats. In contrast, somatic missense mutations displayed an even distribution throughout the coding region of APC (Figure 1a).

Within the MCR several mutation-hotspots were apparent, defined here as any codon with a >30-fold increased mutation frequency compared with the average frequency for the entire gene: codons 1309 (110-fold), 1367 (60-fold), 1378 (32-fold), 1414 (32-fold), 1429 (50-fold), 1450 (146-fold), 1465 (50-fold) and 1556 (133-fold). Although these hotspots accounted for 41.6% (134/322) of MCR mutations, the remaining MCR mutations were dispersed over 103 different codons. In addition, six mutation-hotspots were identified upstream of the MCR: codons 213 (92-fold), 216 (73-fold), 232 (46-fold), 283 (41-fold), 876 (73-fold) and 935 (37-fold). The splice-site mutation c.835-8A > G was also overrepresented (50-fold).
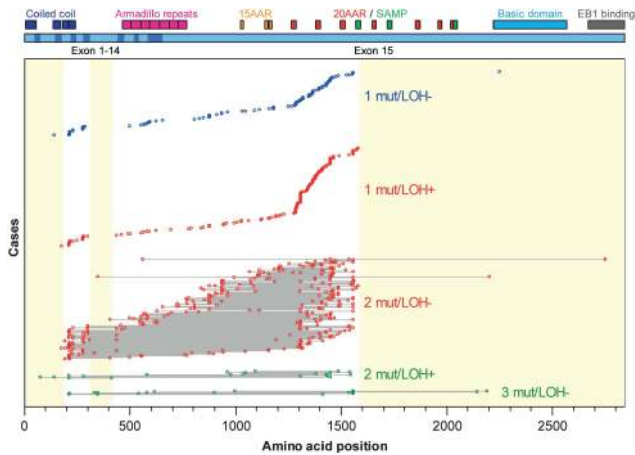
Figure 2. Distribution of truncating *APC* mutations according to the number and types of somatic hits: cases with one hit (1 mutation/LOH−) are shown in blue, with two hits (1 mutation/LOH+ and 2 mutations/LOH−) in red, and with three hits (2 mutations/LOH+ and 3 mutations/LOH−) in green. The y axis shows case number. Regions with few truncating mutations are highlighted.

Figure 3. Distribution of truncating *APC* mutations upstream of the somatic MCR according to number of intact 20AARs left by the MCR hit for CRCs with two mutations, one within the MCR and one 5′ to the MCR (n = 130). Cancers with 2–3 intact 20AARs show an overrepresentation of 5′ non-MCR mutations downstream of codon 767 leaving an intact armadillo-repeat domain, whereas cancers with 1 intact 20AAR show an overrepresentation of 5′ non-MCR mutations upstream of codon 768 disrupting or removing the armadillo-repeat domain. Cases with 3 intact 20AARs are shown in green, 2 intact 20AARs in blue, and 1 intact 20AAR in red.

There were regions of the gene with few observed truncating mutations (Figure 2): 5′ to the alternative start-codon at position 184,[23] the alternatively spliced region of exon 9,[3] and 3′ to the MCR, containing 0.8% (5/621), 1.3% (8/621) and 0.8% (5/621) of mutations but accounting for 6.5%, 3.6% and 44.4% of the coding sequence, respectively ($P < 0.015$ for all comparisons). Notably, cases with three *APC* hits contributed only 6.4% (40/621) of all mutations but accounted for 40.0% (2/5), 50.0% (4/8) and 40% (2/5) of mutations in these regions, respectively, ($P < 0.039$ for all comparisons).
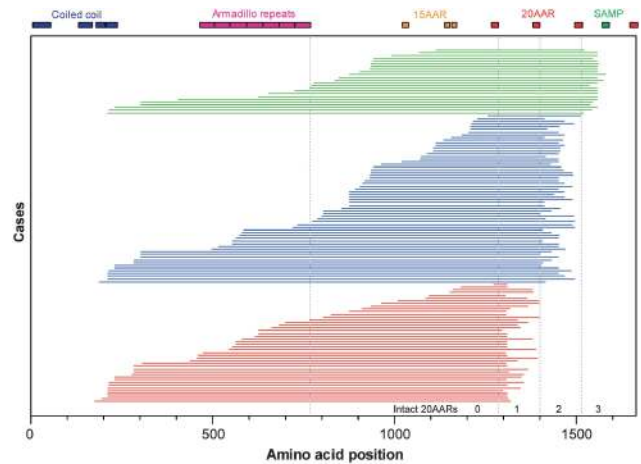
Cases with only one identified truncating mutation and no LOH (Figure 2) showed a distribution of mutations similar to those with two hits, suggesting that these may have undetected second hits in *APC* or perhaps genetic or epigenetic changes in other WNT pathway members.

Interdependence of somatic *APC* hits

For cancers with two hits in *APC* (1 mutation/LOH+, n = 157 or 2 mutations/LOH−, n = 161) resulting genotypes were non-random with respect to mutation types and location (Figure 2). Overall, cancers showed an overrepresentation of genotypes comprising one mutation in the MCR plus either LOH or another mutation 5′ to the MCR: in tumours with 1 mutation/LOH+, 65.6% (103/157) of mutations occurred within the MCR vs 46.6% (150/322) of mutations in tumours with 2 mutations/LOH− ($P < 0.001$); in tumours with 2 mutations/LOH−, 80.7% (130/161) of cases had one mutation within the MCR and the other 5′ to the MCR vs an expected 49.8% for two independent mutations ($P < 0.001$).

The distribution of MCR mutations was further different between cases with 1 mutation/LOH+ and 2 mutations/LOH−. Cases with 1 mutation/LOH+ exhibited a greater proportion of MCR mutations leaving 1 20AAR compared with cases with 2 mutations/LOH− (62.1%, 64/103 vs 34.7%, 52/150, respectively, $P < 0.001$). In addition, cancers with 1 mutation/LOH+ showed different distributions of MCR mutations depending on whether LOH occurred by chromosomal Del (n = 28) or a CN event (n = 75). Cases with CN LOH had more MCR mutations leaving 1 20AAR compared with cases with chromosomal Del (89.3%, 25/28 vs 52.0%, 39/75, respectively, $P < 0.001$), whereas cases with chromosomal Del tended to have MCR mutations leaving 2–3 20AARs.

Finally, for cancers with 2 mutations/LOH− where one mutation was within the MCR and the other 5′ to the MCR

(n = 130), there was significant interdependence between the locations of the two hits (Figure 3). Cases with the MCR mutation leaving 1 20AAR (n = 44) had an overrepresentation of 5′ mutations upstream of codon 768, disrupting the armadillo-repeat domain, compared with cases with MCR mutations leaving 2–3 20AARs (n = 86, 1 20AAR 70.5%, 31/44 vs 2–3 20AARs 37.2%, 32/86, $P < 0.001$). Conversely, cases with the MCR mutation leaving 2–3 20AARs had an overrepresentation of 5′ mutations downstream of codon 767 leaving the armadillo-repeat domain intact.

Different *APC* genotypes in the proximal and distal large intestine

The distributions of truncating *APC* mutations were compared between cancers from the embryonic midgut-derived proximal and hindgut-derived distal large intestine (Figure 4). Cases from the transverse colon were considered to be proximal. The MCR appeared to differ by anatomical location, being approximately codons 1411–1581 for proximal tumours, and codons 1282–1494 for distal tumours, corresponding closely to 2–3 and 1–2 intact 20AARs, respectively. Overall, proximal cancers had an overrepresentation of mutations resulting in 2–3 20AARs compared with distal cancers (2 20AARs: proximal 32.3%, 80/248 vs distal 11.8%, 44/373, $P < 0.001$; 3 20AARs: proximal 16.9%, 42/248 vs distal 2.4%, 9/373, $P < 0.001$), whereas distal cancers displayed an overrepresentation of mutations leaving 0 or 1 20AARs (0 20AARs: proximal 41.5%, 103/248 vs distal 52.3%, 195/373, $P = 0.009$; 1 20AAR: proximal 8.1%, 20/248 vs distal 33.0%, 123/373, $P < 0.001$). In multivariate logistic regression analysis, this association between tumour site and mutation location (0–1 20AARs vs 2–3 20AARs) was independent of cancer MSI status (proximal vs distal odds ratio 0.17, 95% confidence interval 0.11–0.25, $P < 0.001$; MSI vs MSS odds ratio 1.13, 95% confidence interval 0.59–2.15, $P = 0.722$).

Considering colorectal sub-regions, the distribution of mutations appeared dichotomous (Figure 5) with the midgut-derived caecum and ascending colon showing similar enrichment for mutations leaving 2–3 20AARs (caecum 51.9%, 56/108; ascending 48.8%, 40/82), and the hindgut-derived descending colon, sigmoid colon, rectosigmoid and rectum showing similar predominance of mutations leaving 0–1 20AARs (descending 81.1%, 30/37; sigmoid
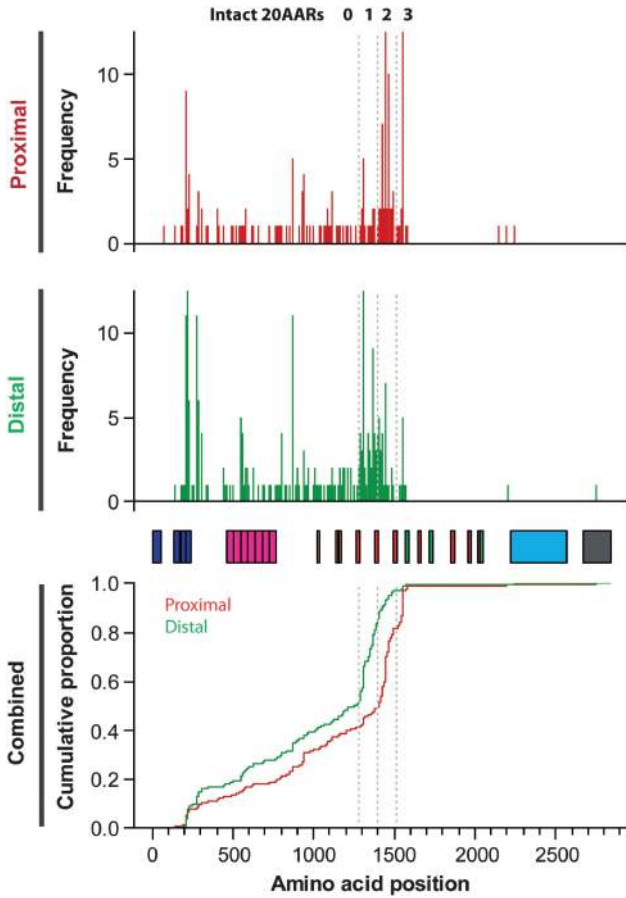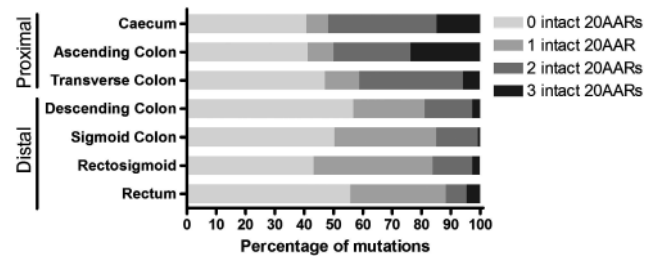
Figure 4. Distribution of truncating *APC* mutations for proximal (red) and distal (green) CRCs. The MCR is approximately codons 1411–1581 for proximal cancers and approximately codons 1282–1494 for distal cancers. Proximal and distal cancers exhibit an overall enrichment for mutations leaving 2–3 and 0–1 intact 20AARs, respectively, as illustrated by the cumulative frequency distributions.

Figure 5. Number of intact 20AARs for truncating mutations in proximal and distal cancers by colorectal sub-region. Cancers from the hepatic flexure were grouped with the ascending colon, and those from the splenic flexure were combined with the descending colon because of small numbers of cases.

Figure 6. Frequency of *APC* genotypes in proximal and distal CRCs with two hits to *APC*. Black asterisk: >5% of proximal cancers. Grey asterisk: >5% of distal cancers. x20AAR, number of intact 20 amino-acid repeats; ARM +, intact armadillo-repeat domain; ARM −, disrupted armadillo-repeat domain; CN, copy-neutral LOH; Del, deletion LOH.
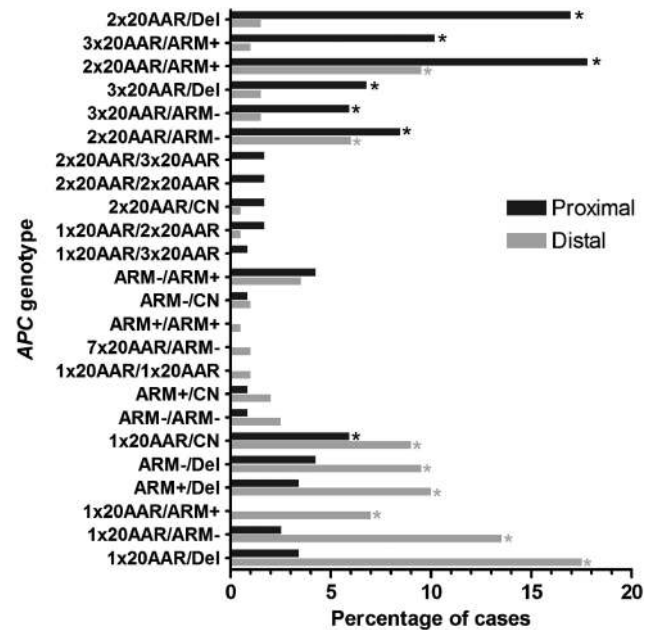
84.4%, 108/128; rectosigmoid 83.8%, 31/37; rectum 87.7%, 114/130). The transverse colon (2/3 midgut and 1/3 hindgut derived) had a distribution more like that of the caecum and ascending colon (2–3 20AARs; 41.2%, 14/34). This dichotomy was supported by logistic regression analysis (Supplementary Table 2).

MSI and MSS cancers showed similar proportions of intact 20AARs for proximal and distal sites (proximal, 2–3 20AARs: MSI 47.5%, 19/40 vs MSS 49.5%, 103/208, $P = 0.864$; distal, 0–1 20AARs: MSI 80.0%, 8/10 vs MSS 85.4%, 310/363, $P = 0.647$). However, when considering proximal cancers there were differences in the types of mutations, with MSI cancers exhibiting an increased frequency of mutations in three nucleotide-repeat sequences: an A5-repeat at codon 1455, an AG5-repeat at codon 1465 and an A6-repeat at codon 1554 (MSI 37.5%, 15/40 vs MSS 10.1%, 21/208, $P < 0.001$).

Although the overall distributions of truncating mutations differed between proximal and distal cancers, similar interdependence between hits for tumours with two hits (1 mutation/LOH + or 2 mutations/LOH −) was seen across anatomical locations. Both proximal and distal cancers showed an overrepresentation of genotypes with one mutation in the MCR plus either LOH ($P < 0.001$ and $P = 0.015$) or another mutation 5′ to the MCR ($P < 0.001$ and $P < 0.001$), an association between LOH and MCR mutations leaving 1 20AAR ($P = 0.027$ and $P < 0.001$), and an association between MCR mutations leaving 1 20AAR and 5′

mutations disrupting the armadillo-repeat domain ($P = 0.049$ and $P = 0.041$).

For cases with two hits, we defined the overall genotype in terms of number of intact 20AARs (1–3 × 20AARs) for MCR mutations, an intact (ARM +) or abolished (ARM −) armadillo-repeat domain for non-MCR mutations, and CN or Del LOH where present (Figure 6). Overall, the most common genotypes in proximal cancers, accounting for 72.0% (85/118) of cases, were 2 × 20AAR/ARM +, 2 × 20AAR/Del, 3 × 20AAR/ARM +, 2 × 20AAR/ARM −, 3 × 20AAR/Del, 1 × 20AAR/CN and 3 × 20AAR/ARM −, and the most common genotypes in distal cancers accounting for 82.0% (164/200) of cases were 1 × 20AAR/Del, 1 × 20AAR/ARM −, ARM + /Del, 2 × 20AAR/ARM +, ARM − /Del, 1 × 20AAR/CN, 1 × 20AAR/ARM + and 2 × 20AAR/ARM −. Notably, the common genotypes in proximal cancers all produced a total of 2 or 3 intact 20AARs, whereas those in distal cancers produced a total of 0, 1 or 2 intact 20AARs. The total numbers of 20AARs presented here are summed from both alleles, ignoring the potential confounding effect of ploidy.

Table 1. Clinicopathological characteristics of 630 patients with sporadic CRC according to the number of somatic *APC* hits (truncating mutation and LOH)

Characteristic	All patients	Number of APC hits					P-value	
		0 Hits (%)	1 Hit (%)	2 Hits (%)	3 Hits (%)	Any hit (%)	0 vs any hit	3 vs 1 or 2 hits
Age, years								
Median	70	73	69	69	73	69	0.002*	0.147
Range	30–99	31–99	30–93	33–92	50–86	30–93		
Gender								
Male	353 (56.0)	70 (19.8)	83 (23.5)	195 (55.2)	5 (1.4)	283 (80.2)	<0.001*	0.011*
Female	277 (44.0)	89 (32.1)	53 (19.1)	123 (44.4)	12 (4.3)	188 (67.9)		
Location								
Proximal colon	273 (43.3)	93 (34.1)	48 (17.6)	118 (43.2)	14 (5.1)	180 (65.9)	<0.001*	0.001*
Distal colon	198 (31.4)	34 (17.2)	47 (23.7)	115 (58.1)	2 (1.0)	164 (82.8)		
Rectum	159 (25.2)	32 (20.1)	41 (25.8)	85 (53.5)	1 (0.6)	127 (79.9)		
Tumour stage								
I	60 (9.5)	12 (20.0)	11 (18.3)	36 (60.0)	1 (1.7)	48 (80.0)	0.200	0.406
II	188 (29.8)	44 (23.4)	37 (19.7)	100 (53.2)	7 (3.7)	144 (76.6)		
III	296 (47.0)	86 (29.1)	68 (23.0)	137 (46.3)	5 (1.7)	210 (70.9)		
IV	86 (13.7)	17 (19.8)	20 (23.3)	45 (52.3)	4 (4.7)	69 (80.2)		
Differentiation								
Well/moderate	458 (76.0)	97 (21.2)	99 (21.6)	249 (54.4)	13 (2.8)	361 (78.8)	0.001*	0.760
Poor	145 (24.0)	52 (35.9)	28 (19.3)	61 (42.1)	4 (2.8)	93 (64.1)		
Unknown	27	10	9	8	0	17		
Lymphovascular invasion								
Absent	215 (63.2)	44 (20.5)	36 (16.7)	124 (57.7)	11 (5.1)	171 (79.5)	0.890	0.583
Present	125 (36.8)	27 (21.6)	25 (20.0)	69 (55.2)	4 (3.2)	98 (78.4)		
Unknown	290	88	75	125	2	202		
Mucinous histology								
No	397 (79.1)	80 (20.2)	88 (22.2)	217 (54.7)	12 (3.0)	317 (79.8)	0.001*	0.192
Yes	105 (20.9)	38 (36.2)	16 (15.2)	46 (43.8)	5 (4.8)	67 (63.8)		
Unknown	128	41	32	55	0	87		
MSI status								
MSS	540 (85.7)	99 (18.3)	127 (23.5)	302 (55.9)	12 (2.2)	441 (81.7)	<0.001*	0.003*
MSI	90 (14.3)	60 (66.7)	9 (10.0)	16 (17.8)	5 (5.6)	30 (33.3)		

Abbreviations: CRC, colorectal cancer; LOH, loss of heterozygosity; MSI, microsatellite unstable; MSS, microsatellite stable. *P<0.05.

APC mutations and nuclear β-catenin expression

For 52 cancers with detected truncating mutations in *APC* and tissue available for immunohistochemistry, the intensity of nuclear β-catenin staining was related to the maximum number of intact 20AARs across mutant alleles, with 0–1 20AARs associated with moderate or strong nuclear β-catenin staining compared with 2–3 20AARs (82.1%, 23/28 vs 54.2%, 13/24, $P = 0.038$; Supplementary Figure 2a). Considering the number of hits (mutation or LOH) detected in *APC*, cases with one hit showed similar levels of nuclear β-catenin staining intensity compared with those with two hits (strong or moderate staining; 64.7%, 11/17 vs 69.2%, 27/39, $P = 0.76$; Supplementary Figure 2b). Cases with no detected *APC* hit ($n = 14$) were more likely to have absent or weak nuclear β-catenin staining compared with those with *APC* mutations (64.3%, 9/14 vs 30.8%, 16/52, $P = 0.031$; Supplementary Figure 2a).

APC hits and clinicopathological features

Associations between somatic *APC* hits and patient characteristics were analysed for truncating mutations and LOH (Table 1). In univariate analysis, any *APC* hit was associated with younger age, male gender, distal colon and rectal tumour location, well/ moderate differentiation, non-mucinous histology and MSS status ($P < 0.002$ for all comparisons). In multivariate logistic regression analysis, younger age, male gender and MSS status remained independently associated with the presence of *APC* mutation (Supplementary Table 3).

Compared with cases with one or two hits ($n = 454$), cases with three hits ($n = 17$) were associated with female gender, proximal tumour location and MSI ($P < 0.012$ for all comparisons). The associations of three hits with proximal tumour location and female gender were independent of MSI status (data not shown).

APC mutations and disease-free survival in stages II–III CRC

Disease-free survival for patients with stages II–III CRC were analysed for association with somatic *APC* mutation status (no *APC* mutation, maximum number of 0–1, or 2–3 20AARs) stratified by tumour location (proximal or distal). For proximal cancers, there was suggestive evidence in univariate analysis for progressively better outcomes for tumours with a maximum of 2–3 20AARs and 0–1 20AARs compared with tumours with no detected *APC* mutation (Supplementary Figure 3a, Supplementary Table 4), although statistical significance was not reached. In multivariate analysis adjusting for potential predictors of patient outcome, 2–3

20AARs and 0–1 20AARs were associated with significantly better survival compared with no *APC* mutation (2–3 20AARs: hazard ratio = 0.50, 95% confidence interval = 0.26–0.97, $P = 0.039$; 0–1 20AARs: hazard ratio = 0.35, 95% confidence interval = 0.13–0.99, $P = 0.047$). The increase in the effect of *APC* genotype on survival in the multivariate analysis was mainly the result of adjusting for MSI status. In multivariate analysis directly comparing tumours with 2–3 20AARs and 0–1 20AARs, the trend to better outcomes in 0–1 20AAR cases was not significant (Supplementary Table 5). For distal cancers, there was no evidence for different outcomes by *APC* mutation status (Supplementary Figure 3b, Supplementary Tables 4 and 5).

DISCUSSION

We have performed a detailed survey of somatic *APC* mutations and LOH in 630 sporadic CRCs. Consistent with previous reports, 75% of cancers harboured somatic truncating mutations or LOH in *APC*, with two hits detected in the majority.[1,2,9] Most truncating mutations occurred 5′ to codon 1582 with pronounced clustering in codons 1282–1581, refining previous definitions of the MCR.[1,12,13] In contrast, missense mutations were infrequent, distributed evenly throughout the coding region and were not associated with the presence of another hit, suggesting that they are mostly non-pathogenic bystanders. Tumours without *APC* hits exhibited distinct clinicopathological features including older age, female gender, proximal tumour location, poor differentiation, mucinous histology, MSI and weaker nuclear β-catenin staining, consistent with these representing a distinct molecular subtype.[20]

Approximately 50% of truncating *APC* mutations occurred in the MCR. Eight hotspots were apparent within this region, accounting for 40% of MCR mutations, however, the remaining MCR mutations were distributed over 103 different codons. This broad distribution indicates selection for mutations throughout the entire MCR. As identified here, the MCR corresponds almost exactly to the codons in which truncating mutations would produce 1–3 intact 20AARs and abolish all SAMP repeats, consistent with the contention that the MCR is defined by these functional domains that are critical for β-catenin regulation.[13,15] Corresponding truncated APC proteins retain residual β-catenin regulatory activity *in vitro* and in animal models.[24,25]

We further identified seven previously unrecognized somatic mutation-hotspots upstream of the MCR. Overall, *APC* mutation-hotspots were explained by two main mechanisms: C-to-T transitions generating TGA stop-codons (40%, 6/15 hotspots) and frameshift mutations at simple nucleotide-repeats (20%, 3/15 hotspots). The former are consistent with spontaneous deamination of 5-methyl-cytosine to thymine at CpG dinucleotides, a mechanism seen in other tumour-suppressor genes including *TP53*.[26]

In FAP patients, germline and somatic hits in *APC* have been shown to be interdependent with regards to the type of somatic hit (truncating mutation or LOH) and the site of somatic mutation (number of intact 20AARs), with selected genotypes proposed to reflect the combined effect of both mutant alleles on WNT/β-catenin signalling.[14–16] Similar associations appear to exist in sporadic CRCs,[13,16] but limited data and incomplete *APC* mutation screening have prevented conclusive analyses. Our data demonstrate clear interdependence between *APC* hits in sporadic CRC. Consistent with previous observations,[13,16] sporadic cancers with two hits tended to have one mutation within the MCR (leaving 1–3 20AARs) and another hit consisting of either LOH or a mutation 5′ to the MCR (leaving 0 20AARs). Considering both hits, 99% (316/318) of cases had no intact SAMP repeats in either allele, and 76% (243/318) had at least one *APC* allele with intact 20AARs. Complete loss of the AXIN binding SAMP repeats thus appears to be required for tumourigenic *APC* genotypes, with further selection for some truncated APC protein

to retain residual ability to downregulate β-catenin via intact 20AARs. The presence of CN LOH was strongly associated with MCR mutations leaving only 1 20AAR consistent with findings for FAP patients with corresponding germline mutations.[14–16] Given that CN LOH results in two copies of the mutant allele, this implies selection for increased allele dosage to reach a total of 2 20AARs summed from both alleles. In addition, we identified novel associations between MCR mutations and mutations 5′ to the MCR. MCR mutations leaving 2–3 20AARs were associated with 5′ mutations downstream of codon 767, leaving an intact armadillo-repeat domain, whereas MCR mutations leaving 1 20AAR were associated with 5′ mutations upstream of codon 768, disrupting the armadillo-repeat domain. These latter associations might reflect further fine-tuning towards a 'just-right' level of WNT/β-catenin signalling beyond selection for particular numbers of 20AARs. The armadillo-repeat domain interacts with multiple proteins including the B56 regulatory subunit of PP2A, reported to both positively and negatively regulate WNT/β-catenin signalling.[6,7] Truncated proteins retaining the armadillo-repeat domain may further retain some capacity to oligomerize through their N-terminal coiled-coil domains, potentially exerting dominant-negative effects.[27]

Proximal and distal CRCs have been suggested to differ in their somatic *APC* mutation spectra and our results support this contention.[20,28] We found that for proximal tumours the MCR corresponded to 2–3 intact 20AARs, whereas for distal tumours the MCR corresponded to 1–2 intact 20AARs. Overall, proximal cancers had an increased frequency of mutations leaving 2–3 20AARs, whereas distal cancers had an increased frequency of mutations leaving 0–1 20AARs. These associations were independent of MSI status. However, the interdependence between hits with regard to the types and locations of hits were similar across anatomical locations, suggesting that only particular genotypes produce the WNT/β-catenin signalling levels required for tumourigenesis, but with different 'optimal' levels for proximal and distal tumours resulting in selection of different genotypes. Favoured genotypes in the proximal colon had a total of 2 or 3 20AARs summed from both alleles, presumably corresponding to lower levels of WNT/β-catenin signalling, whereas favoured genotypes in the distal colorectum had a total of 0, 1 or 2 20AARs, overall corresponding to higher levels of signalling. Supporting this hypothesis, we found a greater overall intensity of nuclear β-catenin staining in tumours with mutations leaving 0–1 20AARs vs 2–3 20AARs. In proximal cancers, there was suggestive evidence for progressively better outcomes for tumours with a maximum of 2–3 20AARs and 0–1 20AARs compared with cases with no *APC* mutation, consistent with 0–1 20AARs representing a suboptimal genotype. However, the direct comparison between the groups did not reach statistical significance. Many factors could contribute to differential genotype selection in proximal and distal cancers, including the bacterial flora, mucin content, metabolism, density of immune cells, DNA repair mechanisms and/or embryologic derivation.[28] Our data support a role for different embryologic origins, showing a dichotomy in the number of 20AARs selected between sub-regions from midgut-derived proximal, and hindgut-derived distal large intestine.

Compared with proximal MSS cancers, proximal MSI cancers showed an overrepresentation of mutations in three nucleotide-repeat sequences within the MCR: codons 1455 (A5), 1465 (AG5) and 1554 (A6), consistent with mismatch-repair deficiency-associated hypermutation. These frameshift mutations retain 2 or 3 20AARs, the favoured numbers for proximal cancers, and this might partly account for the predominance of MSI cancers in the proximal colon as previously proposed.[22]

Additional evidence for the selective pressure to achieve an optimal *APC* genotype is the combinations of changes observed in cancers with three hits, which accounted for 3% of sporadic CRCs. These showed a marked overrepresentation of mutations before

the alternative start-codon 184, in the alternatively spliced region of exon 9, and 3′ to the MCR, regions in which mutations appear suboptimal for tumourigenesis because of production of residual wild-type protein or intact SAMP repeats. This pattern is reminiscent of patients with AFAP who often carry germline mutations in these regions and whose tumours tend to acquire two additional somatic hits, one targeting the wild-type and one the germline mutant allele.[17] Evolution of sporadic cancers from suboptimal genotypes (rather than random acquisition of a suboptimal mutation in a cancer with two hits) is supported by the observation that such cases showed distinct clinico-pathological features including associations with female gender and proximal tumour location that were independent of MSI status. A tendency to develop tumours in the proximal colon is also observed for AFAP patients,[10] suggesting that early tumourigenesis with suboptimal APC mutations—presumably with lesser impact on WNT/β-catenin signalling—is favoured in the proximal colon.

In conclusion, we have found strong evidence that somatic APC hits in sporadic CRCs are selected to achieve 'optimal' genotypes, with interdependent hits targeting domains critical to β-catenin regulation in specific combinations. Selection is evident for types of hits (truncating mutation vs LOH), locations of truncating mutations with respect to number of intact 20AARs and the armadillo-repeat domain, and mechanisms of LOH (Del vs CN). Cancers from the proximal and distal colorectum differ substantially in their distributions of APC genotypes and show a dichotomy for retention of different numbers of 20AARs suggesting different WNT/β-catenin signalling thresholds for tumourigenesis in these embryologically distinct regions. Although MSS cancers are usually considered to be relatively homogeneous, our results suggest that proximal and distal MSS cancers may differ in their biology and perhaps should be considered separately in future studies. Our findings suggest that even moderate modulation of WNT/β-catenin signalling levels could have antitumour effects, and support the rationale for the development of inhibitors of this pathway, which are showing initial promise.[29–31]

MATERIALS AND METHODS

Patients and material
Fresh-frozen tumour and normal tissues were available from 630 sporadic CRC patients treated at the Royal Melbourne Hospital, Western Hospital Footscray, Prince of Wales Hospital Sydney, and Royal Adelaide Hospital, Australia. All patients gave informed consent and this study was approved by the human research ethics committees of all sites. None of the patients had clinical features of FAP, Lynch or other familial cancer syndromes. Clinicopathological features are shown in Table 1.

Mutation detection
Specimen histology was reviewed, macro-dissection performed for cancers to ensure >60% tumour cell content and genomic DNA extracted. Bi-directional Sanger sequencing of APC exons and exon–intron boundaries was performed using 3730xl Genetic Analyzers (Applied Biosystems, Foster City, CA, USA). Detected mutations were confirmed by re-sequencing of tumour and normal DNA from new PCR products. As a result of difficulties with primer design for high-throughput sequencing, APC exon 11 was screened by high-resolution melting curve analysis on a 7500 Fast Real-Time PCR system (Applied Biosystems), with variant samples analysed by sequencing. Primer details are available from the authors.

LOH analysis
LOH analysis at the APC locus was performed using single-nucleotide polymorphism microarray data for tumour and matched normal DNA samples (Human610-Quad BeadChip, Illumina, San Diego, CA, USA). LOH and copy-number states were determined using OncoSNP software (Isis Innovations, Oxford, UK) (Supplementary Figure 1).[32]

Microsatellite instability analysis
MSI status was determined using the Bethesda microsatellite panel.[33] MSI was considered present if instability was seen at ⩾2 markers.

Immunohistochemistry
Tissue-microarrays containing 66 cancers were analysed for nuclear β-catenin by immunohistochemistry (Dako, Carpinteria, CA, USA; clone β-catenin-1, 1:100). Heat-induced antigen retrieval in EDTA (pH 9.0) buffer was followed by blocking in 3% hydrogen peroxide and 2% bovine serum albumin. Detection was performed using the EnVision + Detection System (Dako). Nuclear staining was scored as absent (0), weak (1 +), moderate (2 +) or strong (3 +).

Statistical analysis
Differences between groups were assessed using Fisher's exact test for categorical variables and the Kruskal–Wallis test for continuous variables. Multivariate analysis for associations between APC mutation status and patient characteristics was performed using logistic regression. Outcome analyses were performed for disease-free survival right-censored at 5 years. Disease-free survival was defined as time from surgery to first relapse. Univariate survival distributions were compared using a log-rank test. Cox proportional hazards models were used to estimate survival distributions and hazard ratios. Statistical analyses were two-sided and considered significant if $P < 0.05$.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

REFERENCES
1 Miyoshi Y, Nagase H, Ando H, Horii A, Ichii S, Nakatsuru S et al. Somatic mutations of the APC gene in colorectal tumors: mutation cluster region in the APC gene. Hum Mol Genet 1992; 1: 229–233.
2 Powell SM, Zilz N, Beazer-Barclay Y, Bryan TM, Hamilton SR, Thibodeau SN et al. APC mutations occur early during colorectal tumorigenesis. Nature 1992; 359: 235–237.
3 Groden J, Thliveris A, Samowitz W, Carlson M, Gelbert L, Albertsen H et al. Identification and characterization of the familial adenomatous polyposis coli gene. Cell 1991; 66: 589–600.
4 Nathke IS. The adenomatous polyposis coli protein: the achilles heel of the gut epithelium. Annu Rev Cell Dev Biol 2004; 20: 337–366.
5 Klaus A, Birchmeier W. Wnt signalling and its impact on development and cancer. Nat Rev Cancer 2008; 8: 387–398.
6 Seeling JM, Miller JR, Gil R, Moon RT, White R, Virshup DM. Regulation of beta-catenin signaling by the B56 subunit of protein phosphatase 2A. Science 1999; 283: 2089–2091.
7 Ratcliffe MJ, Itoh K, Sokol SY. A positive role for the PP2A catalytic subunit in Wnt signal transduction. J Biol Chem 2000; 275: 35680–35683.
8 Miyaki M, Konishi M, Kikuchi-Yanoshita R, Enomoto M, Igari T, Tanaka K et al. Characteristics of somatic mutation of the adenomatous polyposis coli gene in colorectal tumors. Cancer Res 1994; 54: 3011–3020.
9 Samowitz WS, Slattery ML, Sweeney C, Herrick J, Wolff RK, Albertsen H. APC mutations and other genetic and epigenetic changes in colon cancer. Mol Cancer Res 2007; 5: 165–170.
10 Nieuwenhuis M, Vasen H. Correlations between mutation site in APC and phenotype of familial adenomatous polyposis (FAP): a review of the literature. Crit Rev Oncol Hematol 2007; 61: 153–161.

11 Béroud C, Soussi T. APC gene: database of germline and somatic mutations in human tumors and cell lines. *Nucleic Acids Res* 1996; **24**: 121–124.

12 Cheadle JP, Krawczak M, Thomas MW, Hodges AK, Al-Tassan N, Fleming N *et al.* Different combinations of biallelic APC mutation confer different growth advantages in colorectal tumours. *Cancer Res* 2002; **62**: 363–366.

13 Rowan AJ, Lamlum H, Ilyas M, Wheeler J, Straub J, Papadopoulou A *et al.* APC mutations in sporadic colorectal tumors: a mutational ''hotspot'' and interdependence of the ''two hits''. *Proc Natl Acad Sci USA* 2000; **97**: 3352–3357.

14 Lamlum H, Ilyas M, Rowan A, Clark S, Johnson V, Bell J *et al.* The type of somatic mutation at APC in familial adenomatous polyposis is determined by the site of the germline mutation: a new facet to Knudson's 'two-hit' hypothesis. *Nat Med* 1999; **5**: 1071–1075.

15 Albuquerque C, Breukel C, van der Luijt R, Fidalgo P, Lage P, Slors FJM *et al.* The 'just-right' signaling model: APC somatic mutations are selected based on a specific level of activation of the beta-catenin signaling cascade. *Hum Mol Genet* 2002; **11**: 1549–1560.

16 Crabtree M, Sieber OM, Lipton L, Hodgson SV, Lamlum H, Thomas HJW *et al.* Refining the relation between 'first hits' and 'second hits' at the APC locus: the 'loose fit' model and evidence for differences in somatic mutation spectra among patients. *Oncogene* 2003; **22**: 4257–4265.

17 Sieber OM, Segditsas S, Knudsen AL, Zhang J, Luz J, Rowan AJ *et al.* Disease severity and genetic pathways in attenuated familial adenomatous polyposis vary greatly but depend on the site of the germline mutation. *Gut* 2006; **55**: 1440–1448.

18 Esteller M, Sparks A, Toyota M, Sanchez-Cespedes M, Capellá G, Peinado MA *et al.* Analysis of adenomatous polyposis coli promoter hypermethylation in human cancer. *Cancer Res* 2000; **60**: 4366–4371.

19 Segditsas S, Sieber OM, Rowan A, Setien F, Neale K, Phillips RKS *et al.* Promoter hypermethylation leads to decreased APC mRNA expression in familial polyposis and sporadic colorectal tumours, but does not substitute for truncating mutations. *Exp Mol Pathol* 2008; **85**: 201–206.

20 Albuquerque C, Baltazar C, Filipe B, Penha F, Pereira T, Smits R *et al.* Colorectal cancers show distinct mutation spectra in members of the canonical WNT signaling pathway according to their anatomical location and type of genetic instability. *Genes Chromosomes Cancer* 2010; **49**: 746–759.

21 Bufill JA. Colorectal cancer: evidence for distinct genetic categories based on proximal or distal tumor location. *Ann Intern Med* 1990; **113**: 779–788.

22 Albuquerque C, Bakker ERM, van Veelen W, Smits R. Colorectal cancers choosing sides. *Biochim Biophys Acta* 2010; **1816**: 219–231.

23 Heppner Goss K, Trzepacz C, Tuohy TMF, Groden J. Attenuated APC alleles produce functional protein from internal translation initiation. *Proc Natl Acad Sci USA* 2002; **99**: 8161–8166.

24 Kielman MF, Rindapää M, Gaspar C, van Poppel N, Breukel C, van Leeuwen S *et al.* Apc modulates embryonic stem-cell differentiation by controlling the dosage of β-catenin signaling. *Nat Genet* 2002; **32**: 594–605.

25 Roberts DM, Pronobis MI, Poulton JS, Waldmann JD, Stephenson EM, Hanna S *et al.* Deconstructing the ßcatenin destruction complex: mechanistic roles for the tumor suppressor APC in regulating Wnt signaling. *Mol Biol Cell* 2011; **22**: 1845–1863.

26 Tornaletti S, Pfeifer GP. Complete and tissue-independent methylation of CpG sites in the p53 gene: implications for mutations in human cancers. *Oncogene* 1995; **10**: 1493–1499.

27 Joslyn G, Richardson DS, White R, Alber T. Dimer formation by an N-terminal coiled coil in the APC protein. *Proc Natl Acad Sci USA* 1993; **90**: 11109–11113.

28 Leedham SJ, Rodenas-Cuadrado P, Howarth K, Lewis A, Mallappa S, Segditsas S *et al.* A basal gradient of Wnt and stem-cell number influences regional tumour distribution in human and mouse intestinal tracts. *Gut* (e-pub ahead of print 27 January 2012; doi:10.1136/gutjnl-2011-301601).

29 Chen B, Dodge ME, Tang W, Lu J, Ma Z, Fan C-W *et al.* Small molecule–mediated disruption of Wnt-dependent signaling in tissue regeneration and cancer. *Nat Chem Biol* 2009; **5**: 100–107.

30 Ewan K, Pajak B, Stubbs M, Todd H, Barbeau O, Quevedo C *et al.* A useful approach to identify novel small-molecule inhibitors of Wnt-dependent transcription. *Cancer Res* 2010; **70**: 5963–5973.

31 Sack U, Walther W, Scudiero D, Selby M, Aumann J, Lemos C *et al.* S100A4-induced cell motility and metastasis is restricted by the Wnt/β-catenin pathway inhibitor calcimycin in colon cancer cells. *Mol Biol Cell* 2011; **22**: 3344–3354.

32 Yau C, Mouradov D, Jorissen RN, Colella S, Mirza G, Steers G *et al.* A statistical approach for detecting genomic aberrations in heterogeneous tumor samples from single nucleotide polymorphism genotyping data. *Genome Biol* 2010; **11**: R92.

33 Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW *et al.* A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res* 1998; **58**: 5248–5257.

Author/s:

Christie, M;Jorissen, RN;Mouradov, D;Sakthianandeswaren, A;Li, S;Day, F;Tsui, C;Lipton, L;Desai, J;Jones, IT;McLaughlin, S;Ward, RL;Hawkins, NJ;Ruszkiewicz, AR;Moore, J;Burgess, AW;Busam, D;Zhao, Q;Strausberg, RL;Simpson, AJ;Tomlinson, IPM;Gibbs, P;Sieber, OM

Title:

Different APC genotypes in proximal and distal sporadic colorectal cancers suggest distinct WNT/beta-catenin signalling thresholds for tumourigenesis

Date:

2013-09-26

Citation:

Christie, M., Jorissen, R. N., Mouradov, D., Sakthianandeswaren, A., Li, S., Day, F., Tsui, C., Lipton, L., Desai, J., Jones, I. T., McLaughlin, S., Ward, R. L., Hawkins, N. J., Ruszkiewicz, A. R., Moore, J., Burgess, A. W., Busam, D., Zhao, Q., Strausberg, R. L. ,... Sieber, O. M. (2013). Different APC genotypes in proximal and distal sporadic colorectal cancers suggest distinct WNT/beta-catenin signalling thresholds for tumourigenesis. ONCOGENE, 32 (39), pp.4675-4682. https://doi.org/10.1038/onc.2012.486.

Persistent Link:

http://hdl.handle.net/11343/264963

License:

CC BY-NC-ND