Identification of 33 candidate oncogenes by screening for base-specific mutations

S Tuupanen1,2, U A Hänninen1,2, J Kondelin1,2, P von Nandelstadh2, T Cajuso1,2, A E Gylfe1,2, R Katainen1,2, T Tanskanen1,2, H Ristolainen1,2, J Böhm3, J-P Mecklin4, H Järvinen5, L Renkonen-Sinisalo5, C L Andersen6, M Taipale2,7, J Taipale2,7, P Vahteristo1,2, K Lehti2, E Pitkänen1,2 and L A Aaltonen *,1,2

1Department of Medical Genetics, Haartman Institute, University of Helsinki, Helsinki 00014, Finland; 2Research Programs Unit, Genome-Scale Biology Research Program, University of Helsinki, Biomedicum, PO Box 63 (Haartmaninkatu 8), Helsinki FIN-00014, Finland; 3Department of Pathology, Jyväskylä Central Hospital, University of Eastern Finland, Keskussairaantalantie 19, Jyväskylä 40620, Finland; 4Department of Surgery, Jyväskylä Central Hospital, University of Eastern Finland, Keskussairaantalantie 19, Jyväskylä 40620, Finland; 5Department of Surgery, Helsinki University Central Hospital, Haartmaninkatu 4, Helsinki 00290, Finland; 6Department for Molecular Medicine (MOMA), Aarhus University Hospital, Brendstrupgårdsvej 21, Aarhus N DK-8200, Denmark and 7Department of Biosciences and Nutrition, Karolinska Institutet, Hälsövägen 7, Huddinge SE-141 83, Sweden

Background: Genes with recurrent codon-specific somatic mutations are likely drivers of tumorigenesis and potential therapeutic targets. Hypermutable cancers may represent a sensitive system for generation and selection of oncogenic mutations.

Methods: We utilised exome-sequencing data on 25 sporadic microsatellite-instable (MSI) colorectal cancers (CRCs) and searched for base-specific somatic mutation hotspots.

Results: We identified novel mutation hotspots in 33 genes. Fourteen genes displayed mutations in the validation set of 254 MSI CRCs: ANTXR1, MORC2, CEP135, CRYBB1, GALNT9, KRT82, PI15, SLC36A1, CNTF, GLDC, MBTPS1, OR9Q2, R3HDM1 and TTPAL. A database search found examples of the hotspot mutations in multiple cancer types.

Conclusions: This work reveals a variety of new recurrent candidate oncogene mutations to be further scrutinised as potential therapeutic targets.

The functional effect of somatic mutations can sometimes be predicted from mutation patterns. Inactivating mutations often truncate the protein product, and frequent occurrence of such changes in a gene is a sign of a possible tumour suppressor function, as exemplified by the APC gene in colorectal cancer (CRC; Fearon, 2011). The most pathognomonic feature of mutations activating a gene product is recurrent occurrence in one or few codons. Such a signal is a strong indicator of selective value. Importantly, these types of changes also provide possibilities for therapeutic approaches through tackling the abnormal activity of the specific mutant protein product. Examples of successful utilisation of such a strategy include, but are not limited to, inhibitors of BRAF Val600Glu and activated EGFR (Lynch et al, 2004; Bollag et al, 2010). This success is in many cases marred by the fact that targeted monotherapies often fail to cure a patient because of pre-existing resistant subclones. Whereas the most common mutation hotspots may at present be largely known, any additional hotspots would provide a valuable platform for cancer biology in pinpointing relevant genes and domains to study, as well as for development of potentially curative combination therapies for stratified cancer care.

Colorectal cancer with microsatellite instability (MSI) defines a subgroup of CRCs with distinct clinical characteristics. These hypermutable CRCs may represent a sensitive system for generation, selection and detection of mutation hotspots. Indeed, in MSI CRC multiple well-established oncogenes, in particular BRAF, KRAS, PIK3CA and CTNNB1, are known to display very specific mutation hotspots (Fearon, 2011). We have previously mined exome-sequencing data of 25 MSI CRCs for hotspot point mutations (Gylfe et al, 2013). That study

*Correspondence: Dr LA Aaltonen; E-mail: lauri.aaltonen@helsinki.fi

Received 16 April 2014; revised 23 June 2014; accepted 8 July 2014; published online 12 August 2014

© 2014 Cancer Research UK. All rights reserved 0007 – 0920/14
did not consider identical base-specific mutations in the discovery set. The rationale for exclusion of this important mutation type was that such mutation signals, according to our early experience with exome data, typically arose from uncalled polymorphic germline variants. Expansion of control data as well as developments in analysis pipeline currently enable much more robust calling of somatic mutations. Thus, we have here taken the study forward to search for novel recurrent somatic cancer mutations hitting exactly the same base position.

MATERIALS AND METHODS

Sample material. The discovery set (Supplementary Table S1) was composed of normal and fresh-frozen tumour tissue DNA samples of 25 sporadic MSI patients gathered from a population-based series of 1042 colorectal adenocarcinoma patients described previously (Aaltonen et al, 1998, Salovaara et al, 2000). The validation set consisted of 167 Finnish samples (42 Lynch syndrome and 125 sporadic tumours, Supplementary Table S2) obtained from the above-mentioned cohort or from a subsequently collected series of additional 472 Finnish CRCs (unpublished data), and 87 Danish MSI CRCs (Supplementary Table S3). In addition, DNA from 10 CRC cell lines (VAC05, CCL231, GP5D, HCA7, HCT-116, LOVO, LS174T, RKO, SNUC2B and DLD1) was available. The study has been reviewed and approved by the Ethics Committee of the Hospital District of Helsinki and Uusimaa, Finland. Either a signed informed consent or authorisation from the National Supervisory for Welfare and Health was obtained for all the samples.

Exome and whole-genome sequencing. Exomic regions from 25 MSI N/T pairs were captured with the Agilent SureSelect all exon kit v1 (Agilent, Santa Clara, CA, USA). Paired-end short-read sequencing was carried out with Illumina Genome Analyzer II machines (Illumina, San Diego, CA, USA) at the Karolinska Institutet, or the Finnish Institute for Molecular Medicine (FIMM) Genome and Technology Center, Finland. Details on read mapping and variant calling are described in the previous publication (Gylfe et al, 2013). Whole-genome sequencing (WGS) on 91 MSS CRCs and the matched normal samples was performed on the Illumina HiSeq 2000 platform with paired-end reads 100 bp in length (median coverage of at least ×40). Sequencing data analysis is described in detail in Pitkanen et al (2014).

Identification and validation of the mutation hotspots. An in-house comparative analysis tool, RikuRator (Katainen, 2013, http://hdl.handle.net/10138/39539), was used to visualise and analyse the exome data. The criterion for the mutation hotspot was that the same base had to be mutated in at least two tumours. All the hotspot mutations identified in the discovery set were validated by Sanger sequencing (N/T pairs). The confirmed somatic hotspot mutations were further sequenced in an additional set of 254 tumours and somatic origin of the detected mutations was confirmed by sequencing the respective normal DNA samples. The novel hotspot mutations that became validated (14 genes) were then screened in 10 CRC cell lines. See Supplementary Data for details.

Computational methods. Ensembl Variant Effect Predictor, including SIFT (http://sift.jcvi.org/) and PolyPhen (http://genetics.bwh.harvard.edu/pph2/), was used to predict the functional consequences of the variants (http://www.ensembl.org). Multispecies protein sequence alignments were performed with ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/).

RESULTS

Exome-sequencing data of 25 sporadic MSI CRCs (Supplementary Table S1) were used as a discovery set to identify recurrent somatic missense mutations. Detailed information about the quality of the data can be found in our previous publication (Gylfe et al, 2013). Across all samples, 27879 somatic coding mutations were identified, of which 21167 were nonsynonymous and 5734 were synonymous (Figure 1). The most frequent single-nucleotide mutation types were C:G>TA (48%), C:G>AT (32%) and T:A>CG (11%) substitutions, with the rest of the substitutions C:G>GC (3.2%), T:A>AT (2.9%) and T:A>G (2.4%) comprising only less than 9% of all mutations. This is a typical mutation spectrum in MSI CRCs (Alexandrov et al, 2013). We next identified somatic mutations hitting exactly the same base in two or more cancers. This produced a list of 42 genes with possible heterozygous somatic hotspot mutations of which mutations in 36 genes were confirmed to be real as well as somatic. Among these were previously known hotspot mutations in BRAF, PIK3CA and CTNNB1 (Table 1). Novel somatic mutation hotspots were found in 33 genes (Supplementary Table S4). The list of validated hotspot mutations was dominated by transition changes at CpG dinucleotides (32/33, 97%). When compared with the exome data, C:G>TA mutation signature was overrepresented in the hotspots (97% vs 48%, Figure 2A). The distribution of the mutations in the discovery set tumours is presented in Figure 1.

Next, the 33 recurrent somatic missense mutations, as well as known hotspots in BRAF, KRAS, PIK3CA and CTNNB1, were screened in a validation set of 254 MSI CRCs (Supplementary Tables S2 and S3). Fourteen of the novel genes displayed hotspot mutations also in the validation set samples: ANTXR1, CEP135, CRYBB1, MORC2, SLC36A1, GALNT9, PI15, KRT82, CNTF, GLDC, MBTPS1, OR9Q2, R3HDM1 and TTPAL (Table 1, Supplementary Table S4). Twenty-eight (11%) of the validation set tumours displayed at least one of these hotspot mutations (Supplementary Figure S1). BRAF, KRAS, CTNNB1 and PIK3CA were mutated at frequencies of 42%, 15%, 6.9% and 5.5–11%, respectively (Table 1). The most frequent novel hotspot mutation encountered in the validation set was ANTXR1 Arg438Cys (7 out of 251, 2.8%; Figure 2A). The next frequent hotspots were CEP135 Arg1115Cys and CRYBB1 Ala171Thr (1.2%). Distribution of mutations in MORC2 differed somewhat from the other genes (3xSer25Leu, Lys27Met, Lys27Arg). Distribution of the mutations in the validation set tumours is presented in Supplementary Figure S1. Proposed functions of the 14 validated genes are summarised in Supplementary Table S5.

Multispecies protein alignments indicated that in 10 out of 14 genes (ANTXR1, MORC2, CEP135, GALNT9, PI15, KRT82, GLDC, MBTPS1, R3HDM1 and TTPAL) the targeted hotspot amino acids were highly conserved. Multispecies protein alignment for ANTXR1 is shown in Figure 2B. Functional predictions with PolyPhen2 and SIFT indicated damaging or deleterious effect for amino-acid substitutions in 11 genes (ANTXR1, MORC2, CEP135, GALNT9, KRT82, PI15, SLC36A1, GLDC, MBTPS1, R3HDM1 and TTPAL; Supplementary Table S4). We next examined the presence of the hotspot mutations in CRC cell lines and MSS CRCs. Sequencing of the 14 validated hotspots in 10 CRC cell lines revealed MORC2 Ser25Ala in HCA7 and OR9Q2 Ala249Thr in DLD1. To study MSS CRCs, we utilised WGS data on 91 MSS CRCs (unpublished data). We could not find any of the 33 hotspot mutations in MSS CRCs. When examining the complete coding sequence, mutations were found in ANTXR1, GALNT9, KRT82, R3HDM1, TTPAL, MTRMR12, NLRP10, DHX58, PTPR5 and KCNB1 (Supplementary Table S4). As ANTXR1 was the strongest hit in our MSI CRC mutation screen, we examined the two MSS CRC mutations in this gene, Arg250His and...
Pro430Leu, by Sanger sequencing and confirmed their somatic nature also with this method.

To examine the presence of the 33 hotspot mutations across cancer types and cell lines, cBioPortal (Cerami et al., 2012), International Cancer Genome Consortium Data Portal (Hudson et al., 2010) and COSMIC database (Forbes et al., 2011) were utilised. In CRCs, identical hotspot mutations were found in ANTXR1, OR9Q2, CTNNBP2, HEXIM2 and PALD1 (Supplementary Table S6). In MORC2, mutations Lys27Arg and Lys27Thr were found. In addition, in KCNB1 and NLRP10 examples of hotspot codon mutations were found, although the exact base changes were different from those seen in our discovery set data. Mutations in ANTXR1, OR9Q2, CRYBB1, MORC2, SLC36A1, KCNB1, SETD9, TMEM53, PTPRS, KRT82, PI15 and GALNT9 were observed in MSI CRCs. In addition, one TCGA CRC sample displayed ANTXR1 Arg439Gln, hitting adjacent to our hotspot. MSI CRC cell lines HCT-15 and HCT-116 displayed hotspot mutations in OR9Q2 and HEXIM2, respectively.

In extracolonic cancers, hotspot mutations in ANTXR1, CRYBB1, PALD1, SLC36A1, KCNB1, SETD9, TMEM53, PTPRS, KRT82, PI15 and GALNT9 were encountered (Supplementary Table S6). We specifically reviewed the TCGA data on endometrial and stomach MSI cancers. Among the 127 MSI-H endometrial cancers, one example of hotspot mutation in TMEM53 was found. Among 64 MSI-H stomach cancers, hotspot mutations in ANTXR1, CRYBB1 and KCNB1 were encountered, each once. Thus, although the hotspot mutations were not common, they occurred in a wide spectrum of different tumour types.

Finally, we examined expression of the ANTXR1 Arg438Cys mutant allele in five tumours, as well as the Pro430Leu and Arg250His from MSS CRCs. The mutant allele was absent in all seven cDNAs. Four tumours with exon 1 variant rs28365986 (Arg7Lys) expressed both alleles.

**DISCUSSION**

Comprehensive knowledge on mutations underlying human cancers is essential for improved cancer diagnostics, personalised care and novel therapeutic interventions. Most cancer genes that...
are mutated at high frequency have already been discovered; however, the comprehensive catalogue of mutations occurring at intermediate (2–20%) or low frequencies is far from complete (Lawrence et al., 2014). Oncogenes are often found to be recurrently mutated at preferred amino-acid-coding positions, and it has been suggested that the pattern rather than the frequency should be used as criteria for defining new driver genes (Vogelstein et al., 2013). Mutation hotspots are markers of selection in tumorigenesis, as well as attractive therapeutic targets. Here, we aimed at identifying new oncogenic drivers of cancer by examining MSI CRCs as a discovery set. These targets might be ideal for detection of novel oncogenic mutations, by forming a sensitive model system generating frequent mutations for possible selection during tumour progression.

We identified 33 candidate oncogenes of which 14 displayed hotspot mutations in a validation set. ANTXR1 (also known as TEM8) stood out as the most frequent target for hotspot mutations (2.8%). ANTXR1 encodes a single-pass cell-surface protein that was originally identified based on its overexpression in the tumour-infiltrating vasculature in human CRC (St Croix et al., 2000). It has been shown to have an important role in tumour angiogenesis (Chaudhary et al., 2012) and has an ability to promote cell migration by linking the extracellular matrix to the actin cytoskeleton (Werner et al., 2006). We found no expression of the ANTXR1 mutant allele in the seven primary tumours examined. The combination of selection of specific mutations and subsequent loss of expression of the mutant allele, as observed in the resulting cancer, is unusual. Whereas lack of expression of the mutant allele in the mature tumours may make the change unattractive as a therapeutic target, the exact mechanisms of tumorigenesis involved should be a particularly intriguing subject of cancer biology research.

In addition to CRC, the newly discovered hotspot mutations appear to have a role in the genesis of a wide variety of cancer types. Hotspot mutations were encountered in many extra-colonic cancers, albeit with low frequency. Whereas some of the common cancer hotspot mutations have been successfully utilised as targets for therapy, targeted monotherapy often fails to cure the patient due to pre-existing resistant subclones (Diaz et al., 2012). The catalogue of mutation hotspots reported in this study should serve as a valuable resource for focused and fruitful cancer biology work, to pave the way towards identification of multiple druggable targets—and combination treatment

### Table 1. Genes with validated somatic hotspot mutations

| Gene  | Genome position       | cDNA change | Protein change | n out of 25 | Finnish n = 167 (%) | Danish n = 87 (%) | Total n = 254 | COSMIC |
|-------|-----------------------|-------------|----------------|-------------|---------------------|------------------|--------------|--------|
| BRAF  | chr7:140453136        | c.17997>T   | Val600Glu      | 8           | 32                  | 68/167 (41)      | 36/82 (44)   | 104/249 | 42     |
| KRAS  | chr12:25398285-4, 25398282 | c.34G>S, c.37G>T | Gly12/Gly13App | 1           | 4                    | 25/165 (15)      | 13/82 (16)   | 38/247  | 15     |
| PIK3CA ex20 | chr17:38652085      | c.3140A>G   | His1047Arg     | 2           | 8                    | 15/139 (11)      | 7/68 (10)    | 22/207  | 11     |
| PIK3CA ex9  | chr17:38936082-3/178936091-2/178936094-6 | c.1624G–5A/c.1633G–4A/c.1636C–8G | Glu542/Glu545/Glu546 | 3  | 12                    | 8/137 (6)        | 4/80 (5)     | 12/217  | 5.5    |
| CTNNB1 | chr3:41266124         | c.121A>G    | Thr41Ala       | 2           | 8                    | 11/140 (7.9)     | 4/76 (5.3)   | 15/216  | 6.9    |
| ANTXR1 | chr2:69409751         | c.1312C>T   | Arg438Cys      | 3           | 12                   | 4/165 (2.4)      | 3/86 (3.5)   | 7/251   | 2.8    |
| CEP135 | chr4:56890689         | c.3343C>T   | Arg1115Cys     | 2           | 8                    | 2/165 (1.2)      | 1/80 (1.3)   | 3/245   | 1.2    |
| CRYBB1 | chr20:26979707        | c.511G>A    | Ala171Thr      | 2           | 8                    | 1/163 (0.6)      | 2/85 (2.4)   | 3/248   | 1.2    |
| MORC2  | chr22:31345795/31345789 | c.74C>T/c.80A>T | Ser25Leu/Lys27Met | 3b     | 12                    | 2/165 (1.2)      | 0/81 (0)     | 2/244   | 0.8    |
| SLC36A1 | chr5:15084477         | c.406G>A    | Ala136Thr      | 3           | 12                   | 2/166 (1.2)      | 0/87 (0)     | 2/253   | 0.8    |
| GALNT9 | chr12:132688193       | c.22C>T     | Arg8Cys        | 2           | 8                    | 2/165 (1.2)      | 0/73 (0)     | 2/238   | 0.8    |
| PIK3CA ex9 | chr12:52788815        | c.1486C>T   | Arg496Cys      | 2           | 8                    | 0/165 (0)        | 1/87 (1.1)   | 1/252   | 0.4    |
| CNTF   | chr11:58390329        | c.104C>T    | Thr35Met       | 2           | 8                    | 0/164 (0)        | 1/87 (1.1)   | 1/251   | 0.4    |
| GLDC   | chr9:6556244          | c.2885G>A   | Arg92Gln       | 2           | 8                    | 0/164 (0)        | 1/85 (1.2)   | 1/249   | 0.4    |
| MBTS1  | chr16:84120995        | c.1102G>A   | Ala368Thr      | 2           | 8                    | 1/166 (0.6)      | 0/87 (0)     | 1/253   | 0.4    |
| ORM9Q2 | chr11:57958707        | c.745G>A    | Ala249Thr      | 2           | 8                    | 1/165 (0.6)      | 0/86 (0)     | 1/251   | 0.4    |
| R3HD1M1 | chr2:13639373         | c.887G>A    | Arg296Gln      | 2           | 8                    | 1/166 (0.6)      | 0/86 (0)     | 1/251   | 0.4    |
| TTPAL  | chr20:43109078        | c.439C>T    | Arg147Cys      | 2           | 8                    | 1/166 (0.6)      | 0/86 (0)     | 1/252   | 0.4    |

*Arg496Cys, Ser455Pro, Ser455His.  
2 x Ser25Leu, 1 x Lys27Met.  
3 x Arg8Leu, Ser25Leu.  
4 x Arg8Leu, Ser25Leu.  
5 x Ala105Thr.  
6 x Arg1115Cys.  
7 x Arg8Cys.
Figure 2. Schematic representation of functional domains and mutations of ANTXR1. (A) Mutation hotspot in ANTXR1 is in the anthrax-binding domain. Other mutations, Arg250His (MSS), Pro430Leu (MSS) and Arg480Cys (MSS) are all located in the same functional domain. (B) The hotspot amino acid is depicted with the black arrow in the multispecies comparison (ClustalW2) and is highly conserved between species.

ACKNOWLEDGEMENTS

We thank Inga-Lill Svedberg, Iina Vuoristo, Alison Ollikainen, Sini Nieminen, Marjo Rajalaako, Sirpa Soisalo, Kirsi Pylvääniemi and Tuula Lehtinen for technical assistance. This work was supported by grants from the Academy of Finland (Finnish Center of Excellence Program 2012–2017), the Finnish Cancer Society, the Sigrid Juselius Foundation and SYSCOL (an EU FP7 Collaborative Project). Personal grants were received from the Academy of Finland (no. 260370 to PV) and from the University of Helsinki (postdoctoral grant for ST).

CONFLICT OF INTEREST

The authors declare no conflict of interest.
Supplementary Information accompanies this paper on British Journal of Cancer website (http://www.nature.com/bjc)