

## ORIGINAL ARTICLE

Next-generation sequencing of endoscopic biopsies identifies *ARID1A* as a tumor-suppressor gene in Barrett's esophagusMM Streppe<sup>1,2,3</sup>, S Lata<sup>4</sup>, M DelaBastide<sup>4</sup>, EA Montgomery<sup>1</sup>, JS Wang<sup>5</sup>, MI Canto<sup>6</sup>, AM Macgregor-Das<sup>1,7</sup>, S Pai<sup>1</sup>, FHM Morsink<sup>3</sup>, GJ Offerhaus<sup>3</sup>, E Antoniou<sup>4</sup>, A Maitra<sup>1,8,9</sup> and WR McCombie<sup>4</sup>

The incidence of Barrett's esophagus (BE)-associated esophageal adenocarcinoma (EAC) is increasing. Next-generation sequencing (NGS) provides an unprecedented opportunity to uncover genomic alterations during BE pathogenesis and progression to EAC, but treatment-naïve surgical specimens are scarce. The objective of this study was to establish the feasibility of using widely available endoscopic mucosal biopsies for successful NGS, using samples obtained from a BE 'progressor'. Paired-end whole-genome NGS was performed on the Illumina platform using libraries generated from mucosal biopsies of normal squamous epithelium (NSE), BE and EAC obtained from a patient who progressed to adenocarcinoma during endoscopic surveillance. Selective validation studies, including Sanger sequencing, immunohistochemistry and functional assays, were performed to confirm the NGS findings. NGS identified somatic nonsense mutations of *AT-rich interactive domain 1A* (*SWI like*) (*ARID1A*) and *PPIE* and an additional 37 missense mutations in BE and/or EAC, which were confirmed by Sanger sequencing. *ARID1A* mutations were detected in 15% (3/20) high-grade dysplasia (HGD)/EAC patients. Immunohistochemistry performed on an independent archival cohort demonstrated *ARID1A* protein loss in 0% (0/76), 4.9% (2/40), 14.3% (4/28), 16.0% (8/50) and 12.2% (12/98) of NSE, BE, low-grade dysplasia, HGD and EAC tissues, respectively, and was inversely associated with nuclear p53 accumulation ( $P = 0.028$ ). Enhanced cell growth, proliferation and invasion were observed on *ARID1A* knockdown in EAC cells. In addition, genes downstream of *ARID1A* that potentially contribute to the *ARID1A* knockdown phenotype were identified. Our studies establish the feasibility of using mucosal biopsies for NGS, which should enable the comparative analysis of larger 'progressor' versus 'non-progressor' cohorts. Further, we identify *ARID1A* as a novel tumor-suppressor gene in BE pathogenesis, reiterating the importance of aberrant chromatin in the metaplasia-dysplasia sequence.

*Oncogene* (2014) 33, 347–357; doi:10.1038/onc.2012.586; published online 14 January 2013

**Keywords:** next-generation sequencing; endoscopic mucosal biopsies; Barrett's esophagus; esophageal adenocarcinoma; *ARID1A*

## INTRODUCTION

Esophageal cancer represents the seventh most frequent cancer-related cause of death in the United States.<sup>1</sup> The incidence of esophageal adenocarcinoma (EAC) has dramatically increased over the past decades. EAC arises from non-dysplastic Barrett's esophagus (BE) following a multistep progression through low-grade dysplasia (LGD) and high-grade dysplasia (HGD), culminating in invasive neoplasia.<sup>2</sup> BE is found in 1.6–6.8% of the general population, and risk factors include presence of gastrointestinal reflux disease, Caucasian race, male gender, obesity and smoking.<sup>3</sup> The annual risk of developing HGD or EAC is 0.26–0.77% among BE patients.<sup>4,5</sup>

One of the hallmarks of carcinomas like EAC is the accumulation of genetic abnormalities that mirror histological progression from dysplasia to cancer.<sup>6,7</sup> Prior studies have profiled global abnormalities of promoter methylation, transcriptomic aberrations and copy number alterations in the multistep progression of BE to EAC.<sup>8–14</sup> For instance, hypermethylation of the promoter regions

of *CDKN2A*, *RUNX3*, *HPP1*, *APC*, *TIMP-3* and *TERT*, deletions on chromosome 9p and 17p, abnormalities in DNA content, and presence of *CDKN2A* and *TP53* mutations may have some promise as predictors of EAC development.<sup>2,15–18</sup> Remarkably, besides the relatively high prevalence of *CDKN2A/p16* and *TP53* mutations during EAC development,<sup>16–18</sup> there is still little known about the genetic landscapes of BE and EAC. It is of particular interest to discover whether subsets of somatic mutations observed in EAC are already present in BE and if such mutations might segregate patients most likely to progress to HGD or EAC ('progressors'), from the overwhelming majority who will never do. The identification of 'progressor' mutations would not only serve as a biomarker for patient stratification, but also as potential 'actionable' mutations that might block progression.

The availability of next-generation sequencing (NGS) technologies has enabled interrogation of the genomes of human cancers at an unprecedented scale.<sup>19–22</sup> Nearly all of the previously published NGS studies have been restricted to invasive cancers,

<sup>1</sup>Department of Pathology, John Hopkins Medical Institutions, Baltimore, MD, USA; <sup>2</sup>Department of Gastroenterology and Hepatology, University Medical Center Utrecht, Utrecht, The Netherlands; <sup>3</sup>Department of Pathology, University Medical Center Utrecht, Utrecht, The Netherlands; <sup>4</sup>Stanley Institute for Cognitive Genomics, Cold Spring Harbor Laboratory, Woodbury, NY, USA; <sup>5</sup>Division of Gastroenterology, Department of Medicine, Washington University School of Medicine, St Louis, MO, USA; <sup>6</sup>Department of Medicine (Division of Gastroenterology and Hepatology), John Hopkins Medical Institutions, Baltimore, MD, USA; <sup>7</sup>Pathobiology Program, John Hopkins Medical Institutions, Baltimore, MD, USA; <sup>8</sup>Department of Oncology, John Hopkins Medical Institutions, Baltimore, MD, USA and <sup>9</sup>Department of Genetic Medicine, John Hopkins Medical Institution, Baltimore, MD, USA. Correspondence: Professor A Maitra, Departments of Pathology, Oncology and Genetic Medicine, Johns Hopkins Medical Institutions, CRBII Building, Room 341, 1550 Orleans Street, Baltimore, MD 21231, USA or Dr WR McCombie, Stanley Institute for Cognitive Genomics, Cold Spring Harbor Laboratory, Woodbury Building, 500 Sunnyside Boulevard, Woodbury, NY 11797, USA.

E-mail: amaitra1@jhmi.edu or mcombie@cshl.edu

Received 19 July 2012; revised 28 October 2012; accepted 29 October 2012; published online 14 January 2013

using either cell lines, xenografts or surgically resected carcinomas.<sup>20,23,24</sup> The application of NGS to microscopic precursor lesions is still in its infancy, although targeted capture and NGS of larger precursor lesions has now been shown to be feasible.<sup>25</sup> One of the challenges in the context of BE and EAC remains the availability of treatment-naïve surgical resection samples, as many patients with EAC receive neoadjuvant chemo-radiation therapy, or locally ablative treatments in the pre-operative period.<sup>26,27</sup> In contrast, endoscopic mucosal biopsies are widely available and readily banked by gastroenterologists as a component of endoscopic surveillance and/or clinical trials. Even in instances when endoscopic biopsies are obtained from patients who later develop advanced, surgically unresectable cancers, the pre-cancerous sample serves as an invaluable substrate for identifying potential 'progressor' mutations. We have previously demonstrated the utility of archived snap-frozen biopsies for generating the transcriptomic profiles of BE and EAC using serial analysis of gene expression, as well as for generating genome-wide copy number and methylome profiles using array-based approaches.<sup>10,28</sup>

In this study, we utilized endoscopic mucosal biopsies obtained from an EAC patient to generate the first genome-wide profiles of BE and EAC. Our study fulfills two important objectives of significance vis-à-vis BE pathogenesis and progression: first, from a technological standpoint, it establishes the feasibility of using endoscopic biopsies for successfully performing genomic profiling of BE, which should facilitate future studies, such as comparing the landscapes of 'progressors' versus 'non-progressors'. Second, our study identifies the *AT-rich interactive domain 1A (SWI like) (ARID1A)* gene as a novel tumor suppressor in BE. The protein encoded by *ARID1A* is a key component of the highly conserved switch/sucrose non-fermentable chromatin remodeling complex.<sup>29</sup> Using sequencing and immunohistochemical studies on archival specimens and functional assays in cell lines, we establish the frequency of *ARID1A* mutations and *ARID1A* loss in the multistep progression of EAC, and the consequences of inactivating gene function on the neoplastic phenotype. Finally, we determine the downstream effectors of *ARID1A* that are likely to contribute to the oncogenic phenotype caused by *ARID1A* downregulation.

## RESULTS

Whole-genome NGS was performed on normal squamous epithelium (NSE) (germ line control), BE and EAC tissues obtained from one individual during upper endoscopy. Single-nucleotide variants (SNVs) were identified using the Genome Analyzer Toolkit (GATK) and categorized into four tiers, as described by Mardis *et al.*<sup>30</sup> Table 1 contains SNVs in coding regions of annotated exons and splice sites (tier 1), and these variants have been validated by Sanger sequencing. SNVs in regulatory regions (tier 2), in non-repeat masked, non-regulatory regions (tier 3) and all remaining SNVs are listed in Supplementary Tables S1–S3, respectively. Among the validated tier 1 SNVs, 2 were called with low and 38 with high confidence by GATK. In addition, several tier 2/3/4 SNVs were validated by Sanger sequencing.

SNVs in coding regions of annotated exons and splice sites (tier 1) A total of 40 tier 1 SNVs were detected by NGS and confirmed by Sanger sequencing (Table 1). Among these, 37 caused a functional amino-acid change (missense), 1 impacted the 5' splice site and 2 were nonsense mutations. The nonsense mutations were both present in genes located on chromosome 1p, in the coding regions of *ARID1A* and *PP1E*, and were detected in BE and EAC. Among the missense SNVs were monoallelic mutations of *TP53*, *TFAP4* and *ITGB3*. Interestingly, the overwhelming majority of the tier 1 SNVs were found in both BE and EAC.

Many somatically altered genes in which we have found SNVs have either a direct or indirect interaction with a cancer pathway (Figure 1). Four mutated genes, *ARID1A*, *ITGB3*, *TP53* and *TFAP4* are among the most connected genes in this network. Other important tier 1 genes include transcription regulators (*IGF2BP2*, *LBX2*) and G-protein coupled receptors (*TASR1*, *GPR98*). The known tier 1 gene functions in cancer development and/or progression are listed in Supplementary Table S4.

### *ARID1A* is mutated in a subset of HGD/EAC samples

In addition to biopsies from our index patient, 14 frozen HGD/EAC and matched NSE biopsy samples, and 5 formalin-fixed paraffin-embedded EAC and matched tumor-negative lymph nodes (control) tissues were subjected to mutational analysis of *ARID1A* using Sanger sequencing. In total, one nonsense mutation (g.27056230C>A (index patient)), and two indel mutations (g.27099353\_27099354delCCinsAA and g.27023892delC) were detected in 20 patients (3/20 = 15%). Both indel mutations resulted in the reading of a premature stop codon.

### *ARID1A* protein loss occurs early in the EAC carcinogenesis

*ARID1A* protein expression was examined by immunohistochemistry in an independent cohort of 98 EAC patients (Figure 2). Nuclear staining patterns were categorized into two groups; present and lost expression. In some lesions exhibiting *ARID1A* loss, the loss was only observed in one clone of the lesion. *ARID1A* loss was observed in at least one lesion of fourteen EAC patients (14/98, 14.3%). Detailed information regarding in which lesion(s) of these patients *ARID1A* loss occurred is given in Figure 3a. Overall, *ARID1A* loss occurred in 0% (0/76), 4.9% (2/40), 14.3% (4/28), 16.0% (8/50), 12.2% (12/98) and 6.5% (2/31) of NSE, BE, LGD, HGD, EAC and lymph node metastasis tissues, respectively (Figure 3b).

Nuclear accumulation of p53 is significantly less common in the presence of concomitant *ARID1A* loss

The distribution of age at diagnosis, gender and race were not statistically significant different between *ARID1A*-negative and -positive cases. *ARID1A*-negative cases lived by average 3.7 months longer ( $P = 0.608$ ). No correlations between *ARID1A* status and tumor features were found (Supplementary Table S5).

As it has been suggested that mutations in *ARID1A* and *TP53* are mutually exclusive in cancer, and that the former might substitute for loss of p53 function,<sup>31,32</sup> we examined the status of p53 in our cohort, using nuclear accumulation as readout for genetic abnormality (Figures 2 and 3c). We observed that specimens with *ARID1A* loss demonstrated significantly less frequent nuclear p53 accumulation ( $P = 0.028$ , Supplementary Table S5). Among the 12 patients with *ARID1A* loss in EAC, 41.7% (5/12) exhibited nuclear p53 accumulation versus 72.9% (62/85) of the *ARID1A* retained cases (Figure 3a).

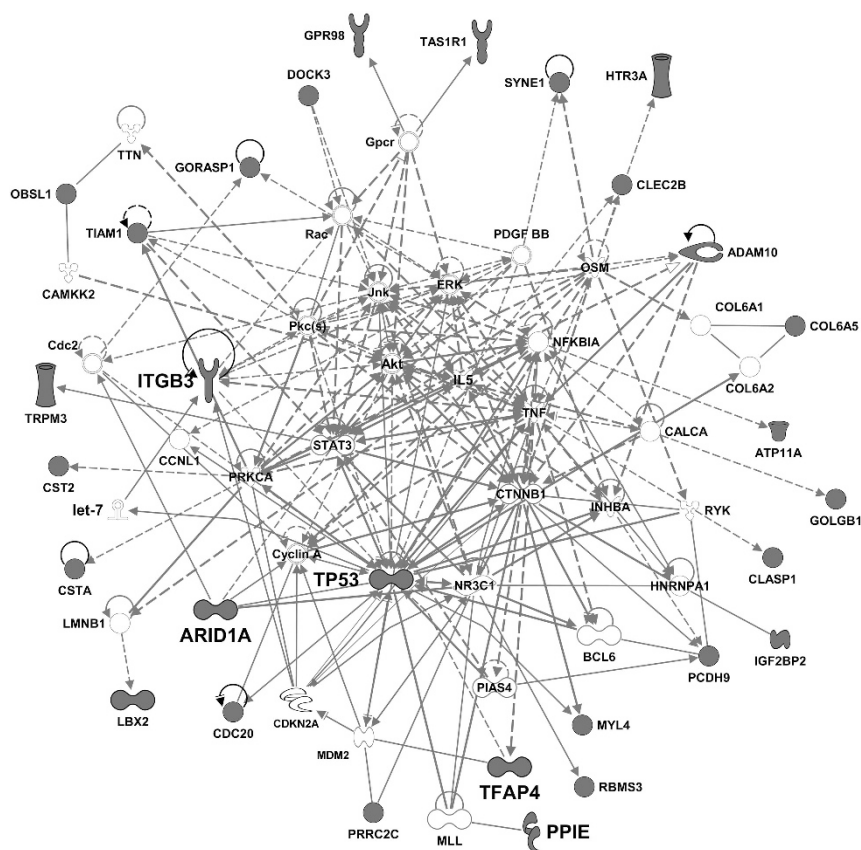
### *ARID1A* regulates cell growth/proliferation and invasion of EAC cells

*ARID1A* was knocked down (*ARID1A* KD) in OE33 cells by transfecting *ARID1A* small interfering RNA into the cells, and knockdown efficiencies of 75% and 81% at mRNA and protein level, respectively, were confirmed (Figures 4a and f). Direct visualization of *ARID1A* KD in OE33 cells was established by immunocytochemistry (Supplementary Figure S1D, E). Loss of *ARID1A* was associated with significantly increased cell growth ( $P < 0.001$ ; Figure 4b). This increased growth potential was validated by a 5-bromo-2-deoxyuridine (BrdU) proliferation assay, which confirmed significantly increased proliferation as the underlying mechanism ( $P < 0.001$ ; Figures 4d and e, Supplementary Figure S1C). No significant differences in migratory

**Table 1.** Tier 1 single-nucleotide variants

Chromosome	Genomic position	Reference base	SNV	Gene	Existing variant	Genotype	Impact	Phast Con	GERP	SIFT	PPH2	Mut assessor	Condel score	Condel table
1	1904401	G	A	KIAA1751		BE + EAC	Missense	0.167	2.11	0.07	0.939		0.937	Deleterious
1	6634991	G	A	TAS1R1		EAC	Missense	0.006	2.29	0.68	0		0.888	Deleterious
1	27056230	C	A	ARID1A		BE + EAC	Nonsense	0.998	5.81					
1	40207046	C	T	PPIE		BE + EAC	Nonsense	1	1.54					
1	43826412	G	A	CDC20		BE + EAC	Missense	0.992	4.33	0.06	0.727	3.185	0.83	Deleterious
1	171510723	G	A	PRRC2C		BE + EAC	Missense	1	5.1		0			
1	176664951	G	A	PAPPA2		BE + EAC	Missense	0.802	5.16	0.15	0.999		0.979	Deleterious
2	74725151	A	T	LBX2		BE + EAC	Missense	0.602	2.54	0.17	0.952		0.934	Deleterious
2	122125361	C	T	CLASP1		BE + EAC	Missense	0.853	4.15	0.08	0.081		0.584	Deleterious
2	220433041	G	A	OBSL1	rs117926118	BE + EAC	Missense	1	4.97	0.05	0.999		0.978	Deleterious
2	237405897	C	T	IQCA1		BE + EAC	Missense	0.9	5.16	0	1		1	Deleterious
3	29938905	G	A	RBM53		BE + EAC	Missense	1	4.86	0.18	1		0.979	Deleterious
3	39144218	A	G	GORASP1		BE + EAC	Missense	1	4.62	0	0.984		0.992	Deleterious
3	51400025	C	A	DOCK3		BE + EAC	Missense	1	5.21	0.02	0.078		0.842	Deleterious
3	121383392	G	T	GOLGB1		BE + EAC	Missense	0.946	3.28	0	0.993		0.997	Deleterious
3	122060293	C	T	CSTA		BE + EAC	Missense	0	-2.78	1	0		0	Neutral
3	130159097	G	A	COL6A5		BE + EAC	Missense	0	-7.43	0.37	0.21		0.31	Neutral
3	185376142	T	G	IGF2BP2		BE + EAC	Missense	0.998	4.21	0.06	0.233		0.781	Deleterious
5	89941786	C	T	GPR98		BE + EAC	Missense	1	5.1	0.01	0.44		0.862	Deleterious
6	152532698	C	T	SYNE1		BE + EAC	Missense	1	5.57	0.999	0.999		0.999	Deleterious
7	117024830	C	T	ASZ1		BE + EAC	Missense	1	4.34	0.05	0.184		0.794	Deleterious
9	73736231	C	A	TRPM3		BE + EAC	Missense	0.997	4.24	0	0.025		0.907	Deleterious
10	72541595	C	T	C10orf27	rs115775127	BE	Missense	1	4.84	0	0.999		0.999	Deleterious
11	113857660	C	T	HTR3A		BE + EAC	Missense	0.759	NA	0	0.976		0.988	Deleterious
12	10005934	C	T	CLEC2B		BE + EAC	Missense	0.948	2.89	0.42	0.907		0.888	Deleterious
13	67799896	C	T	PCDH9		BE + EAC	Missense	0.949	5.43	0.53	0.001		0.294	Neutral
13	113516819	T	A	ATP11A		BE + EAC	Missense	0.049	4.86	0.49	0		0.512	Deleterious
15	58974427	T	A	ADAM10		BE + EAC	Missense	1	5.03	0	0.999		0.999	Deleterious
15	83935682	C	T	BNC1		BE + EAC	Missense	1	5.08	0	0.999		0.999	Deleterious
15	86838559	C	T	AGBL1		BE	Missense	0.955	5.28	0	0.999		0.999	Deleterious
16	4312620	G	A	TFAP4		BE + EAC	Missense	1	3.64	0	0		1	Deleterious
17	7578526	C	A	TP53	COSM10647	BE + EAC	Missense	1	5.13	0	0.998	3.885	0.999	Deleterious
17	45299814	A	G	MYL4	rs16941677	BE + EAC	Missense	1	4.69	0.06	0.082		0.793	Deleterious
17	45384961	C	A	ITGB3		BE + EAC	Missense	1	3.06	0.01	0.987		0.988	Deleterious
19	15688320	G	A	LOC100287986		BE + EAC	Missense	1	2.32					
19	11618576	G	A	ECST		BE + EAC	Missense	0.045	3.04	0	1	NA	1	Deleterious
19	40710449	C	T	LOC100289661		BE + EAC	Missense	0.994	-3.28					
20	23807069	C	T	CST2		BE + EAC	Splice5	0.484	1.43					
20	42088667	T	G	LOC100287002		BE + EAC	Missense	0.807	2.35	0	0.999	2.275	0.859	Deleterious
21	32639065	G	A	TIAM1		BE + EAC	Missense	1	4.79	0	0.999		0.999	Deleterious

Sanger validated SNVs in coding sequences or splice junctions. The SNVs are either EAC-specific, BE-specific or shared between BE and EAC. We indicated the most damaging Condel annotation across all the transcripts. Abbreviations: ARID1A, AT-rich interactive domain 1A (SWI like); BE, Barrett's esophagus; EAC, esophageal adenocarcinoma; GERP, genomic evolutionary rate profiling; NA, not applicable; PPH2, polymorphism phenotyping v2; SNV, single-nucleotide variant; SIFT, sorting intolerant from tolerant.



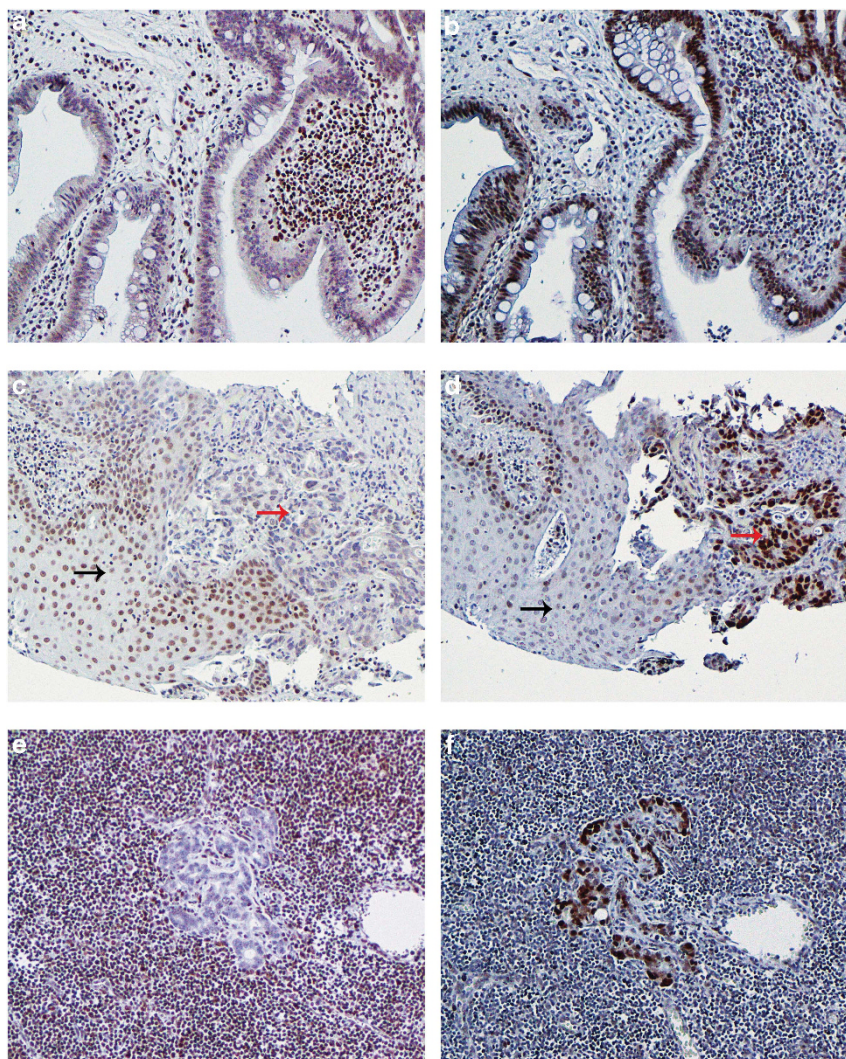
or clonogenic potential were observed between *ARID1A* KD and mock-transfected (mock) OE33 cells (Supplementary Figures S1A and S1B). However, a phenotype of significantly increased invasive potential (210%,  $P < 0.001$ ) was revealed in *ARID1A* KD compared with mock OE33 cells (Figure 4c).

To identify genes that are aberrantly expressed on ARID1A inactivation, we performed Affymetrix microarray on ARID1A KD and mock OE33 cells. Ninety-four genes (52 upregulated and 42 downregulated) with an average fold change  $>1.75$  were recognized as differentially expressed on ARID1A KD (Supplementary Figure S2, Supplementary Table S6). As we had observed a phenotype of significantly increased cell growth, proliferation and invasion on ARID1A loss in OE33 cells, we queried publicly available databases of gene function to identify putative candidate genes that have the ability to control these cellular processes. We selected 16 coding genes, which met these criteria; *HPDG*, *CYP1B1*, *FGFBP1*, *S100A4*, *CEACAM5*, *PIM1*, *ACPP*, *BIRC3*, *CDKN1C* and *NRG1* were upregulated in ARID1A KD OE33 cells, whereas *MT2A*, *SDHD*, *ATG5*, *RSG16*, *GSPT1* and *RHOB* were downregulated (Supplementary Figure S2, Supplementary Table S7). Expression levels of these genes were validated by quantitative reverse transcriptase-PCR in four biological replicate experiments. Specifically, transcripts corresponding to *HPDG*, *MT2A*, *CYP1B1*, *SDHD*, *FGFBP1*, *S100A4*, *CEACAM5*, *ACPP*, *ATG5*, *RSG16* and *NRG1* were statistically significant aberrantly expressed in ARID1A KD (Figure 4f, Supplementary Table S7).

In this study, we show that it is feasible to perform high depth NGS on snap-frozen endoscopic mucosal biopsies. To the best of our knowledge, this is the first example of whole-genome sequencing of matched precursor and invasive carcinoma obtained from a single patient. We found that the mutational profiles of BE and EAC in our NGS patient are remarkably similar, suggesting that genetic alterations in the metaplasia–carcinoma sequence of BE might occur much earlier than the histological changes of frank dysplasia. Consistent with this, Wang *et al.*<sup>33</sup> have previously shown that the transcriptomic profiles of BE and EAC are comparable, and our own group has shown that global hypomethylation observed in EAC actually occurs at the stage of BE itself.<sup>10</sup> Future studies utilizing NGS on endoscopic mucosal biopsies of progressors versus non-progressors should enable us to compare the genetic landscapes of these two BE groups.

Phenotypic heterogeneity of precursor lesions and cancers, and in particular the lack of sensitive sequencing techniques that are capable of detecting genetic aberrations in subclones of neoplastic cells, have delayed the development of accurate risk stratification tools for BE. The ‘cancer stem cell’ hypothesis and genetic heterogeneity potentially underlie phenotypic heterogeneity.<sup>34–36</sup> Studies have shown that only sub-populations of cancer cells, so-called ‘cancer stem cells’, have unlimited proliferative and clonogenic potential, and that these cells are most likely to initiate metastases formation.<sup>35</sup> Genetic heterogeneity is not limited to cancers from different cell origin and anatomic sites, but is also seen among patients with a particular cancer type, and even within individual tumors.<sup>34,36</sup> Many solid cancers exhibit multiple subclones, and although these





**Figure 2.** Immunohistochemistry for ARID1A (left panels) and p53 (right panels) in representative cores of BE (a, b, 20 ×), EAC (c, d, 20 ×) and lymph node metastasis (LNM; e, f, 20 ×) lesions. In all lesions, nuclear ARID1A expression is completely lost, whereas there is nuclear accumulation of p53. The NSE in image (c, d, black arrows) served as control for staining as ARID1A is present and p53 is absent. The red arrows (c, d) indicate EAC.

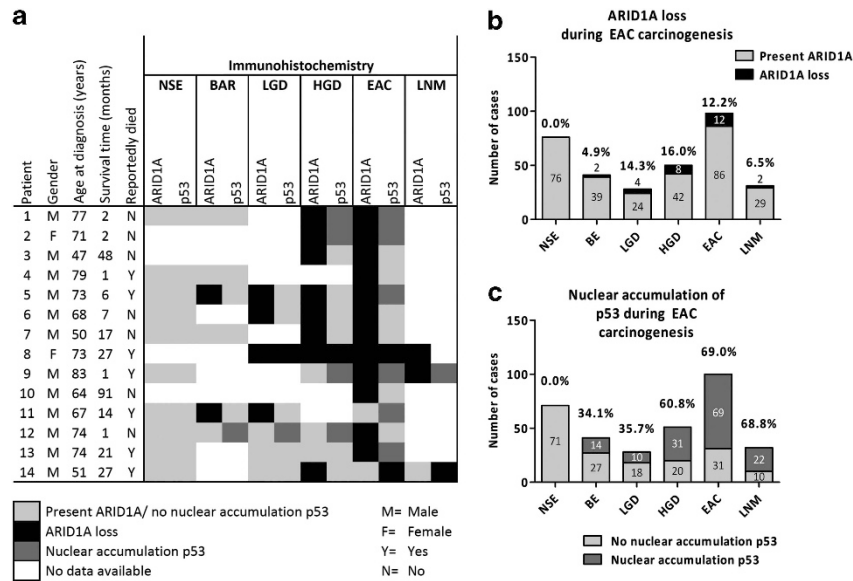
clones probably have the same driver mutations, they undergo clonal evolution acquiring additional (epi)genetic aberrations.<sup>36</sup> The expansion of the most dominant subclone, as shown in breast cancer, is likely to result in clinical symptoms and finally diagnosis.<sup>37</sup> However, as a consequence of ongoing lineage formation and branching, each subclone individually may be capable of initiating metastasis formation.<sup>34,36</sup> NGS might enable us to detect alterations in subclones of BE and EAC that are specific to neoplastic and metastatic clones, respectively, and may result into risk stratification tools and drug discoveries.

In terms of the specifics of our genomic analysis, we observed somatic nonsense mutations of *PPIE* and *ARID1A*, both located on chromosome 1p. Inactivating mutations in *ARID1A* have been detected in several solid cancers including digestive tract cancers (Table 2, summary of *ARID1A* mutations in solid carcinomas).<sup>20,38,39</sup> However, *ARID1A* mutations had not been reported in EAC till recently; while our manuscript was under review, a NGS study reported a nonsense mutation in *ARID1A* in 1 out of the 11 EAC patients (a summary of frequent exomic mutations identified in this and our NGS study is provided in Table 3).<sup>40</sup> We detected an *ARID1A* nonsense mutation in our index patient, as well as indel

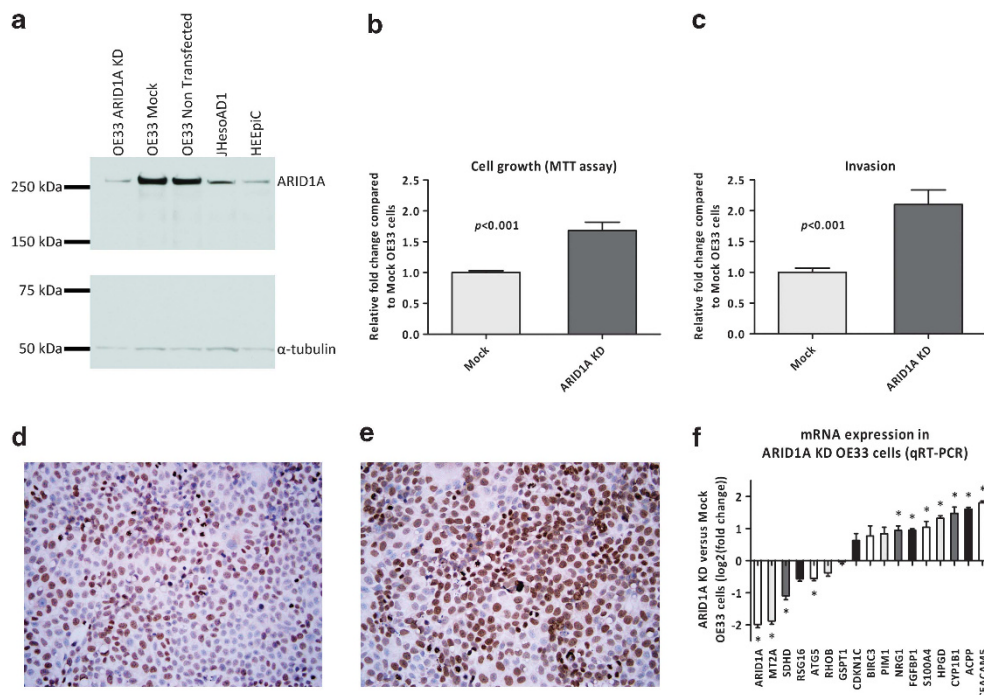
mutations in two additional patients, suggesting that the mutational rate of *ARID1A* in EAC is ~15% (3/20).

We further found that *ARID1A* loss occurs in 14.3% (14/98) of the EAC patients. We examined several stages of the metaplasia–carcinoma sequence in these patients and concluded that *ARID1A* is lost early in the carcinogenesis of EAC. As studies have reported that *ARID1A* mutations are negatively associated with *TP53* mutations,<sup>31,32,41</sup> we determined the p53 status in our BE progression cohort. We observed that specimens with *ARID1A* loss demonstrated significantly less frequent nuclear accumulation of p53. The fact that our index patient harbored mutations in both *ARID1A* and *TP53*, also suggest that mutations in these genes are not always mutually exclusive.

Our *in vitro* studies suggested that *ARID1A* has an important role in the regulation of cell growth and invasion, which implicates *ARID1A* is a tumor-suppressor gene in EAC. Proliferation assays performed in four gastric adenocarcinoma cell lines by Zang *et al.*<sup>41</sup>, also indicated that *ARID1A* has the capacity to inhibit cell proliferation in the stomach. Global expression profiling identified novel downstream genes of *ARID1A*, which are all involved in regulating cell proliferation, survival and/or invasion. Among the



**Figure 3.** (a) Immunohistochemistry results for lesions in which complete ARID1A loss was detected. Nuclear accumulation of p53 was observed in 5/12 EAC ARID1A loss patients. ARID1A loss was observed in 0% (0/76), 4.9% (2/41), 14.3% (4/28), 16% (8/50), 12.2% (12/98) and 6.5% (2/31) of the NSE, BE, LGD, HGD, EAC and lymph node metastasis (LNM), respectively (b). There was a stepwise increase in presence of nuclear accumulation of p53 seen during neoplastic progression of BE (c).



**Figure 4.** Robust ARID1A expression was detected at the expected size in OE33, JHesAD1 and HEEpC cells. ARID1A protein was efficiently knocked down (81% compared with mock OE33) in OE33 cells (a). A significant increase in cell growth (b) and promotion of invasion (c) was observed in the ARID1A KD OE33 cells. The brown staining in photo (d, 20 $\times$ ) and (e, 20 $\times$ ), which are example photos, depicts incorporation of BrdU in mock and ARID1A KD cells, respectively. Overall, 48.8% and 67.9% of the mock and ARID1A KD cells, respectively, were proliferative 36–48 h after transfection ( $P < 0.001$ ). (f) Shows quantitative reverse transcriptase (qRT)-PCR validation results (four biological replicates) of the microarray experiment. The error bar in the graph indicates the standard error of the mean, whereas the asterisk depicts a  $P < 0.05$  between ARID1A KD and mock OE33 cells.

upregulated genes in ARID1A KD OE33 cells are *CYP1B1*, *S100A4* and *CEACAM5*. *CYP1B1* is overexpressed in multiple carcinomas, and has an important role in tumor formation.<sup>42,43</sup> *In vitro* experiments performed by Martinez et al.<sup>44</sup> suggest that *CYP1B1* expression is correlated with docetaxel-resistance and increased

cell survival in breast cancer cells.<sup>43,44</sup> *S100A4* is a member of the S100 family of calcium-binding proteins, and *S100A4* upregulation leads to enhanced cell proliferation. It furthermore regulates cell cycle progression, apoptosis, and promotes tumor invasion and metastases formation.<sup>45,46</sup> *S100A4* overexpression has been



**Table 2.** ARID1A mutations in solid carcinomas

Cancer type	Mutational rate	Reference(s)
Breast cancer	3.2% (3/95) 3.5% (4/114) 0% (0/82)	Cornen <i>et al.</i> <sup>63</sup> Jones <i>et al.</i> <sup>38</sup> Mamo <i>et al.</i> <sup>64</sup>
Colorectal carcinoma	10.1% (12/119)	Jones <i>et al.</i> <sup>38</sup>
Esophageal adenocarcinoma	9.1% (1/11) 15% (3/20)	Agrawal <i>et al.</i> <sup>40</sup> This paper
Gastric adenocarcinoma	10% (10/100) 27.2% (6/22) 8% (9/110)	Jones <i>et al.</i> <sup>38</sup> Wang <i>et al.</i> <sup>32</sup> Zang <i>et al.</i> <sup>41</sup>
Hepatocellular carcinoma	10% (12/120)	Fujimoto <i>et al.</i> <sup>65</sup>
High-grade serous ovarian carcinomas	0% (0/76)	Wiegand <i>et al.</i> <sup>39</sup>
Lung carcinoma	5.6% (2/36)	Jones <i>et al.</i> <sup>38</sup>
Ovarian clear cell carcinoma	57% (24/42) 46.2% (55/119)	Jones <i>et al.</i> <sup>20</sup> Wiegand <i>et al.</i> <sup>39</sup>
Ovarian endometrioid carcinoma	30.3% (10/33)	Wiegand <i>et al.</i> <sup>39</sup>
Pancreatic ductal adenocarcinoma	6.9% (2/29) 8.4% (12/119)	Birnbaum <i>et al.</i> <sup>66</sup> Jones <i>et al.</i> <sup>38</sup>
Prostate carcinoma	8.7% (2/23)	Jones <i>et al.</i> <sup>38</sup>
Transitional cell carcinoma of the bladder	18.6% (18/97)	Gui <i>et al.</i> <sup>67</sup>

Abbreviation: ARID1A, AT-rich interactive domain 1A (SWI like).

**Table 3.** Summary frequent mutations in EAC discovered by NGS

Gene/gene complex	Mutation rate	Type of mutation (absolute amount)				Reference(s)
		Nonsense	Indel	Missense	Splice site	
APC	16.7%	1	1	0	0	Agrawal <i>et al.</i> <sup>40</sup>
ARID1A	16.7%	2	0	0	0	Agrawal <i>et al.</i> <sup>40</sup> , this paper
BRSK1	16.7%	0	1	1	0	Agrawal <i>et al.</i> <sup>40</sup>
FBN3	16.7%	1	0	1	0	Agrawal <i>et al.</i> <sup>40</sup>
MUC16	16.7%	1	0	1	0	Agrawal <i>et al.</i> <sup>40</sup>
HMCN1	25.0%	0	0	3	0	Agrawal <i>et al.</i> <sup>40</sup>
HTR1A	25.0%	0	0	3	0	Agrawal <i>et al.</i> <sup>40</sup>
ANK2	25.0%	0	0	2	1	Agrawal <i>et al.</i> <sup>40</sup>
SYNE1	33.3%	0	0	4	0	Agrawal <i>et al.</i> <sup>40</sup> , this paper
KIF2B/CLASP1	41.7%	0	0	5 <sup>a</sup>	0	Agrawal <i>et al.</i> <sup>40</sup> , this paper
TP53	75.0%	2	1	3	3	Agrawal <i>et al.</i> <sup>40</sup> , this paper

An overview of prevalent mutations in EAC identified using NGS by Agrawal *et al.*<sup>40</sup> and our group. In total, 12 EAC and matched NSE tissue samples were utilized for NGS. Genes in which nonsense and indel mutations were detected in at least 2/12 (16.7%) patients as well as genes in which missense or splice site mutations occurred in >3 cases (25%) are listed in this table. Abbreviations: ARID1A, AT-rich interactive domain 1A (SWI like); EAC, esophageal adenocarcinoma; NGS, next-generation sequencing; NSE, normal squamous epithelium of the esophagus. <sup>a</sup>CLASP1 mutation was found in only one patient.

determined in 67% of the EACs, and was associated with the presence of lymph node metastases.<sup>47</sup> The third significantly upregulated gene is *CEACAM5*, better known as CEA, and is commonly used as a serum marker for colorectal adenocarcinomas. *CEACAM5* is upregulated in nearly all colorectal adenocarcinomas, and 60% and 4.6% of the EAC and BE tissues, respectively.<sup>48,49</sup> It has been extensively studied *in vitro*, and among others inhibits cell differentiation and anoikis.<sup>50,51</sup>

*PPIE* mutations have not been reported before in EAC. Functionally, *PPIE* has been shown to directly bind to the third PHD finger of *mixed lineage leukemia 1* (*MLL1*), which results into a switch in *MLL1* function from transcriptional activator to repressor.<sup>52</sup> Overexpression of *PPIE* results in repression of oncogenic *homeobox* (*HOX*) gene expression,<sup>52,53</sup> whereas mutant *PPIE* is not able to bind to *MLL1* and downregulate *HOX* genes.<sup>53</sup> In addition, *PPIE* is capable of downregulating the *MLL1* target genes *CDKN1B* and *C-MYC*.<sup>54</sup> A nonsense mutation in *PPIE* combined with deletion or epigenetic silencing of the second allele would be postulated to cause complete inactivation of *PPIE*,

leading to a dysfunctional switch between activation and repression of *MLL1* target genes, which might promote cancer formation and/or progression.

In conclusion, our studies established the feasibility of using widely available and banked endoscopic biopsies for NGS, which should facilitate similar studies in larger cohorts of samples. Furthermore, we identified *ARID1A* as a novel tumor-suppressor gene in the BE-associated EAC sequence, reiterating the importance of aberrant chromatin in this process.

## MATERIALS AND METHODS

### Next-generation sequencing

Libraries for NGS were generated using DNA obtained from archived snap-frozen endoscopic mucosal biopsies from a treatment-naïve 51-year-old Caucasian male with a pT2NxMx EAC. Libraries were prepared from three independent samples obtained during upper endoscopy: NSE (germ line control), BE and EAC. Reference samples were taken directly adjacent to where the research biopsies had been obtained. Two expert gastrointestinal pathologists (EAM and AM) confirmed that the lesion biopsies

were pure (no mixtures of neoplastic, preneoplastic and normal cells were observed) and contained a cellularity of >80%. Total DNA was extracted from the frozen biopsies using the DNeasy kit (Qiagen, Germantown, MD, USA) according to the manufacturer's manual. The patient subsequently received neoadjuvant chemo-radiation therapy, and underwent surgical resection, which demonstrated residual foci of moderately differentiated EAC. He remains disease-free 3 years following the index biopsies.

### Library preparation NGS

One milligram of genomic DNA was fragmented using Covaris S2 (Caliper Inc., Waltham, MA, USA). Sequencing libraries were constructed using the NEBNext DNA library preparation kit (New England Biolabs, Ipswich, MA, USA) according to the manufacturer's instructions. DNA was size-selected to an average size of 560 bp, which corresponds to an average insert size of 300 bp. Libraries were quantified on an Agilent bioanalyzer (Agilent, Santa Clara, CA, USA) using a DNA 1000 chip, and sequencing flow cells were prepared using a CBOT (Illumina Inc., San Diego, CA, USA). NGS was performed on Illumina HiSeq 2000 (paired-end 100-bp runs) and on Illumina GAII-X (paired-end 150-bp runs).

### Alignment

Illumina paired-end reads (ranging from 101 to 151 bp) of EAC, BE and NSE were aligned separately to the human NCBI Build 37 reference sequence using Novoalign software ([www.novocraft.com](http://www.novocraft.com)). The aligned sequence files were sorted and merged using SAMtools.<sup>55,56</sup> Picard (<http://picard.sourceforge.net>) was used to remove PCR duplicates from merged bam files. The average genome coverage for EAC, BE and NSE samples were estimated to be 56.2X, 50.3X and 37.9X, respectively. GATK was used for base quality score recalibration, and SNV discovery across was performed for all three samples.<sup>57</sup> Relaxed filtering parameters (clusterWindowSize 10 clusterSize 3 min confidence score filter of 50) were used to label the SNVs as high or low confidence.

### Identification of SNVs in EAC and BE

For initial SNV calling, minimal filtering for quality was used to capture as many variants as possible at the initial stage. More subsequent filtering and validation was done, as described below, on the SNVs annotated in or near coding regions. Using GATK, 4090303 SNVs in the EAC, and 4072159 SNVs in the BE tissue were identified. As tissue samples are heterogeneous, relaxed GATK filters were used to minimize the number of false negatives. Subsequently, the SNVs were filtered by comparing the SNVs detected in BE and EAC with those identified in NSE. A total of 3 913 704 of the EAC SNVs and 3 904 615 of the BE SNVs were also detected in the NSE sample, suggesting that these are the patient's inherited SNVs. The remaining 176 599 EAC SNVs and 167 544 BE SNVs were annotated using SeattleSeq (<http://snp.gs.washington.edu/SeattleSeqAnnotation131/>, accessed on 19 July 2011). 7280 EAC-specific SNVs and 6566 BE-specific SNVs were located in the coding sequences of genes (missense, nonsense, splice3 and splice5), in regulatory regions (UTR3, UTR5, near gene 3 (2 kbp or less upstream of coding sequences), or near gene 5 (2 kbp or less downstream of coding sequences)), and these SNVs were subjected to further manual review.

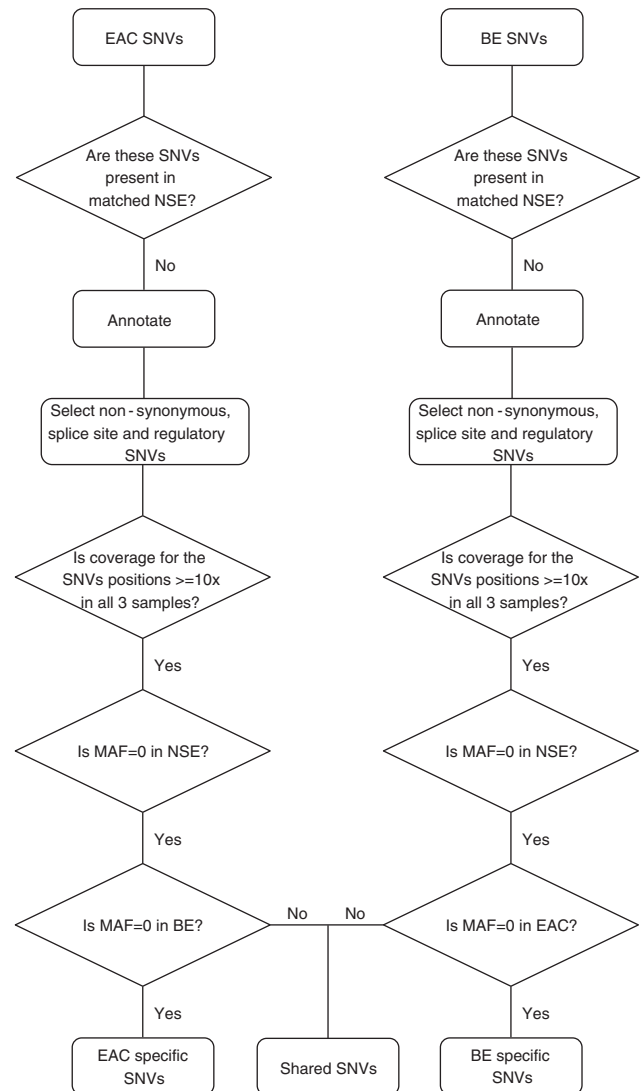
### Comparison of EAC and BE SNVs

In order to identify mutations that are unique to EAC or BE or are shared between these lesions, the lists of EAC and BE SNVs were compared. An in-house Perl script was used to inspect the coding, splice site and regulatory SNVs, and to classify the SNVs into EAC-specific, BE-specific and shared somatic SNVs. Subsequently, the minor allele frequency (MAF) was calculated for every SNV location in each of the three samples using an in-house Perl script. The detailed comparison strategy is depicted as a flowchart (Figure 5).

The SNVs that met the following criteria were selected as candidate SNVs and validated by Sanger sequencing:

- (1) The SNVs have a coverage of  $\geq 10 \times$  in NSE, BE and EAC.
- (2) MAF = 0 in NSE.
- (3) MAF is >0 only in EAC (EAC-specific SNV), or MAF is >0 only in BE (BE-specific SNV), or MAF >0 in both the EAC and BE (shared SNV).

After manual inspection, 46 EAC-specific somatic SNVs (7 missense, 4 UTR3, 2 UTR5, 20 near gene 3 and 13 near gene 5), 49 BE-specific somatic SNVs (11 missense, 1 frameshift, 5 UTR3, 4 UTR5, 11 near gene 3 and 17



**Figure 5.** A flowchart depicting the comparison strategy to identify EAC-specific, BE-specific and shared SNVs.

near gene 5) and 410 somatic SNVs shared between EAC and BE (64 missense, 3 nonsense, 1 Splice5, 49 UTR3, 21 UTR5, 126 near gene 3 and 146 near gene 5) were identified. Finally, the SNVs were categorized into four tiers, as described by Mardis *et al.*<sup>30</sup>; SNVs located in coding regions of annotated exons and splice sites (tier 1), in conserved genomic regions (tier 2), in non-repeat masked, non-regulatory regions (tier 3) and unclassified SNVs (tier 4).

### Sanger sequencing

Tier 1 SNVs and selected tier 2/3/4 SNVs were validated using Sanger sequencing. Primers were designed to amplify the SNV sites using Primer3 (Supplementary Table S8).<sup>58</sup> Specific PCR conditions are available on request. Sanger sequencing was performed using the PCR primers and the Big Dye terminator sequencing kit (Life Technologies, Grand Island, NY, USA) on an Applied Biosystem 3730XL DNA sequencer (Applied Biosystems, Carlsbad, CA, USA).

### Pathway analysis

Tier 1 genes were used as input data for the Ingenuity pathway analysis software (<http://www.ingenuity.com>) (analysis settings available on request). Six networks were identified, and merged.



## ARID1A mutational analysis

HGD/EAC and matched NSE frozen biopsies from 15 patients were obtained during upper endoscopy under a Johns Hopkins IRB-approved protocol. In addition, formalin-fixed paraffin-embedded EAC and matched tumor-negative lymph node tissue from five EAC patients were acquired during surgical resection at the University Medical Center in Utrecht. Lesions and normal tissues were manually microdissected, and DNA was extracted using the DNeasy kit (Qiagen). All exons were sequenced using customized primers (Supplementary Table S9, conditions available on request). Bidirectional Sanger sequencing was performed on an Applied Biosystem 3730XL DNA sequencer.

## Immunohistochemical assessment of loss of ARID1A expression in BE progression

In order to validate aberrations of ARID1A protein expression in the multistep progression of EAC, we utilized samples obtained from a relatively large independent cohort of BE patients, which have been engineered into tissue microarrays by our group, as previously described.<sup>59</sup> Briefly, the tissue microarrays included NSE, BE, LGD, HGD, EAC and lymph node metastasis tissues from 98 EAC patients, who had undergone surgical resection of EAC and had not received neoadjuvant chemo-radiation therapy. Patient characteristics including survival time were collected. Pathological features such as histological differentiation grade, presence of metastases and tumor involvement of surgical margins were extracted from the pathological surgical resection reports.

Immunohistochemistry for ARID1A and p53 was performed as previously described by our group.<sup>59</sup> Primary antibodies against ARID1A (HPA005456, dilution 1:100, Sigma Aldrich, St Louis, MO, USA) and p53 (DO-1, dilution 1:400, Santa Cruz, Santa Cruz, CA, USA) were used.

Staining patterns were scored by two authors (MMS and AM). For both ARID1A and p53, nuclear expression was assessed, and categorized in two groups; present and absent accumulation. Using the log-rank test, correlations between ARID1A status and patient outcome were tested. The Student's *t*-test and  $\chi^2$  test were run to determine associations between ARID1A status and patient characteristics and tumor features. A *P*-value <0.05 was considered as statistically significant.

## Cell culture and RNA interference

The EAC cell lines OE33 (European Collection of Cell Cultures, Wiltshire, UK) and JHesoAD1,<sup>60</sup> and the human esophageal epithelial cell line HEEpIC (Sciencell, Carlsbad, CA, USA) were grown. OE33 and JHesoAD1 cells were cultured in 1640 RPMI supplemented with 15–20% fetal bovine serum, whereas HEEpIC cells were grown in epithelial cell medium-2 (Sciencell). OE33 cells were transfected using lipofectamine 2000 (Life Technologies) and ON-TARGETplus SMARTpool ARID1A small interfering RNA (50 nM, 48-h incubation, Thermo Fisher Scientific, Lafayette, CO, USA) in order to knockdown ARID1A transcripts. ON-TARGETplus Non-targeting Pool small interfering RNA (Thermo Fisher Scientific) was used for mock transfection.

## Western blotting

ARID1A expression was validated at protein level by western blot analysis in ARID1A KD OE33, mock OE33, non-transfected OE33, JHesoAD1 and HEEpIC cells. Western blot analysis was performed as previously described by our group.<sup>59</sup> Primary antibodies against ARID1A (A301–041A, Bethyl Laboratories, Montgomery, TX, USA, dilution 1:500) and  $\alpha$ -tubulin (protein loading control, Santa Cruz, dilution 1:30 000) were used. Band intensities were measured using ImageJ software (National Institutes of Health, Bethesda, MD, USA).

## Cell growth, migration and invasion assays

MTT assays were performed to evaluate whether ARID1A KD cells have a growth advantage compared with mock OE33 cells. Migration and invasion assays were performed as previously described.<sup>61</sup> Five randomly selected 20  $\times$  fields were counted per insert, and experiments were performed in triplicate and repeated three times.

## Proliferation assay

Proliferation was examined in two biological replicated experiments. OE33 cells were seeded and transfected in cell culture chamber slides (BD Biosciences, San Jose, CA, USA). After 36 h, media was replaced by media containing BrdU. Twelve hours thereafter, BrdU was detected using the BrdU staining kit (Life Technologies). Five random 20  $\times$  fields per

experiment were examined. Staining intensity was categorized as follows; moderate-intense, and no-light. Percentages of moderate-intense staining were calculated. A parallel ARID1A immunocytochemistry experiment was performed to confirm ARID1A KD.

## Soft agar colony formation assay

Non-transfected, mock and ARID1A KD OE33 cells, mixed in 0.5% agarose in normal growth media, were seeded on a 1% agarose layer. The experiments were performed in triplicate. Colonies were formed after 2 weeks, and were subsequently stained and fixed in 0.005% crystal violet/10% methanol. Colonies were manually counted, and the experiment was repeated twice.

## Analysis functional assays

Cell/colony counts were corrected for viability differences by performing a parallel MTT assay. The Mann–Whitney *U*-test was used to determine statistically significant differences between ARID1A KD and mock OE33 cells.

## Identification of ARID1A effector genes

Gene expression levels in ARID1A KD were compared with those in mock OE33 cells to identify downstream targets of ARID1A by performing an Affymetrix Human PrimeView Gene Expression Array across two biological replicates. Data were robust micro-array average (RMA) normalized and converted to Log2 notation with Partek Genomics Suite. ARID1A KD and mock experiments were compared with one-way analysis of variance. Potential downstream genes of ARID1A were selected for quantitative reverse transcriptase–PCR validation using the following criteria:

1. An average relative fold change >1.75 in ARID1A KD compared with mock OE33 cells was calculated.
2. The aberrantly expressed gene has a known function in cancer cell proliferation and/or invasion. Available databases of gene function including Ingenuity pathway analysis and Pubmed were queried to identify putative candidate genes that have the ability to control cell proliferation, and/or invasion.

The microarray data are accessible through GEO Series accession number GSE38380 (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE38380>).

## Quantitative real-time PCR

RNA was extracted using the RNeasy kit (Qiagen). Quantitative reverse transcriptase–PCR was carried out using SYBR-green reagents (Applied Biosystems) and customized primers (Supplementary Table S10). Gene expression was normalized to *glyceraldehyde 3-phosphate dehydrogenase* expression. Relative gene expression levels were calculated using the  $2^{-\Delta\Delta CT}$  method.<sup>62</sup> Differences between ARID1A KD and mock OE33 cells were tested using the Mann–Whitney *U*-test. *P*<0.05 was considered as statistically significant.

## CONFLICT OF INTEREST

WRM has participated in Illumina sponsored meetings over the past 4 years and received travel reimbursement and honoraria for presenting at these events. The remaining authors declare no conflict of interest.

## ACKNOWLEDGEMENTS

The work at Johns Hopkins Medical Institutions was funded by the Jerry D'Amato foundation and the NIH (1K23DK068149) and the work at WRM's laboratory was supported by a Cancer Center Support Grant (CA045508) from the NCI and NIH Shared Instrument grant S10 RR023702-01. MMS has been sponsored by Fulbright/ the Netherland America Foundation and the René Vogels Foundation. The patient's consent was obtained under an IRB-approved protocol.

**Author contributions:** MMS acquired, analyzed the data, performed statistical analysis and drafted the manuscript. SL performed the bioinformatics analysis. SL, MD, AMM, SP and EA acquired, analyzed and interpreted data, and critically revised the manuscript. JSW, MIC, FHMM and GJO provided materials and critically revised the manuscript. EAM, WRM and AM interpreted data, critically revised the manuscript, obtained funding and supervised the study. We would like to sincerely thank Conover Talbot Jr from the Institute for Basic Biomedical Sciences at Johns Hopkins School of Medicine for his help with the bioinformatics analysis of the Affymetrix Gene Expression Array data.

## REFERENCES

- Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. *CA Cancer J Clin* 2012; **62**: 10–29.
- Reid BJ, Li X, Galipeau PC, Vaughan TL. Barrett's oesophagus and oesophageal adenocarcinoma: time for a new synthesis. *Nat Rev Cancer* 2010; **10**: 87–101.
- Gilbert EW, Luna RA, Harrison VL, Hunter JG. Barrett's esophagus: a review of the literature. *J Gastrointestinal Surg* 2011; **15**: 708–718.
- Hvid-Jensen F, Pedersen L, Drewes AM, Sorensen HT, Funch-Jensen P. Incidence of adenocarcinoma among patients with Barrett's esophagus. *N Engl J Med* 2011; **365**: 1375–1383.
- Yousef F, Cardwell C, Cantwell MM, Galway K, Johnston BT, Murray L. The incidence of esophageal cancer and high-grade dysplasia in Barrett's esophagus: a systematic review and meta-analysis. *Am J Epidemiol* 2008; **168**: 237–249.
- Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011; **144**: 646–674.
- Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. *Nat Med* 2004; **10**: 789–799.
- Sato F, Meltzer SJ. CpG island hypermethylation in progression of esophageal and gastric cancer. *Cancer* 2006; **106**: 483–493.
- Jin Z, Cheng Y, Gu W, Zheng Y, Sato F, Mori Y et al. A multicenter, double-blinded validation study of methylation biomarkers for progression prediction in Barrett's esophagus. *Cancer Res* 2009; **69**: 4112–4115.
- Alvarez H, Opalinska J, Zhou L, Sohal D, Fazzari MJ, Yu Y et al. Widespread hypomethylation occurs early and synergizes with gene amplification during esophageal carcinogenesis. *PLoS Genet* 2011; **7**: e1001356.
- Paulson TG, Maley CC, Li X, Li H, Sanchez CA, Chao DL et al. Chromosomal instability and copy number alterations in Barrett's esophagus and esophageal adenocarcinoma. *Clin Cancer Res* 2009; **15**: 3305–3314.
- Li X, Galipeau PC, Sanchez CA, Blount PL, Maley CC, Arnaudo J et al. Single nucleotide polymorphism-based genome-wide chromosome copy change, loss of heterozygosity, and aneuploidy in Barrett's esophagus neoplastic progression. *Cancer Prev Res* 2008; **1**: 413–423.
- di Pietro M, Lao-Sirieix P, Boyle S, Cassidy A, Castillo D, Saadi A et al. Evidence for a functional role of epigenetically regulated midcluster HOXB genes in the development of Barrett esophagus. *Proc Natl Acad Sci USA* 2012; **109**: 9077–9082.
- Goh XY, Rees JR, Paterson AL, Chin SF, Marioni JC, Save V et al. Integrative analysis of array-comparative genomic hybridisation and matched gene expression profiling data reveals novel genes with prognostic significance in oesophageal adenocarcinoma. *Gut* 2011; **60**: 1317–1326.
- Kaz AM, Grady WM. Epigenetic biomarkers in esophageal cancer. *Cancer Lett* (e-pub ahead of print 7 March 2012).
- Paulson TC, Galipeau PC, Xu L, Kissel HD, Li X, Blount PL et al. p16 mutation spectrum in the premalignant condition Barrett's esophagus. *PLoS ONE* 2008; **3**: e3809.
- Galipeau PC, Li X, Blount PL, Maley CC, Sanchez CA, Odze RD et al. NSADs modulate CDKN2A, TP53, and DNA content risk for progression to esophageal adenocarcinoma. *PLoS Med* 2007; **4**: e67.
- Reid BJ. p53 and neoplastic progression in Barrett's esophagus. *Am J Gastroenterol* 2001; **96**: 1321–1323.
- Wood LD, Parsons DW, Jones S, Lin J, Sjoblom T, Leary RJ et al. The genomic landscapes of human breast and colorectal cancers. *Science* 2007; **318**: 1108–1113.
- Jones S, Wang TL, Shih IeM, Mao TL, Nakayama K, Roden R et al. Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. *Science* 2010; **330**: 228–231.
- Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K et al. Mutational processes molding the genomes of 21 breast cancers. *Cell* 2012; **149**: 979–993.
- Link DC, Schuettelpelz LG, Shen D, Wang JL, Walter MJ, Kulkarni S et al. Identification of a novel TP53 cancer susceptibility mutation through whole-genome sequencing of a patient with therapy-related AML. *JAMA* 2011; **305**: 1568–1576.
- Clark MJ, Homer N, O'Connor BD, Chen Z, Eskin A, Lee H et al. U87MG decoded: the genomic sequence of a cytogenetically aberrant human cancer cell line. *PLoS Genet* 2010; **6**: e1000832.
- Kumar A, White TA, MacKenzie AP, Clegg N, Lee C, Dumpit RF et al. Exome sequencing identifies a spectrum of mutation frequencies in advanced and lethal prostate cancers. *Proc Natl Acad Sci USA* 2011; **108**: 17087–17092.
- Wu J, Matthaei H, Maitra A, Dal Molin M, Wood LD, Eshleman JR et al. Recurrent GNAS mutations define an unexpected pathway for pancreatic cyst development. *Sci Translational Med* 2011; **3**: 92ra66.
- Wang KK, Wongkeesong M, Buttar NS. American Gastroenterological Association medical position statement: role of the gastroenterologist in the management of esophageal carcinoma. *Gastroenterology* 2005; **128**: 1468–1470.
- van Hagen P, Hulshof MC, van Lanschot JJ, Steyerberg EW, van Berge Henegouwen MI, Wijnhoven BP et al. Preoperative chemoradiotherapy for esophageal or junctional cancer. *N Engl J Med* 2012; **366**: 2074–2084.
- Alvarez H, Montgomery EA, Karikari C, Canto M, Dunbar KB, Wang JS et al. The Axl receptor tyrosine kinase is an adverse prognostic factor and a therapeutic target in esophageal adenocarcinoma. *Cancer Biol Ther* 2010; **10**: 1009–1018.
- Huang J, Zhao YL, Li Y, Fletcher JA, Xiao S. Genomic and functional evidence for an ARID1A tumor suppressor role. *Genes Chromosomes Cancer* 2007; **46**: 745–750.
- Mardis ER, Ding L, Dooling DJ, Larson DE, McLellan MD, Chen K et al. Recurring mutations found by sequencing an acute myeloid leukemia genome. *New Engl J Med* 2009; **361**: 1058–1066.
- Li M, Zhao H, Zhang X, Wood LD, Anders RA, Choti MA et al. Inactivating mutations of the chromatin remodeling gene ARID2 in hepatocellular carcinoma. *Nat Genet* 2011; **43**: 828–829.
- Wang K, Kan J, Yuen ST, Shi ST, Chu KM, Law S et al. Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer. *Nat Genet* 2011; **43**: 1219–1223.
- Wang S, Zhan M, Yin J, Abraham JM, Mori Y, Sato F et al. Transcriptional profiling suggests that Barrett's metaplasia is an early intermediate stage in esophageal adenocarcinogenesis. *Oncogene* 2006; **25**: 3346–3356.
- Ashworth A, Lord CJ, Reis-Filho JS. Genetic interactions in cancer progression and treatment. *Cell* 2011; **145**: 30–38.
- Reya T, Morrison SJ, Clarke MF, Weissman IL. Stem cells, cancer, and cancer stem cells. *Nature* 2001; **414**: 105–111.
- Yates LR, Campbell PJ. Evolution of the cancer genome. *Nat Rev Genet* 2012; **13**: 795–806.
- Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW et al. The life history of 21 breast cancers. *Cell* 2012; **149**: 5.
- Jones S, Li M, Parsons DW, Zhang X, Wesseling J, Kristel P et al. Somatic mutations in the chromatin remodeling gene ARID1A occur in several tumor types. *Hum Mutat* 2012; **33**: 100–103.
- Wiegand KC, Shah SP, Al-Agha OM, Zhao Y, Tse K, Zeng T et al. ARID1A mutations in endometriosis-associated ovarian carcinomas. *N Engl J Med* 2010; **363**: 1532–1543.
- Agrawal N, Jiao Y, Bettgowda C, Hutfless SM, Wang Y, David S et al. Comparative genomic analysis of esophageal adenocarcinoma and squamous cell carcinoma. *Cancer Discovery* 2012; **2**: 899–905.
- Zang ZJ, Cutcutache I, Poon SL, Zhang SL, McPherson JR, Tao J et al. Exome sequencing of gastric adenocarcinoma identifies recurrent somatic mutations in cell adhesion and chromatin remodeling genes. *Nat Genet* 2012; **44**: 570–574.
- Luby TM. Targeting cytochrome P450 CYP1B1 with a therapeutic cancer vaccine. *Expert Rev Vaccines* 2008; **7**: 995–1003.
- McFadyen MC, Melvin WT, Murray GI. Cytochrome P450 enzymes: novel options for cancer therapeutics. *Mol Cancer Ther* 2004; **3**: 363–371.
- Martinez VG, O'Connor R, Liang Y, Clynes M. CYP1B1 expression is induced by docetaxel: effect on cell viability and drug resistance. *Br J Cancer* 2008; **98**: 564–570.
- Mishra SK, Siddique HR, Saleem M. S100A4 calcium-binding protein is key player in tumor progression and metastasis: preclinical and clinical evidence. *Cancer Metastasis Rev* 2011; **31**: 163–172.
- Sherbet GV, Lakshmi MS. S100A4 (MTS1) calcium binding protein in cancer growth, invasion and metastasis. *Anticancer Res* 1998; **18**: 2415–2421.
- Lee OJ, Hong SM, Belkhir A, Moskaluk C, El-Rifai W. Overexpression of calcium binding protein S100A4 in Barrett's tumorigenesis. *Gastroenterology* 2006; **130**: A273–A273.
- Berinstein NL. Carcinoembryonic antigen as a target for therapeutic anticancer vaccines: a review. *J Clin Oncol* 2002; **20**: 2197–2207.
- Griffin M, Sweeney EC. The relationship of endocrine cells, dysplasia and carcinoembryonic antigen in Barrett's mucosa to adenocarcinoma of the oesophagus. *Histopathology* 1987; **11**: 53–62.
- Ordóñez C, Screaton RA, Iltis C, Stanners CP. Human carcinoembryonic antigen functions as a general inhibitor of anoikis. *Cancer Res* 2000; **60**: 3419–3424.
- Camacho-Leal P, Stanners CP. The human carcinoembryonic antigen (CEA) GPI anchor mediates anoikis inhibition by inactivation of the intrinsic death pathway. *Oncogene* 2008; **27**: 1545–1553.
- Fair K, Anderson M, Bulanova E, Mi H, Tropshug M, Diaz MO. Protein interactions of the MLL PHD fingers modulate MLL target gene regulation in human cells. *Mol Cell Biol* 2001; **21**: 3589–3597.
- Wang Z, Song J, Milne TA, Wang GG, Li H, Allis CD et al. Pro isomerization in MLL1 PHD3-bromo cassette connects H3K4me readout to CYP33 and HDAC-mediated repression. *Cell* 2010; **141**: 1183–1194.
- Park S, Osmer U, Raman G, Schwantes RH, Diaz MO, Bushweller JH. The PHD3 domain of MLL acts as a CYP33-regulated switch between MLL-mediated activation and repression. *Biochemistry-US* 2010; **49**: 6576–6586.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N et al. The sequence alignment/map format and SAMtools. *Bioinformatics* 2009; **25**: 2078–2079.

- 56 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A *et al*. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; **20**: 1297–1303.
- 57 DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C *et al*. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011; **43**: 491–498.
- 58 Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* 2000; **132**: 365–386.
- 59 Streppel MM, Vincent A, Mukherjee R, Campbell NR, Chen SH, Konstantopoulos K *et al*. Mucin 16 (cancer antigen 125) expression in human tissues and cell lines and correlation with clinical outcome in adenocarcinomas of the pancreas, esophagus, stomach, and colon. *Hum Pathol* 2012; **43**: 1755–1763.
- 60 Alvarez H, Koorstra JB, Hong SM, Boonstra JJ, Dinjens WN, Foratiere AA *et al*. Establishment and characterization of a bona fide Barrett esophagus-associated adenocarcinoma cell line. *Cancer Biol Ther* 2008; **7**: 1753–1755.
- 61 Gupta S, Pramanik D, Mukherjee R, Campbell NR, Elumalai S, de Wilde RF *et al*. Molecular determinants of retinoic acid sensitivity in pancreatic cancer. *Clin Cancer Res* 2012; **18**: 280–289.
- 62 Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc* 2008; **3**: 1101–1108.
- 63 Cornen S, Adelaide J, Bertucci F, Finetti P, Guille A, Birnbaum DJ *et al*. Mutations and deletions of ARID1A in breast tumors. *Oncogene* 2012; **31**: 4255–4256.
- 64 Mamo A, Cavallone L, Tuzmen S, Chabot C, Ferrario C, Hassan S *et al*. An integrated genomic approach identifies ARID1A as a candidate tumor-suppressor gene in breast cancer. *Oncogene* 2012; **31**: 2090–2100.
- 65 Fujimoto A, Totoki Y, Abe T, Boroevich KA, Hosoda F, Nguyen HH *et al*. Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat Genet* 2012; **44**: 760–764.
- 66 Birnbaum DJ, Adelaide J, Mamessier E, Finetti P, Lagarde A, Monges G *et al*. Genome profiling of pancreatic adenocarcinoma. *Genes Chromosomes Cancer* 2011; **50**: 456–465.
- 67 Gui Y, Guo G, Huang Y, Hu X, Tang A, Gao S *et al*. Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nat Genet* 2011; **43**: 875–878.

Supplementary Information accompanies the paper on the Oncogene website (<http://www.nature.com/onc>)