Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Barnetson RA, Tenesa A, Farrington SM, et al. Identification and survival of carriers of mutations in DNA mismatch-repair genes in colon cancer. N Engl J Med 2006;354:2751-63.

Supplementary Table 1

Variables investigated as possible predictors of mismatch repair gene defects for the first stage (see Supplementary Table 3A for univariate analysis)

- Patient age at onset (AAO)

- Gender (SEX)
  Female
  Male

- Differentiation of the tumour (DIFFEREN)
  Well
  Poor
  Moderate

- Histology (HISTOLOG)
  Standard
  Mucinous

- Location of the tumour (LOCATION)
  Right side
  Left side

- Other synchronous or metachronous tumours (SYNMET_TUM)
  No
  Yes

- 1st degree relative with colorectal cancer (CRC1FH)
  No
  Yes and AAO of the relative was before 50 years
  Yes and AAO of the relative was at or after 50 years

- Number of 1st degree relatives with colorectal cancer (CRC1NO)

- 2nd degree relative with colorectal cancer (CRC2FH)
  No
  Yes and AAO of the relative was before 50 years
  Yes and AAO of the relative was at or after 50 years

- Number of 2nd degree relatives with colorectal cancer (CRC2NO)

- 1st degree relative with endometrial cancer (END1FH)
  No
  Yes and AAO of the relative was before 50 years
  Yes and AAO of the relative was at or after 50 years

- Number of 1st degree relatives with endometrial cancer (END1NO)

- 2nd degree relative with endometrial cancer (END2FH)
  No
  Yes and AAO of the relative was before 50 years
  Yes and AAO of the relative was at or after 50 years

- Number of 2nd degree relatives with endometrial cancer (END2NO)
- 1st degree relative with gastric cancer (STO1FH)
  No
  Yes and AAO of the relative was before 50 years
  Yes and AAO of the relative was at or after 50 years

- Number of 1st degree relatives with gastric cancer (STO1NO)
- 2nd degree relative with gastric cancer (STO2FH)
  No
  Yes and AAO of the relative was before 50 years
  Yes and AAO of the relative was at or after 50 years

- Number of 2nd degree relatives with gastric cancer (STO2NO)
- 1st degree relative with kidney cancer (KID1FH)
  No
  Yes and AAO of the relative was before 50 years
  Yes and AAO of the relative was at or after 50 years

- Number of 1st degree relatives with kidney cancer (KID1NO)
- 2nd degree relative with kidney cancer (KID2FH)
  No
  Yes and AAO of the relative was before 50 years
  Yes and AAO of the relative was at or after 50 years

- Number of 2nd degree relatives with kidney cancer (KID2NO)
- 1st degree relative with any other cancer (OTH1FH)
  No
  Yes and AAO of the relative was before 50 years
  Yes and AAO of the relative was at or after 50 years

- Number of 1st degree relatives with any other cancer (OTH1NO)
- 2nd degree relative with any other cancer (OTH2FH)
  No
  Yes and AAO of the relative was before 50 years
  Yes and AAO of the relative was at or after 50 years

- Number of 2nd degree relatives with any other cancer (OTH2NO)

Supplementary Table 2

Tumour characteristics by DNA mismatch repair genotype

A *Tumour location*

	Carriers (%)		Non-Carriers (%)	
Rectum	7	(18.4)	285	(35.2)
Rectosigmoid	2	(5.3)	66	(8.1)
Sigmoid	5	(13.2)	183	(22.6)
Descending colon	2	(5.3)	37	(4.6)
Distal colon (total)	**16**	**(42)**	**571**	**(70.5)**
Splenic flexure	0		21	(2.6)
Transverse colon	3	(7.9)	30	(3.7)
Ascending colon/hepatic flexure	10	(26.3)	68	(8.4)
Caecum	9	(23.7)	110	(13.6)
Appendix	0		10	(1.2)
Proximal colon (total)	**22**	**(58)**	**239**	**(29.5)**
Site assessable (total)	38		810	
Site not assessable	0		22	

B *Presence of synchronous and metachronous tumours*

	Carriers (%) (n=38)		Non carriers (%) (n=832)		Total
Synchronous colorectal cancer	4	(10.5)	12	(1.4)	16
Synchronous other HNPCC-related cancer	0		3	(0.4)	3
Synchronous other invasive cancer	0		5	(0.6)	5
Previous colorectal cancer	3	(7.9)	2	(0.2)	5
Previous other HNPCC-related cancer	1	(2.6)	5	(0.6)	6
Previous other invasive cancer	2	(5.3)	24	(2.9)	26

C *Pathological features by carrier status*

	Carriers	(%)	Non carriers	(%)
Tumour staging*				
Stage I	7	(18.4)	145	(17.6)
Stage II	19	(50.0)	244	(29.6)
Stage III	8	(21.1)	300	(36.4)
Stage IV	4	(10.5)	135	(16.4)
Stage assessable (total)	38		824	
Stage not assessable	0		8	
Degree of differentiation				
Well/Mod	27	(73.0)	650	(83.8)
Poor	10	(27.0)	126	(16.2)
Not assessable	1		56	
Histological type				
Mucinous	13	(35.1)	121	(15.2)
Standard	23	(62.2)	667	(83.6)
Signet ring cell	0		7	(0.9)
Neuroendocrine diff.	1	(2.7)	3	(0.4)
Not assessable	1		34	

Supplementary Table 2 legend

Clinico-pathological data were collected from hospital clinical records, pathology reports and surgical audit returns, including data on tumour staging, topographical location, histological type and degree of differentiation. Inclusion criteria required proven invasive adenocarcinoma. Tumours that had not broached the epithelial basement membrane were considered benign and excluded, regardless of the degree of cytological atypia. Summary tables presenting the relationship between DNA mismatch repair gene mutation status and: A - location of the index colorectal cancer; B - proportion of patients with concurrent (synchronous) colorectal cancer or other Lynch Syndrome associated tumours and/or with prior (metachronous) colorectal cancer, other Lynch Syndrome associated tumours or any other invasive cancer type; C - pathological features of the index colorectal cancer - tumour stage categorised using the American Joint Committee on Cancer (AJCC) stage groupings of the TNM system; the predominant degree of tumour differentiation; the predominant histological sub-type of the tumour. It was not possible to assess every variable for all tumours and so the number of tumours that were assessable is provided. For the sake of clarity, the various tumour types have not been entered in section B. The tumours of each type and the relevant genotypes (non-carriers unless stated) included in the synchronous or metachronous non-HNPCC other tumour category were: germ-cell tumour of testis (n=3), cutaneous malignant melanoma (n=2) and cutaneous squamous carcinoma (n=3, including 1 carrier), breast carcinoma (n=10, including 1 carrier), endometrial carcinoma (n=1, including 1 carrier), gastric carcinoma (n=1), ovarian carcinoma (n=3), brain tumour (n=3), laryngeal carcinoma (n=1), limb sarcoma (n=1 carrier), invasive cervix uteri carcinoma (n=2), leukaemia (n=1), lymphoma (n=3), vaginal squamous carcinoma (n=1). Some patients had more than one tumour type.

Sample availability is incomplete because of the reliance on provision from hard-pressed NHS pathology laboratories. For the MSI tumour analysis presented in Table 3 in the text, there were 352 available tumour/normal pairs (30 carriers, 322 non-carriers) that were assessable unambiguously. For immunohistochemistry tumour analysis presented in Table 3 in the text, there were 390 patients (31 carriers, 359 non-carriers) with assessable paraffin-embedded sections, 275 of whom had tumours assessable for all three DNA mismatch repair proteins (19 carriers, 256 non-carriers). Immunostaining data was not available for each gene for every available tumour/normal section because of sub-optimal staining or poor tumour/normal representation.

Supplementary Table 3A Univariate analysis used to construct the two-stage predictive model

Predictor	P value	Category	Reference category	OR (95% CI)
Variables tested for 1st stage				
Age at onset	<0.001	Continuous variable	N/A	0.90 (0.86-0.94)
Other synchronous/metachronous tumours	<0.001	Yes	No	8.02 (3.00-21.45)
Location of index primary colorectal cancer	<0.001	Distal	Proximal	0.27 (0.14-0.52)
Gender	0.03	Male	Female	2.24(1.10-4.58)
Histological subtype	0.003	Mucinous	Standard type	3.13 (1.54-6.35)
Stage (AJCC stage groupings I-IV)	0.06	-	-	-
Degree of differentiation	0.08	-	-	-
1st degree relative with colorectal cancer <50yrs	<0.0001	Affected relative aged <50yrs	No affected relative	36 (13-100)
		Affected relative aged ≥50yrs	No affected relative	4.24 (1.39-13.95)
2nd degree relative with colorectal cancer <50yrs	0.009	Affected relative aged <50yrs	No affected relative	8.07 (2.48-26.29)
		Affected relative aged ≥50yrs	No affected relative	1.91 (0.65-5.62)
1st degree relative with endometrial cancer <50yrs	0.011	-	-	-
2nd degree relative with endometrial cancer <50yrs	0.34	-	-	-
1st degree relative with gastric cancer <50yrs	0.011	-	-	-
2nd degree relative with gastric cancer <50yrs	0.004	Affected relative aged <50yrs	No affected relative	9.42 (2.39-37.19)
		Affected relative aged ≥50yrs	No affected relative	4.71 (1.31-16.99)
1st degree relative with upper urinary tract cancer <50yrs	0.06	-	-	-
2nd degree relative with upper urinary tract cancer <50yrs	0.15	-	-	-
1st degree relative with other cancer <50yrs	0.06	-	-	-
2nd degree relative with other cancer <50yrs	0.12	-	-	-
HNPCC criteria				
Bethesda	0.001	Positive	Negative	11 (2-45)
Amsterdam modified	<0.001	Positive	Negative	33 (15-73)
Variable tested in 2nd stage				
MSI status	<0.001	MSI-L	MSS	42 (8-206)
		MSI-H	MSS	113 (257-513)
		MSI-H or MSI-L	MSS	76 (18-329)
Immunohistochemistry	<0.001	Negative	Positive	81 (26-252)

Supplementary Table 3B Comparison of effectiveness of the two-stage predictive model at various cutoff values with Bethesda and Amsterdam criteria

	After Stage 1							After Stage 2							
	Prospective Series			Replication Series				MSI and IHC tumour analysis						Identified by Stage 2 (% of total series)	
Cut-off	**Sensitivity**	**Specificity**	**Positive predictive value**	**Sensitivity**	**Specificity**	**Positive predictive value**	**Number passing to Stage 2 (%)**	**Sensitivity**		**Specificity**		**Positive predictive value**			
								MSI	**IHC**	**MSI**	**IHC**	**MSI**	**IHC**	**MSI**	**IHC**
0.005	0.95 (0.82-0.99)	0.14 (0.12-0.17)	0.05 (0.03-0.07)	1.00	0.00	0.23	728 (86%)	0.88 (0.78-0.98)	0.81 (0.68-0.93)	0.87 (0.85-0.89)	0.94 (0.92-0.95)	0.24 (0.17-0.31)	0.38 (0.27-0.48)	138 (16%)	81 (9.3%)
0.01	0.89 (0.80-0.99)	0.42 (0.38-0.45)	0.07 (0.05-0.09)	0.94	0.075	0.23	506 (60%)	0.83 (0.71-0.95)	0.76 (0.62-0.89)	0.91 (0.89-0.93)	0.95 (0.94-0.97)	0.30 (0.22-0.39)	0.42 (0.31-0.54)	104 (12%)	68 (7.8%)
0.05	**0.68 (0.51-0.82)**	**0.86 (0.83-0.88)**	**0.19 (0.13-0.26)**	**0.94**	**0.27**	**0.27**	**140 (17%)**	**0.65 (0.50-0.80)**	**0.62 (0.46-0.77)**	**0.97 (0.96-0.98)**	**0.99 (0.98-1.00)**	**0.53 (0.39-0.68)**	**0.80 (0.66-0.95)**	**46 (5.2%)**	**30 (3.4%)**
0.15	0.55 (0.39-0.71)	0.96 (0.95-0.97)	0.40 (0.26-0.53)	0.86	0.64	0.41	53 (6%)	0.52 (0.36-0.68)	0.49 (0.33-0.65)	1.00 (0.99-1.00)	1.00 (1.00-1.00)	0.87 (0.73-1.00)	1.00 (0.85-1.00)	23 (2.6%)	19 (2.2%)
0.2	0.47 (0.31-0.63)	0.97 (0.96-0.98)	0.45 (0.30-0.60)	0.80	0.73	0.46	40 (5%)	0.44 0.29-0.60)	0.44 (0.28-0.60)	1.00 (0.99-1.00)	1.00 (1.00-1.00)	0.8 (0.71-1.00)	1.00 (0.84-1.00)	20 (2.3%)	17 (2.0%)
0.25	0.39 (0.24-0.55)	0.98 (0.97-0.99)	0.45 (0.28-0.62)	0.74	0.76	0.47	33 (4%)	0.36 (0.21-0.51)	0.36 (0.21-0.51)	1.00 (0.99-1.00)	1.00 (1.00-1.00)	0.84 (0.67-1.00)	1.00 (0.81-1.00)	16 (1.8%)	14 (1.6%)
0.35	0.39 (0.24-0.55)	0.99 (0.98-1.00)	0.71 (0.52-0.91)	0.74	0.80	0.52	21 (2%)	0.36 (0.21-0.51)	0.36 (0.21-0.51)	1.00 (0.99-1.00)	1.00 (1.00-1.00)	0.87 (0.71-1.00)	1.00 (0.81-1.00)	16 (1.8%)	14 (1.6%)
0.45	0.34 (0.19-0.49)	1.00 (1.00-1.00)	1.00 (0.79-1.00)	0.66	0.85	0.56	13 (2%)	NA	NA	NA	NA	NA	NA	NA	NA
Family history based criteria															
Bethesda	0.95 (0.82-0.99)	0.38 (0.34-0.41)	0.06 (0.05-0.09)	-	-	-	555 (64%)	0.88 (0.77-0.98)	0.84 (0.73-0.96)	0.92 (0.90-0.94)	0.95 (0.94-0.97)	0.32 (0.23-0.41)	0.44 (0.33-0.56)	103 (12%)	72 (8.3%)
AMS	0.42 (0.26-0.59)	0.98 (0.97-0.99)	0.47 (0.30-0.65)	-	-	-	34 (4%)	0.39 (0.24-0.55)	0.39 (0.24-0.54)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (0.82-1.00)	1.00 (0.82-1.00)	15 (1.7%)	15 (1.7%)

Supplementary Table 3 Legend

A. Variables tested in the univariate analysis as predictors of mutation status with relevant p values and odds ratios (OR). OR are only shown for variables that were significant at the 5% level and with sufficient data to provide OR with narrow 95% confidence intervals. For family history variables, age refers to the age at diagnosis of the relative. MSI status is shown for MSI-H, MSI-L and combined MSI-H&L categories. For Bethesda and Amsterdam criteria, positive refers to fulfilment of current criteria and negative is lack of qualifying variables.

B. Age, tumour location, synchronous/metachronous tumours, first degree relative with colorectal cancer, first degree relative with endometrial cancer were then incorporated as Stage 1 of the model (http://www1.hgu.mrc.ac.uk/Softdata/MMRpredict.php). Sensitivity, specificity and positive predictive value (PPV) are shown for Stage 1, and Stage 2 (tumour MSI status/immunohistochemical analysis). Stage 1 of the model was then tested using an independent series of 155 patients aged <45yrs at diagnosis (mean age 34yrs). These patients were identified retrospectively and systematically recruited from the same Scottish population as that for the prospective series (see methods). Negative predictive value was never less than 97% after the second stage with any model cut-off or with Bethesda and Amsterdam criteria and so is not shown. Mutation analysis is taken as gold standard. A cut-off probability of x means a person is categorised as mutation carrier if his/her predicted probability of being mutation carrier is $\geq$x (non-carrier <x). The model cut-off relates to the stringency of the prediction, whereas Bethesda and Amsterdam criteria are dichotomised. MSI-L and MSI-H status had similar predictive value in the univariate analysis (Panel A), at each model cut-off level, as well as for Bethesda and Amsterdam criteria. Hence, we pooled MSI-L and MSI-H such that MSI is dichotomised to either MSI-H&L stable or unstable. Indeed, this performs better than MSI-H alone (Panel A). Immunohistochemistry status refers to loss of expression of any gene, not necessarily the relevant gene for the germline mutation that was identified. In the univariate analysis, MLH1, MSH2 and MSH6 loss are individual significant predictors of mutation status. However, there is a strong relationship between loss of expression of MSH2 and MSH6. Hence, adding MSH6 data to MLH1 and MSH2 in the model does not significantly better explain the data. However, we used the category of “any gene loss” because it is more useful in practice since it takes into account poor quality immunohistochemistry that might be obtained for one antibody but good quality slides for another showing clear loss of expression.

Figure 1 ROC curves of model performance for the prospective and replication series

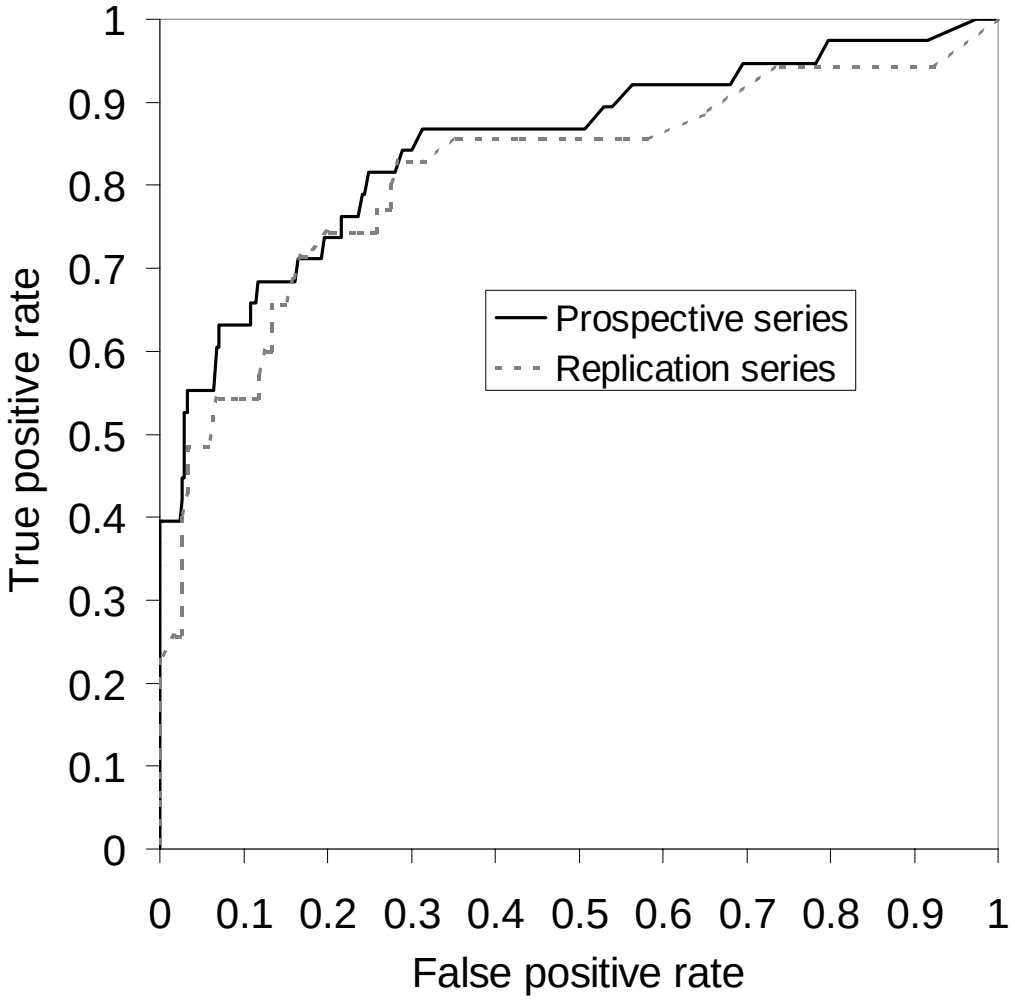

Figure 1 legend

Discrimination power of the model to detect mutation carriers in the prospective and replication case series. The area under each ROC curve was estimated using a nonparametric approach[41] as implemented in SPSS 11.0.