Polymorphic CAG repeat and protein expression of androgen receptor gene in colorectal cancer

Rui Huang\textsuperscript{1}, Guiyu Wang\textsuperscript{1}, Yanni Song\textsuperscript{2}, Feng Wang\textsuperscript{3}, Bing Zhu\textsuperscript{4}, Qingchao Tang\textsuperscript{1}, Zheng Liu\textsuperscript{1}, Yinggang Chen\textsuperscript{1}, Qian Zhang\textsuperscript{1}, Shan Muhammad\textsuperscript{1}, Xishan Wang\textsuperscript{1,5,*}

1 Department of Colorectal cancer Surgery, The Second Affiliated Hospital of Harbin Medical University, 246 Xuefu Road, Harbin 150086, China
2 Department of Tumor Surgery, The Third Affiliated Hospital of Harbin Medical University, 150 Haping Road, Harbin 150081, China
3 Department of General Surgery, Affiliated Hospital of Inner Mongolia Medical University, 1Tongdaobei Street, Hohhot 010050, China
4 Department of General Surgery, Affiliated Liaoyang Central Hospital of China Medical University, 148 Zhonghua Road, Liaoyang 111000, China
5 Colorectal Cancer Institute of Harbin Medical University, 194 Xuefu Road, Harbin 150086, China

*Correspondence to: Xishan Wang, MD, Department of Colorectal surgery, The second Affiliated Hospital of Harbin Medical University, 246 Xuefu Road, Harbin 150086; Colorectal Cancer Institute of Harbin Medical University, 194 Xuefu Road, Harbin 150086, China. Fax: +86-451-86297230. E-mail: xishan1208@126.com

Note: Rui Huang, Guiyu Wang, and Yanni Song contributed equally to this work and should be considered co-first authors.

Running Title: (CAG)n and protein expression of AR in colorectal cancer
**Key Words:** Colorectal cancer; Androgen receptor; Polymorphism; CAG repeat length; Prognosis

**Abbreviations list:** AR: androgen receptor; CRC: colorectal cancer; OS: overall survival; AJCC: American Joint Committee on Cancer; TNM: tumor node metastasis; OR: odd ratio; HR: hazard ratio
Financial Information

Our manuscript is not funded by National Institutes of Health, Medical Institute, or the Medical Research Council; this service is not applicable to us.
Abstract

Although somatic alterations in CAG repeats in the androgen receptor (AR) gene have been suggested to predispose to colorectal cancer (CRC), less is known about AR in CRC carcinogenesis. Due to lack of relevant analysis on CAG repeat length and AR expression in CRC, we aimed to investigate the prognostic value of polymorphic CAG and protein expression of the AR gene in CRC patients. A case-control study was carried out on 550 CRC patients and 540 healthy controls to investigate whether polymorphic CAG within the AR gene is linked to increased risk for CRC. Polymorphic CAG and AR expression were analyzed to clarify their relationship with clinicopathological and prognostic factors in CRC patients. The study showed that the AR gene in CRC patients had a longer CAG repeat sequence than those in the control group, as well as increased risk for CRC among females ($P = 0.013$), males ($P = 0.002$), and total CRC population ($P < 0.001$), respectively. AR expression exhibited a significant difference in long CAG repeat sequence among males ($P < 0.001$), females ($P < 0.001$), and total CRC study population ($P < 0.001$). Both long CAG repeat sequence and negative AR expression were associated with a short 5-year overall survival (OS) rate in CRC. Long CAG repeat sequences and absence of AR expression were closely related to the development of CRC. Both long CAG and decreased AR expression were correlated with the poor 5-year OS in CRC patients.

Introduction

The principal factor in colorectal carcinogenesis still remains unclear. Colorectal cancer (CRC) could be developed through different genetic pathways (1). Colorectal tumors relate to these genetic changes involving the chromosomal instability pathway and some hereditary syndromes, such as familial adenomatous polyposis (2). Recent studies have showed that androgens regulate cellular growth and differentiation in several hormone-dependent tissues, including colorectal tissue (3,4). The androgen receptor (AR) is a ligand-dependent transcription factor that is involved in controlling cellular proliferation and differentiation (5). AR is located on chromosome Xq11-12 and contains a variable number of CAG repeats, which are polymorphic (CAG 8–35) in a normal human population (6). The CAG repeat length of AR inversely affects its transactivation potential, either as a directly altered
receptor function (7, 8) or indirectly reduced AR messenger at RNA and protein levels (9). Somatic alterations (10) and instability (11) of the CAG repeat in the AR gene are correlated with CRC development. However, the genetic influences and protein expression of AR in CRC carcinogenesis remain unknown.

In this study, we examined the association between polymorphisms of the AR gene and risk for CRC, as well as the associations between protein expression and polymorphic CAG to the prognostic significance of CRC. We hypothesized that the risk of CRC would vary with polymorphic CAG repeat lengths of the AR gene. In addition, we also evaluated the associations and interactions between long and short CAG repeat sequences in different expression levels of AR protein.

Materials and methods
Study populations
In this study, patients and controls were from different hospitals, namely, The Second Affiliated Hospital of Harbin Medical University (the second AHMU), Third Affiliated Hospital of Harbin Medical University (the third AHMU) in Heilongjiang Province northeast China, and Affiliated Hospital of Inner Mongolia Medical University in Inner Mongolia Autonomous Region in northwest China. All patients and controls have provided informed consent. The study has been approved by the Heilongjiang and Inner Mongolia Regional Ethics Committee.

Controls
The subjects included in the control group were matched by gender and age, and selected using of a sampling frame, targeting the same population from the same place. Blood samples were collected from 540 cancer-free healthy donors (310 males and 230 females) without a family history of CRC from February 2012 to August 2012.

Patients
Data were collected in 610 consecutive patients with CRC who were operated from January 2004 to December 2008 and underwent AR analysis in biomolecular laboratories from different universities. Patients with inflammatory bowel disease, hereditary nonpolyposis colon cancer, familial polyposis, and metastatic carcinoma were excluded from
this study. All patients received surgery, without giving any adjuvant therapy before surgery. Histopathological assessment confirmed the diagnosis of CRC.

Supplementary Table S1 showed distribution of clinical and pathological features of patients with CRC. The pooled 550 CRC patients were used to study the germ line CAG repeat length. A total of 290 patients who met the same AR(CAG)n repeat length of two paired tumor tissues and blood samples were used as test cohorts in the Affiliated Hospital of Harbin Medical University. While, 260 patients who met the same AR(CAG)n repeat length of two paired tumor tissues and blood samples were used as validation cohorts in the Affiliated Hospital of Inner Mongolia Medical University. The pooled 550 CRC patients comprised of 238 females and 312 males (ratio of 1:1.31). In this study, the average age in the patients group was 59.6 years old, range 34 to 88 years. The tumor grade and stage were classified according to the seventh edition of AJCC (American Joint Committee on Cancer) staging system (12). Among the 550 patients with CRC, both nontumoral (from surrounding “normal” mucosa at least 10 cm away from the tumor) and malignant tissue specimens were collected from 150 patients. The medical and follow-up data of patients, including age, gender, lymph node metastases, type of T stage, diameter and location of primary tumor, TNM stage, degree of primary tumor differentiation, growth pattern, and adjuvant chemotherapy, were retrieved from hospital records and interviews. Adjuvant chemotherapy consisted of two regimens for FOLFOX (5-Fluorouracil, Folinic Acid, and Oxaliplatin) and XELOX (capecitabine plus oxaliplatin).

DNA extraction

The specimens came from both patients group and controls group, including formalin-fixed paraffin-embedded (FFPE) tumor tissues and paired blood samples. DNA extraction complied with the instruction using Tissue DNA Kit (OMEGA, USA) or Blood Genomic DNA Kit (AXYGEN, Union City, CA94587 USA), respectively. Microdissection was used to exclude the normal tissue before FFPE tumor tissues extracted DNA.

PCR-based GeneScan analysis of AR-CAG repeats length

After DNA was extracted from specimens, a region containing the polymorphism of repeat CAG sequence was amplified by PCR. Primers (13) were AR-Forward 5’-GTTTCTGTGGGGCCTCTACGATGG-3’ and AR-Reverse 5’-
GTTTCTGCGCGAAGTGATCCAGAA- 3’. AR-Forward was fluorescently labeled with 6-carboxy-fluoresceine (FAM). PCR was carried out as we have previously been described (14). Data were analyzed by GeneScan Software (Applied Biosystems) and Genetyper Software (Applied Biosystems).

**Immunohistochemistry**

As described previously (14), 4µm sections from formalin-fixed paraffin-embedded tissues were dewaxed and rehydrated according to routine procedure. Sections were permeabilized in 0.5% Triton-X100 for 20 min and incubated overnight with primary antibody (1:100; ab74272; Abcam), followed by biotinylated anti rabbit immunoglobulin secondary antibody for 30 min. Next, sections were visualized by DAB (3,3’-diaminobenzidine tetrahydrochloride) and counterstained with hematoxylin. Positive controls for AR were from prostate carcinoma tissues. PBS (Phosphate-buffered saline) replaced the primary antibody as the negative control. AR positive expression was determined according to the percentage and intensity (15,16).

**Statistical analysis**

The relationship among AR status, polymorphism of CAG repeat and clinicopathological factors was analyzed with a chi-square test. Univariate logistic regression with CRC patients were used to assess the association of AR expression with AR(CAG)n repeat length in individual patients. Kaplan-Meier method was used for univariate survival analyses. Cox proportional hazard regression model was used for estimating the relative importance of unitary and multiple prognostic factors on survival. Date of diagnosis (colonoscopy or surgery) to death or the date last known alive was calculated as overall survival (OS). SPSS 18.0 statistical software (Austin, Texas, USA) for statistical analysis, $P < 0.05$ was considered significant.

**Results**

**Comparison of CAG repeat length between the CRC and control groups**

The PCR products obtained ranged from 206 bp to 287 bp in length (equivalent to 8 and 35 CAG repeats, respectively). Given that the median number of CAG repeats in 550 CRC patients was 21.5±2.5, our study selected 22 CAG repeats of 248 bp in length as cut-points.
Heterozygous women harboring an allele with more than 22 CAG repeats were classified as having “long” CAG repeat sequences. Figs. 1 (A, B, C, and D) showed “short” and “long” CAG repeat sequences in males, homozygous females, and heterozygous females, respectively. Figs. 2 (A, B, and C) showed the distribution of CAG repeat sequences between the control group and CRC patients (males, females, and total population, respectively). Fig S1 (A and B) showed distribution of “short” and “long” CAG repeat sequences in cases and controls from the different centers. Supplementary Table S2 illustrated that, alleles in the control group mainly comprised of 22 or fewer CAG repeats (males, 55.4%; females, 54.3%; total population, 54.9%), whereas alleles in the CRC group (males, 56.6%; females, 56.5%; total population, 56.5%) consisted of 22 to 35 repeats. Supplementary Table S2 also showed that long CAG repeat sequences increased the risk for CRC among males (OR, 1.293; 95% CI, 1.090–1.534; \( P = 0.002 \)), females (OR, 1.255; 95% CI, 1.035–1.523; \( P = 0.013 \)) and total population (OR, 1.277; 95% CI, 1.123–1.451; \( P < 0.001 \)).

**Analysis of AR protein expression in CRC and associativity of between AR and CAG repeat length**

The positive expression of AR protein was mainly weak, and moderate staining was observed in CRC cell nuclei. AR staining demonstrated the following results: 315 cases (male, 175 cases; female, 140 cases) showed negative immunostaining (Fig. 3A), and 235 cases (male, 137 cases; female, 98 cases) showed positive immunostaining (Fig. 3B). Supplementary Table S3 showed that AR expression was positive in 42.7% of the 550 tumor tissues and 81.3% of the 150 paired “normal” mucosa tissues (Fig. 3C) (\( P < 0.001 \)). Supplementary Table S4 showed the relationship between the CAG repeat length and AR expression in CRC patients. Long CAG repeat sequences exhibited a significant difference between the negative and positive AR subgroups in females (\( P < 0.001 \)), males (\( P < 0.001 \)), and total CRC population (\( P < 0.001 \)). Univariate logistic regression with CRC patients indicated the association of AR expression with AR(CAG)n repeat length (OR, 0.048; 95% CI, 0.031–0.075; \( P < 0.001 \)) in individual patients.

In this study, their distributions were similar between the test and validation cohorts. By the time of the final analysis on 30 December 2013, 110 patients (37.9%) in the test cohort died of CRC, with a median survival time of 46.4 months (5-year OS rate 62.1%). Among
the patients in the validation cohort, 103 (39.6%) died of CRC, with a median survival time of 46.0 months (5-year OS rate 60.4%). The 5-year OS rate for the pooled sample of two cohorts was 61.3%.

We also investigated the associations of CAG repeat lengths and AR expression in some clinical and pathological features, such as age, gender, tumor size, type of T stage, location of primary tumor, lymph node metastases, TNM stage, degree of differentiation, growth pattern, and adjuvant chemotherapy (Table 1). AR expression and long CAG repeat sequences were related to large tumors (≥5cm) (AR, \( P = 0.001 \); CAG, \( P = 0.001 \)) and degree of differentiation (AR, \( P < 0.001 \); CAG, \( P = 0.028 \)), respectively. AR expression was related to pathological T stage (\( P = 0.001 \)). Long CAG repeat sequences showed a higher percentage in pathological N1-2 stage (\( P < 0.001 \)).

**Survival analysis of AR expression and CAG repeat length in CRC**

We initially wanted to verify whether various CAG repeat lengths and AR expression would contribute to survival. We found that CAG repeat length (Figs. 2D and E) and AR expression (Figs. 3D and E) significantly affected the survival (5-year OS) in the test and validation cohorts. We also found that CAG repeat lengths (Figs. 2F, G, and H) and AR expression (Figs. 3F, G, and H) significantly affected the survival (5-year OS) in males, females and total CRC population, respectively.

We analyzed the 5-year OS according to AR expression, CAG repeat length, and clinicopathological parameters (Table 2) in the test cohort, validation cohort, and pooled samples. Univariate analysis revealed the associations of various factors, including largest tumor diameter (\( P = 0.005 \)), pathological T (pT) stage (\( P = 0.002 \)), pathological N (pN) stage (\( P = 0.019 \)), TNM stage (\( P = 0.024 \)), AR expression (\( P = 0.001 \)) and CAG repeat length (\( P = 0.001 \)), with CRC prognosis. Tumors with large diameters (≥5cm), poor pT (T3-4)/pN (N1-2) stage, high TNM III stage or negative AR expression, or long CAG repeat lengths were indicators of poor CRC prognosis. No significant differences in gender, age, differentiation, tumor location, or growth pattern were observed (\( P > 0.05 \)).

We further investigated these adverse factors in the Cox multivariate survival model. Multivariate analysis for 5-year OS (Table 2) showed that four variables, namely, large tumor diameter ≥5 cm (HR, 1.372; 95% CI, 1.044–1.803; \( P = 0.023 \)), pT3-4 stage (HR,
1.434; 95% CI, 1.075–1.912; \( P = 0.014 \)), negative AR expression (HR, 1.480; 95% CI, 1.105–1.982; \( P = 0.008 \)) and long CAG repeat lengths (HR, 1.364; 95% CI, 1.042–1.785; \( P = 0.024 \)), were significantly correlated with prognosis whereas pN and TNM stages were not considered independent risk factors (\( P > 0.05 \)).

**Survival analysis of AR expression and CAG repeat length in adjuvant chemotherapy CRC**

We analyzed 296 patients with stages IIb and III CRC who underwent postoperative adjuvant chemotherapy to improve risk stratification. Fig. 4 showed the relationship between AR expression (Fig. 4A) as well as CAG repeat length (Fig. 4B), and recurrence or metastasis in 296 postoperative adjuvant chemotherapy patients. Fig. 4A showed that 100 patients with negative AR expression (58.1%) and 56 patients with positive AR expression (45.2%) received postoperative adjuvant chemotherapy, and a statistical difference was observed in recurrence or metastasis in negative and positive AR expression (\( P = 0.018 \)). Fig. 4B showed that 84 patients with long CAG repeat sequences (51.4%) and 52 patients with short CAG repeat sequences (39.1%) received postoperative adjuvant chemotherapy, and a statistical difference was observed in recurrence or metastasis in long and short CAG repeat sequence (\( P = 0.022 \)). Besides, we found that AR-negative subgroup (Fig 4 C) and long CAG repeat sequences (Fig. 4D) significantly affected the poor survival (5-year OS) in postoperative adjuvant chemotherapy patients with CRC (\( P = 0.002 \) and \( P = 0.004 \), respectively).

**Discussion**

Steroid hormone receptors may participate in the differentiation, proliferation, and progression of CRC tissues (3, 17, 18). Many studies have focused on estrogen and progesterone, which were believed to have an important function in the development of CRC (19-21). Rudolph A (22) found that ER expression is independently related to poor survival. However, the function of androgen and AR in CRC etiology is poorly understood. The function of AR in the colon and rectum carcinogenesis remains unclear. The significance of androgen in tumors was confirmed by different approaches of study (3). Recent studies suggested a protective function of androgens in the colon (23, 24). AR expressions have
been detected in the colorectal mucosa of experimental animals and humans (25, 26). Positive AR expressions were observed in 42.7% of CRC foci and 81.3% of normal mucosa in our study; but this percent was different from that of two studies, which reported lower AR expression in neoplastic mucosa than that in normal colonic mucosa (17, 27). We found a drastic difference between normal and neoplastic mucosa when analyzing AR protein expression.

We hypothesized that polymorphic CAG repeat length can affect AR activity. First, previous studies showed that the CAG repeat length of AR inversely affects its transactivation potential, either as a directly altered receptor function (7) or indirectly reduced AR messenger in RNA and protein levels (9). Second, the transactivation function of AR is dependent on the ligands and androgens. The receptor function also participates in the control of cellular differentiation and proliferation in hormone dependent tissues (4). For example, the androgen signal provides a mitogenic effect in the prostate, where decreasing CAG repeat length of the AR gene increases the cancer risk. By contrast, androgens in breast tissues function in an anti-mitogenic response, in which long CAG repeats indicated a higher risk for male (14) and female (15) patients with breast cancer. Finally, we observed that the CRC group had a longer CAG repeat length compared to the control group. We identified the association between the absence of AR expression and long CAG repeat sequence in our study. We studied 550 CRC patients without somatic alterations in CAG repeats, and demonstrated that long CAG repeats of the AR genotype were associated with higher (1.293 fold in males, 1.255 fold in females and 1.277 fold in total CRC population, respectively) risk for CRC in a Chinese population, whereas a short CAG repeat sequence may provide a certain degree of protection against CRC. Our study was different from that of Slattery ML (13), who found that the AR genotype can decrease the risk for rectal cancer between patients and controls if the genotype possesses a short CAG repeat length. We also obtained a different cut-off point of CAG repeat length, which might attribute to different regions and ethnicities.

We tested the CAG repeat length of each CRC patient in all tumor tissues and paired blood samples. We found that 90% (550 of 610) of CRC samples showed no somatic AR CAG repeat alterations. Our study was similar to that of Ferro P (10), who found 10%
somatic alterations in CAG repeats within the AR gene in colon tumors. First of all, our study investigated the prognostic significance of AR (CAG)n repeat sequence. Our results indicate that patients with different CAG repeat statuses had significantly different survival rates. Therefore, a relatively long CAG repeat sequence was independently associated with poor survival in CRC.

Interestingly, we also found that the absence of AR expression was associated with long CAG repeat sequences that appeared in tumor tissues, and the ratio was 82.9% in males, 92.1% in females, and 86.9% in total patients, respectively. Univariate logistic regression with CRC patients indicated the association of AR negative expression with AR(CAG)n long repeat length in individual patients. The absence of AR expression and long CAG repeat sequences were observed in tumor with large diameter (≥5 cm), moderate/poor differentiation, T3-4 stage, and N1-2 stage compared with positive AR expression and short CAG repeat sequences in CRC patients, but they were not correlated with other clinicopathological features. Further research is necessary to determine the exact mechanism underlying the absence of AR expression and presence of long CAG repeat sequences, and their association with a shorter survival rate in CRC patients. This phenomenon illustrates that CAG repeat lengths could affect the function of AR protein, and alter the biometric traits of CRC. The results left us with a new question on whether or not we may start with polymorphisms to discover the regulatory mechanism of the transcription change of the AR gene, so that we can interfere with this regulatory mechanism to inhibit tumor development.

In this study, we investigated functional polymorphic CAG and protein expression of the AR gene in postoperative adjuvant chemotherapy patients with CRC. The absence of AR expression and long CAG repeat sequences were independently associated with poor survival of postoperative adjuvant chemotherapy patients with CRC. This study indicates that the negative AR subgroup and long CAG repeat sequences were insensitive and tolerant to adjuvant chemotherapy (FOLFOX and XELOX), therefore these patients were prone to recurrence or metastasis and poor prognostic effect.

Although our study generates some important findings, it also has some limitations. Perhaps, there is variation in androgen levels in foods throughout the year. Additionally, dietary data need obtain using a very detailed questionnaire. Besides, we selected cut-points
for the AR gene based on the median data since there is limited information on the most appropriate cut-points to evaluate. Other study limits include that studies examining other polymorphisms and functionality of these genes are needed. Much still remains to be done and more accurate data need to be generated in order to yield more valuable findings in the future.

All in all, our study indicated that absence of AR expression and long CAG repeat sequence were closely related to the development of CRC. Both long CAG and decreased AR expression were correlated with the poor 5-year OS in CRC patients.

Conflict of interest The authors declare that they have no competing interests.

Acknowledgments We would like to thank Ke-Fei Wu and Hong-Tao Song for assistance with immunohistochemistry staining analysis. We thank Jing Song Dr and Jiu-Feng Wei PhD for technical assistance.

Reference
26. Waliszewski P, Błaszczzyk M, Wolinska-Witort E, Drews M, Snochowski M, Hurst RE. Molecular study of sex steroid receptor gene expression in human colon and in colorectal...

Table 1 and 2
Table 1 AR expression and AR (CAG) n repeat length in CRC patients in relation to clinical and pathological features

<table>
<thead>
<tr>
<th>Factor</th>
<th>No. of AR (N=550)</th>
<th>No. of CAG Repeat length</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>AR (-) (%)</td>
<td>AR (+) (%)</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 60</td>
<td>277</td>
<td>160 (51%)</td>
<td>117 (50%)</td>
</tr>
<tr>
<td>≥ 60</td>
<td>273</td>
<td>155 (49%)</td>
<td>118 (50%)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>312</td>
<td>175 (56%)</td>
<td>137 (58%)</td>
</tr>
<tr>
<td>female</td>
<td>238</td>
<td>140 (44%)</td>
<td>98 (42%)</td>
</tr>
<tr>
<td>Largest tumor diameter</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 5cm</td>
<td>289</td>
<td>147 (47%)</td>
<td>142 (60%)</td>
</tr>
<tr>
<td>≥ 5cm</td>
<td>261</td>
<td>168 (53%)</td>
<td>92 (40%)</td>
</tr>
<tr>
<td>Pathological T stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1−2</td>
<td>248</td>
<td>123 (39%)</td>
<td>125 (53%)</td>
</tr>
<tr>
<td>T3−4</td>
<td>302</td>
<td>192 (61%)</td>
<td>110 (47%)</td>
</tr>
<tr>
<td>Pathological N stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N0</td>
<td>237</td>
<td>127 (40%)</td>
<td>110 (46%)</td>
</tr>
<tr>
<td>N1−2</td>
<td>313</td>
<td>188 (60%)</td>
<td>125 (54%)</td>
</tr>
<tr>
<td>Tumor location</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rectum</td>
<td>301</td>
<td>169 (54%)</td>
<td>132 (65%)</td>
</tr>
<tr>
<td>colon</td>
<td>249</td>
<td>146 (46%)</td>
<td>103 (35%)</td>
</tr>
<tr>
<td>Differentiation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>well</td>
<td>238</td>
<td>114 (36%)</td>
<td>124 (53%)</td>
</tr>
<tr>
<td>moderate</td>
<td>229</td>
<td>146 (46%)</td>
<td>83 (35%)</td>
</tr>
<tr>
<td>poor</td>
<td>83</td>
<td>55 (18%)</td>
<td>28 (12%)</td>
</tr>
<tr>
<td>TNM stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>179</td>
<td>108 (34%)</td>
<td>71 (30%)</td>
</tr>
<tr>
<td>II</td>
<td>149</td>
<td>87 (28%)</td>
<td>62 (32%)</td>
</tr>
<tr>
<td>III</td>
<td>222</td>
<td>120 (38%)</td>
<td>102 (30%)</td>
</tr>
<tr>
<td>Growth pattern</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protruded type</td>
<td>267</td>
<td>158 (50%)</td>
<td>109 (46%)</td>
</tr>
<tr>
<td>Ulcerative type</td>
<td>283</td>
<td>157 (50%)</td>
<td>126 (54%)</td>
</tr>
</tbody>
</table>

TNM: Tumor Node Metastasis; AR: androgen receptor

Table 2 Univariate and multivariate analysis of prognosis factors (5Y-OS) for CRC
## Patients

<table>
<thead>
<tr>
<th>Variables</th>
<th>Test cohort (N=290)</th>
<th>Validation cohort (N=260)</th>
<th>Pooled sample (N=550)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR</td>
<td>95% CI</td>
<td>P</td>
</tr>
<tr>
<td><strong>Univariate analysis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (&lt;60/≥ 60)</td>
<td>.368</td>
<td>.975</td>
<td>.935</td>
</tr>
<tr>
<td>Gender( male/ female)</td>
<td>.171</td>
<td>.102</td>
<td>.497</td>
</tr>
<tr>
<td>Largest tumor diameter (≥5cm/ &lt;5cm)</td>
<td>1.53</td>
<td>1.05-2.22</td>
<td>.027</td>
</tr>
<tr>
<td>Pathological T stage (T3-4/T1-2)</td>
<td>2.08</td>
<td>1.37-3.17</td>
<td>.001</td>
</tr>
<tr>
<td>Pathological N stage (N1-2/N0)</td>
<td>1.98</td>
<td>1.34-2.93</td>
<td>.001</td>
</tr>
<tr>
<td>Tumor location (rectum/colon)</td>
<td>.993</td>
<td></td>
<td>.641</td>
</tr>
<tr>
<td>Differentiation (well-moderate/ poor)</td>
<td>.327</td>
<td></td>
<td>.697</td>
</tr>
<tr>
<td>TNM stage (III/ I + II )</td>
<td>.082</td>
<td>1.82</td>
<td>1.25-2.65</td>
</tr>
<tr>
<td>Growth pattern (Ulcerative/Protruded)</td>
<td>.993</td>
<td></td>
<td>.611</td>
</tr>
<tr>
<td>AR (+/-)</td>
<td>2.20</td>
<td>1.49-3.26</td>
<td>.001</td>
</tr>
<tr>
<td>CAG Repeat (Long/Short)</td>
<td>1.63</td>
<td>1.13-2.35</td>
<td>.009</td>
</tr>
<tr>
<td><strong>Multivariate analysis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Largest tumor diameter (≥5cm/ &lt;5cm)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Pathological T stage (T3-4/T1-2)</td>
<td>1.81</td>
<td>1.22-2.69</td>
<td>.003</td>
</tr>
<tr>
<td>Pathological N stage (N1-2/N0)</td>
<td>.232</td>
<td></td>
<td>.158</td>
</tr>
<tr>
<td>AR (+/-)</td>
<td>2.05</td>
<td>1.38-3.04</td>
<td>.001</td>
</tr>
<tr>
<td>CAG Repeat (Long/Short)</td>
<td>1.60</td>
<td>1.07-2.39</td>
<td>.021</td>
</tr>
</tbody>
</table>

**HR:** Hazard Ratio; **TNM:** Tumor Node Metastasis; **AR:** androgen receptor

---

**Figure legends**
Figure 1. AR gene polymorphism analysis using the PCR-based Genescan. (A) Allelotype of short CAG repeat sequence (≤22 CAG repeat, 248 bp) in male and homozygous female; (B) Allelotype of long CAG repeat sequence (>22 CAG repeat, 248 bp) in homozygous female; (C) Allelotype of short CAG repeat sequence (double allele ≤22 CAG repeat, 248 bp) in heterozygous female; (D) Allelotype of long CAG repeat sequence (any allele >22 CAG repeat, 248 bp) in heterozygous female.

Figure 2. Distribution of CAG repeat length among the CRC patients and control groups and the Kaplan–Meier curves of CRC patients affected by CAG repeat length. (A–C) Distribution of CAG repeat length among males, females and total number of CRC patients compared with the control groups, respectively; (D and E) 5-year OS of CRC patients affected by CAG repeat length in the test and validation cohorts; (F–H) 5-year OS of males, females and pooled CRC patients affected by CAG repeat length, respectively.

Figure 3. Immunohistochemical staining of the AR in tumor tissue compared with paired normal tissue and the Kaplan–Meier curves of CRC patients affected by AR expression status. (A) Negative staining in tumor tissue; (B) Positive staining in tumor tissue; (C) Strong positive staining in paired normal tissue; (D and E) 5-year overall survival of CRC patients affected by AR expression status in test and validation cohorts; (F–H) 5-year overall survival of male, female and pooled CRC patients affected by AR expression status, respectively.

Figure 4. Associations between AR expression and CAG repeat length with adjuvant chemotherapy CRC patients. (A) Recurrence or metastasis is illustrated according to AR expression in 296 postoperative adjuvant chemotherapy patients. (B) Recurrence or metastasis is illustrated according to CAG repeat length in 296 postoperative adjuvant chemotherapy patients. (C and D) The prognostic effect of AR expression and CAG repeat length are illustrated in 296 CRC patients who received postoperative adjuvant chemotherapy.
Fig 1

A

B

C

D
Fig 2

A. P = 0.013 (Short: 22; Long: 32)

Frequency

|    | 8  | 9  | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 |
|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| cases (N=500) | 0  | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 | 110 | 120 | 130 | 140 | 150 | 160 | 170 | 180 | 190 | 200 | 210 | 220 | 230 | 240 |
| controls (N=540) | 0  | 5  | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 | 55 | 60 | 65 | 70 | 75 | 80 | 85 | 90 | 95 | 100 | 105 | 110 | 115 | 120 | 125 | 130 |

Number of ARS(AG)n Repeats in Total Population

B. P = 0.009 (Short: 22; Long: 32)

Frequency

|    | 8  | 9  | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 |
|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| cases (N=140) | 0  | 5  | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 | 55 | 60 | 65 | 70 | 75 | 80 | 85 | 90 | 95 | 100 | 105 | 110 | 115 | 120 | 125 | 130 |
| controls (N=160) | 0  | 2  | 4  | 6  | 8  | 10 | 12 | 14 | 16 | 18 | 20 | 22 | 24 | 26 | 28 | 30 | 32 | 34 | 36 | 38 | 40 | 42 | 44 | 46 | 48 | 50 | 52 | 54 |

Number of ARS(AG)n Repeats in Female Population

C. P = 0.001 (Short: 22; Long: 32)

Frequency

|    | 8  | 9  | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 |
|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| cases (N=112) | 0  | 5  | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 | 55 | 60 | 65 | 70 | 75 | 80 | 85 | 90 | 95 | 100 | 105 | 110 | 115 | 120 | 125 | 130 |
| controls (N=126) | 0  | 2  | 4  | 6  | 8  | 10 | 12 | 14 | 16 | 18 | 20 | 22 | 24 | 26 | 28 | 30 | 32 | 34 | 36 | 38 | 40 | 42 | 44 | 46 | 48 | 50 | 52 | 54 |

Number of ARS(AG)n Repeats in Total Population

D. P = 0.009

F.

P = 0.001

E. P = 0.048

G.

P = 0.049

H. P = 0.014

Survival Analysis of Short (CAG)n (N=148) vs Long (CAG)n (N=164)

Survival Analysis of Short (CAG)n (N=280) vs Long (CAG)n (N=290)

Survival Analysis of Short (CAG)n (N=198) vs Long (CAG)n (N=212)

Survival Analysis of Short (CAG)n (N=132) vs Long (CAG)n (N=148)
Fig 4 A

Recurrence or Metastasis

AR- (N=172) 58.1%

AR+ (N=124) 41.9%
P = 0.018

Number

B

Recurrence or Metastasis

Short (CAGm (N=133) 51.4%

Long (CAGm (N=163) 48.6%
P = 0.022

Number

C

P = 0.002

Time (months)

D

P = 0.004

Time (months)
Polymorphic CAG repeat and protein expression of androgen receptor gene in colorectal cancer

Rui Huang, Guiyu Wang, Yanni Song, et al.

Mol Cancer Ther Published OnlineFirst January 30, 2015.

Updated version
Access the most recent version of this article at:
doi:10.1158/1535-7163.MCT-14-0620

Supplementary Material
Access the most recent supplemental material at:
http://mct.aacrjournals.org/content/suppl/2015/01/31/1535-7163.MCT-14-0620.DC1

Author Manuscript
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.