Germline Polymorphisms and Length of Survival of Nasopharyngeal Carcinoma: An Exome-Wide Association Study in Multiple Cohorts

Yun-Miao Guo, Jie-Rong Chen, Yan-Chun Feng, Chua, Melvin L. K., Zeng, Yanni, Edwin Pun Hui, Allen K. C. Chan, Lin-Quan Tang, Lin Wang, Qian Cui, Hui-Qiong Han, Chun-Ling Luo, Guo-Wang Lin, Yan Liang, Yang Liu, Zhong-Lian He, Yu-Xiang Liu, Pan-Pan Wei, Chu-Jun Liu, Wan Peng, Bo-Wei Han, Xiao-Yu Zuo, Enya H. W. Ong, Eugenia L. L. Yeo, Kar Perng Low, Gek San Tan, Tony K. H. Lim, Jacqueline S. G. Hwang, Bo Li, Qi-Sheng Feng, Xiaojun Xia, Yun-Fei Xia, Josephine Ko, Wei Dai, Maria Li Lung, Anthony T. C. Chan, Dennis Y. M. Lo, Mu-Sheng Zeng, Hai-Qiang Mai, Jianjun Liu, Yi-Xin Zeng, Jin-Xin Bei, Hai-Qiang Mai, Jianjun Liu, Yi-Xin Zeng, Jin-Xin Bei

2020-03-06

ChinaRxiv

Abstract

Germline polymorphisms have been linked with differential survival outcomes in cancers but have not been well studied in nasopharyngeal carcinoma (NPC). Here, two-phase association study is conducted to discover germline polymorphisms that are associated with the prognosis of NPC. The discovery phase includes two consecutive hospital cohorts of patients with NPC from Southern China. Exome-wide genotypes at 246,173 single nucleotide polymorphisms (SNPs) are determined, followed by survival analysis for each SNP under Cox proportional hazards regression model. Candidate SNP is replicated in another two independent cohorts from Southern China and Singapore. Meta-analysis of all samples (n = 5,553) confirm that the presence of rs1131636-T, located in the 3′-UTR of RPA1, confers an inferior overall survival (HR = 1.33, 95% CI = 1.20-1.47, P = 6.31 × 10-8). Bioinformatics and biological assays show that rs1131636 has regulatory effects on upstream RPA1. Functional studies further demonstrate that RPA1 promoted the growth, invasion, migration, and radioresistance of NPC cells. Additionally, miR-1253 has been identified as a suppressor for RPA1 expression, likely through regulation of its binding affinity to rs1131636 locus. Collectively, these findings provide a promising biomarker aiding in stratifying patients with poor survival, as well as a potential drug target for NPC.

Full Text

Germline Polymorphisms and Length of Survival of

Nasopharyngeal Carcinoma: An Exome-Wide Association

Study in Multiple Cohorts

Yun-Miao GUO1, Jie-Rong CHEN1,2, Yan-Chun FENG1, Melvin L. K. CHUA3,4, Yanni ZENG1,5,6,7, Edwin Pun HUI8,

Allen K. C. CHAN9, Lin-Quan TANG1, Lin WANG1, Qian CUI2, Hui-Qiong HAN10, Chun-Ling LUO1, Guo-Wang

LIN1, Yan LIANG1, Yang LIU1, Zhong-Lian HE1, Yu-Xiang LIU1, Pan-Pan WEI1, Chu-Jun LIU1, Wan PENG1, Bo-

Wei HAN1, Xiao-Yu ZUO1, Enya H. W. ONG3,4, Eugenia L. L. YEO3,4, Kar Perng LOW3,4, Gek San TAN11, Tony K.

H. LIM12, Jacqueline S. G. HWANG12, Bo LI13,14, Qi-Sheng FENG1, Xiaojun XIA1, Yun-Fei XIA1, Josephine KO15,

Wei DAI15, Maria L. LUNG15, Anthony T. C. CHAN8, Dennis Y. M. LO9, Mu-Sheng ZENG1, Hai-Qiang MAI1*,

Jianjun LIU2,16, *Yi-Xin ZENG1* & Jin-Xin BEI1,17*

1Sun Yat-sen University Cancer Centre, State Key Laboratory of Oncology in South China, Collaborative Innovation

Centre for Cancer Medicine, Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Guangzhou 510060, P. R. China. 2Guangdong Provincial People' s Hospital, Guangdong Academy of Medical Sciences, Guangzhou 510080, P. R. China.

3Division of Radiation Oncology and Division of Medical Sciences, National Cancer Centre Singapore, Singapore

169610. 4Duke-NUS Medical School, Singapore 169857.

5Faculty of Forensic Medicine, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou 510080, P. R.

China. 6Guangdong Province Translational Forensic Medicine Engineering Technology Research Centre，Zhongshan School

of Medicine, Sun Yat-sen University, Guangzhou 510080, P. R. China. 7Guangdong Province Key Laboratory of Brain Function and Disease, Zhongshan School of Medicine, Sun Yat-sen

University, Guangzhou 510080, P. R. China. 8Department of Clinical Oncology, State Key Laboratory of Translational Oncology, The Chinese University of

Hong Kong, Prince of Wales Hospital, Shatin, Hong Kong SAR. 9Department of Chemical Pathology, Li Ka Shing Institute of Health Sciences, State Key Laboratory of Translational

Oncology, The Chinese University of Hong Kong, Hong Kong SAR. 10Department of Oncology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, P. R. China.

11Translational Pathology Centre, Department of Molecular Pathology, Singapore General Hospital, Singapore 169856.

- Corresponding author: Jin-Xin Bei (E-mail: beijx@sysucc.org.cn); Yi-Xin Zeng (E-mail: zengyx@sysucc.org.cn); Jianjun Liu (E-mail: liuj3@gis.a-star.edu.sg); Hai-Qiang Mai (E-mail: maihq@sysucc.org.cn).

Department of Anatomical Pathology, Singapore General Hospital, Singapore 169856. 13Department of Biochemistry and Molecular Biology, Zhongshan School of Medicine, Sun Yat-sen University,

Guangzhou 510080, P. R. China. 14RNA Biomedical Institute, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou 510120, P. R. China.

15Department of Clinical Oncology, University of Hong Kong, Hong Kong SAR.

16Genome Institute of Singapore, Genome Building, Singapore 138672.

17Center for Precision Medicine, Sun Yat-sen University, Guangzhou 510080, P. R. China.

Author contributions

Y.-M. G., J.-R. C., Y.-C. F., M. L. K. C., and Y. Z. contributed equally to this work. J.-X. B., Y.-X. Z., J. L., H.-Q. M.

jointly supervised this work. J.-X.B., Y.-X.Z., J.L., and H.-Q.M. designed and oversaw the study and obtained financial

support. Y.-M.G., J.-R.C., L.-Q.T., L.W., Q.C., H.-Q.H., Y.L., Z.-L.H., Y.-X.L., P.-P.W., E.H.W.O., E.L.L.Y., K.P.L.,

T.K.H.L., J.S.G.H., and Q.-S.F. initiated or participated in sample recruitment and sample preparation. Y.-M.G.,

M.L.K.C., Y.Z., Y.L., C.-J.L., G.-W.L., B.-W.H., X.-Y.Z., and J.-X.B. performed the statistical analyses and

interpreted the data. J.-R.C., Y.-C.F., C.-L.L., and W.P. performed the experiment and analysed the data. E.P.H.,

A.K.C.C., J.K., W.D., M.L.L., A.T.C.C., and D.Y.M.L. participated in annotation of results. Y.-M.G., J.-X.B., Y.-X.Z.,

H.-Q.M., and J.L. drafted the manuscript. M.L.K.C., B.L., X.-J.X., Y.-F.X., M.-S.Z., and J.L. revised the manuscript.

All authors critically reviewed the manuscript and approved the final version.

Abstract

Germline polymorphisms have been linked with differential survival outcomes in cancers but have not been well studied in

nasopharyngeal carcinoma (NPC). Here, two-phase association study is conducted to discover germline polymorphisms

that are associated with the prognosis of NPC. The discovery phase includes two consecutive hospital cohorts of patients

with NPC from Southern China. Exome-wide genotypes at 246,173 single nucleotide polymorphisms (SNPs) are

determined, followed by survival analysis for each SNP under Cox proportional hazards regression model. Candidate SNP

is replicated in another two independent cohorts from Southern China and Singapore. Meta-analysis of all samples (n =

5,553) confirm that the presence of rs1131636-T, located in the 3′-UTR of RPA1, confers an inferior overall survival (HR

= 1.33, 95% CI = 1.20-1.47, P = 6.31 × 10-8). Bioinformatics and biological assays show that rs1131636 has regulatory

effects on upstream RPA1. Functional studies further demonstrate that RPA1 promoted the growth, invasion, migration,

and radioresistance of NPC cells. Additionally, miR-1253 has been identified as a suppressor for RPA1 expression, likely

through regulation of its binding affinity to rs1131636 locus. Collectively, these findings provide a promising biomarker

aiding in stratifying patients with poor survival, as well as a potential drug target for NPC.

Keywords: Nasopharyngeal carcinoma; Prognosis; Overall survival; Single nucleotide polymorphism; RPA1

1. Introduction

Nasopharyngeal carcinoma (NPC) is an Epstein-Barr virus related malignancy with unique ethnic and

geographic distribution, which is prevalent in Southern China, South-eastern Asia and Northern

Africa1. Radiotherapy is the primary therapeutic modality for NPC because of the radiosensitive

nature of its tumour cells and the deep-seated anatomical position. Survival outcomes in patients with

NPC have improved substantially2, largely due to the widespread implementation of intensity-

modulated radiotherapy (IMRT) and the addition of platinum-based chemotherapy in patients with

loco/regionally advanced disease3-5. Nonetheless, current treatment strategy and risk stratification for

clinical outcomes remain confined to the conventional tumour-node-metastasis (TNM) classification

system for NPC6. Patients with the same clinical stage have different outcomes after receiving similar

treatments, indicating heterogeneity among patients with the same clinical stage as categorized by the

TNM system and the limitation of the system in predicting the treatment outcomes7. Disease

recurrence and metastasis are major causes of treatment failure and thus poor survival in NPC.

Therefore, it's important to identify effective biomarkers or indicators and reveal the underlying

mechanisms for precise treatment planning and accurate prognosis of patients with NPC.

Genetic polymorphisms among individuals contribute to their phenotypic differences, such as

disease predisposition, treatment response, and survival8. In NPC, genetic polymorphisms have been

associated with its predisposition and development9. Nonetheless, the exact mechanisms underpinning

how genetic polymorphisms drive tumour behaviour and eventual clinical outcomes are not well

elucidated. Up to date, a few association studies have reported that genetic variants in candidate genes

such as MCP-1, HLA-G, TP53, CELF2, and CXCL12 are associated with prognosis of NPC10-14.

However, the robustness of these findings remains limited by the small study sample sizes and the

restriction of a candidate gene approach.

Here, to investigate germline polymorphisms that might contribute to NPC prognosis, we

conducted a large exome-wide association study of 31,870 common single nucleotide polymorphisms

(SNPs) with survival in two consecutive cohorts from Southern China involving 3,257 patients with

NPC and observed an association of a germline polymorphism in RPA1 gene with survival. The

association was replicated in two additional cohorts involving 2,296 patients with NPC recruited from

Southern China and Singapore. We further demonstrated the functional relevance of RPA1 on tumour

aggression and radioresistance as well as the mechanism of the germline SNP involved in regulating

RPA1 expression.

2. Methodology (Design/Approach)

2.1 Patient recruitment and follow-up

A two-stage design including discovery and replication was adopted for this study. At the

discovery stage, patients with NPC from two consecutive cohorts were recruited at the Sun Yat-sen

University Cancer Centre (SYSUCC; Guangzhou, China); the first cohort of 1,485 patients was

enrolled from March 3, 2003 to December 28, 2007 (SYSUCC-1), and the second cohort of 1,791

patients was enrolled from January 1, 2008 to April 25, 2012 (SYSUCC-2). At the replication stage,

two more independent cohorts were recruited, including 1,751 patients from SYSUCC between April

22, 2008 and June 1, 2015 (SYSUCC-3), and 545 patients from the National Cancer Centre Singapore

(Singapore) between January 1, 2008 and June 14, 2018 (NCCS). For evaluation of gene expression,

additional 132 and 10 patients were recruited at SYSUCC (January 11, 2011 to November 28, 2013)

and NCCS (January 1, 2008 and June 14, 2018), respectively.

All subjects were histologically diagnosed with NPC and subsequently treated at the recruitment

sites. Individuals with history of cancer, radiotherapy, chemotherapy, or any other anti-tumour therapy

prior to first diagnosis were excluded. Clinical staging of NPC was determined according to the 7th

edition of the International Union Against Cancer (UICC) and American Joint Committee on Cancer

(AJCC) staging system15. The follow-up data was collected at three-month intervals for the first two

years and six-monthly thereafter. The date of last follow-up for the discovery samples (SYSUCC-1

and -2) was August 22, 2016; and for the replication cohorts, it was March 15, 2019 for SYSUCC-3,

and February 25, 2019 for NCCS. Local recurrence was confirmed by fiberoptic endoscopy, MRI and

biopsy, whereas distant metastases were diagnosed by clinical symptoms, physical examination, and

imaging methods include computed tomography (CT), bone scan, abdominal sonography, and/or 18F-

fluorodeoxyglucose positron emission tomography-CT (18F-FDG-PET-CT). Serum immunoglobulin

A (IgA) antibodies to Epstein-Barr virus (EBV) capsid antigen (IgA-VCA) and to the EBV early

antigen (IgA-EA) were available for 1,251 individuals from SYSUCC-1 cohort. Written informed

consent was obtained from all patients enrolled in this study, and the study was approved by the local

ethics committees at SYSUCC and NCCS.

2.2 Genotyping and quality control

EDTA-anticoagulant peripheral whole blood sample was obtained from each patient at the time

of diagnosis and before any treatment. Genomic DNA sample was extracted from the blood sample

using Qiagen blood midi or maxi kits (Qiagen, Duesseldorf, Germany). For the two cohorts in the

discovery stage (SYSUCC-1 and SYSUCC-2), genome-wide genotyping of 246,173 SNPs was done

at the Genome Institute of Singapore (Singapore) and the CapitalBio Technology Co., Ltd (Beijing,

China), by using Infinium HumanExome BeadChips (Illumina, San Diego, California, USA), which

contains a total of 246,173 SNPs and mostly exonic SNPs of rare to low frequency in population. As

for SYSUCC-3 cohort in the replication stage, the genotypes of candidate SNP were retrieved from

the genome-wide data, which were obtained by using Infinium Global Screening Array BeadChip

(Illumina, San Diego, California, USA) followed by imputation using the IMPUTE2 program (version

2.3.2) 16 with default parameters. For the NCCS cohort in the replication stage, the genotypes of the

candidate SNP were directly typed by TaqManTM assay (Applied Biosystems, Waltham,

Massachusetts, USA) on CFX384 Real-Time system (Bio-Rad, Hercules, California, USA) according

to the manufacturer' s instruction.

Stringent quality control filters were applied to remove genotype markers with poor quality or

low frequency in both cohorts in the discovery stage, using the PLINK program (version 1.07) 17.

SNPs with genotyping successful rate of less than 95% were excluded, as were SNPs of non-

autosomal, those with minor allele frequency (MAF) less than 1%, and those deviating from Hardy-

Weinberg equilibrium (HWE) test ($P < 1 \times 10\text{-}6$). Similarly, quality control filters were done among

individuals. Samples with low genotyping efficiency less than 95% and with discordant genders to

their recorded ones (implying possible unintended technical error in the sample processing) were

excluded; moreover, samples showing extremes of heterozygosity (inbreeding coefficient estimate F <

-0.01, indicating possible cross-contamination) were also excluded; furthermore, for pairs of

individuals with cryptic relatedness according to identity-by-descent (IBD) calculation implemented

in PLINK (possibly due to unintended technical error or biological relatives). Lastly, ethnic outliers as

revealed by principal component analysis (PCA) of genetic ancestry were removed. Consequently, a

total of 3,257 samples and 31,870 SNPs passing quality control were subjected for further analyses

(Figure S1, Supporting Information).

2.3 Cell culture

The human embryonic kidney cells 293T (HEK293T) and two NPC cell lines (S18 and 5-8F)

were cultivated in Dulbecco' s Modified Eagle Medium (DMEM; Gibco, Grand Island, NY, USA)

supplemented with 10% Fetal Bovine Serum (FBS; Gibco, Grand Island, NY, USA) and maintained

under standard cell culture conditions at 37°C in a water-saturated atmosphere of 5% CO2. HEK293T

was purchased from American Type Culture Collection (ATCC). The two NPC cell lines, 5-8F and

S18, were established and maintained at the SYSUCC. 5-8F was characterized as a subclone with high

tumorigenic and metastatic ability from the parental cell line SUNE-1, which was derived from a

patient with NPC treated at SYSUCC18. S18 was a clone, showing great migration and invasion ability,

picked up from a single cell colony of CNE-2, which was derived from a patient with NPC from

Southern China19,20. Both 5-8F and S18 have lost their EBV genome/episomes upon in vitro

propagations and became EBV negative21.

2.4 RNA extraction and reverse transcription (RT) PCR analysis

Total RNA was isolated using Trizol reagent (Invitrogen, Carlsbad, CA, USA) and reverse

transcription was performed using oligo (dT) priming and M-MLV Reverse Transcriptase according

to the manufacturer' s instructions (Promega, Madison, WI, USA). Real-time quantitative PCR was

performed in the StepOne Real-Time PCR System (Applied Biosystems, Waltham, Massachusetts,

USA) using primer pairs according to gene of interest (Table S1, Supporting Information) and SYBR

Green as DNA dye (Takara, Tokyo, Japan). miRNA level was quantified by stem-loop RT-qPCR

method using Bulge-Loop hsa-miR-1253 Primer Set (RIBOBIO, China). Then the PCR products were

cloned into pMD™19-T Vector (TAKARA, Japan). The plasmids contained miRNA was determined

using Sanger sequencing.

2.5 Luciferase assays

DNA fragments containing RPA1 3′-UTR with either C (RPA1-3UTR*C) or T allele (RPA1-

3UTR*T) at rs1131636 were obtained by using reverse transcribed (RT)-PCR and mRNA extracted

from NPC cells as template (5-8F cell line; primers information were listed at Table S1, Supporting

Information). To examine the targeting effect of miR-1253 on RPA1 3′-UTR region spanning

rs1131636 as predicted by using TargetScan22, a mutant construct of RPA1 3′-UTR was obtained by

replacing the seeding sequence with 5′-CAAAGCGTTTCTTTAgTaCaCtCTTCAATTAATGC-3′

(RPA1-3UTR*MUT). Each of the DNA products were separately subcloned into a luciferase reporter

plasmid at the downstream of luciferase coding region (psiCHECK2; Promega, Madison, WI, USA)

and the sequence validity was confirmed by Sanger sequencing. Next, 5 × 104 cells were seeded into

each well of 24-well plate and simultaneously transfected with respective construct and either miR-

1253 mimics or scrambled control miRNAs (GenePharma, Shanghai, China) using lipofectamine 2000

(Invitrogen, Carlsbad, CA, USA). After being cultured for 36 hours, luciferase activity of the cells was

measured with a Dual-Luciferase Assay Kit (Promega, Madison, WI, USA) according to the

manufacturer' s protocol.

2.6 Stable knockdown and overexpression of RPA1 in cells

Lentiviral expression system was deployed to manipulate the expression of RPA1 in cells. For

knockdown with shRNAs, double-strand oligos (GenePharma, Shanghai, China; sequences were

provided at Table S1, Supporting Information) were inserted to lentiviral vector pLKO.1 following the

protocols from Addgene (http://www.addgene.org/). For overexpression, full length of human RPA1

cDNA was obtained using RT-PCR and mRNA extracted from NPC cells as template (5-8F; primer

information was listed at Table S1, Supporting Information), and subsequently cloned into the pCDH

vector (pCDH-Flag-RPA1); and as a negative control, the gene encoding green fluorescent protein

(GFP) was also constructed into the pCDH vector separately (pCDH-GFP). For lentivirus production,

lentiviral construct was co-transfected with packaging vectors 8.9 and VSVG into HEK293T cells

using Lipofextamine 2000 (Invitrogen, Carlsbad, CA, USA) following the manufacturer' s instructions.

Viral supernatant was collected at 48 hours post transfection and filtered through 0.45μm PVDF filters.

For transfection, lentivirus with polybrene (8μg/ml; Sigma, San Antonio, TX, USA) were added to the

cells; and subsequently the cells were maintained for 48 hours, followed by selection of infected cells

with 2μg/ml puromycin (Solarbio, Beijing, China) for one week.

2.7 Immunoblotting analysis

Proteins were extracted from the whole cell lysates as described previously23. In brief, whole cell

lysates were prepared in cold low-salt lysis buffer with protease inhibitor cocktail (Roche, Basel,

Switzerland), followed by centrifugation at 12,000 rpm for 5 mins at 4℃. Proteins in the supernatant

were then subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis and transferred to

PVDF membranes (Millipore, Billerica, MA, USA). Membranes were subsequently blocked with 5%

defatted milk and incubated with primary anti-RPA1 antibody (sc-28304; Santa Cruz, CA, USA), anti-

Flag antibody (F1804; Sigma-Aldrich, St. Louis, MO, USA) and anti-Actin antibody (A1978; Sigma-

Aldrich, St. Louis, MO, USA), separately.

2.8 Immunohistochemistry

Paraffin-embedded section was deparaffinated and rehydrated, followed by antigen retrieval and

blockage with 5% normal goat serum and subsequent incubation with primary antibody against RPA1

(Santa-Cruz, sc-28304) at 4℃ for overnight. Immunostaining were performed with horseradish

peroxidase (HRP) conjugates by using Dako REAL™ EnVision™ Detection System. Staining

condition were evaluated independently by two pathologists. Protein expression level was analysed

according to staining frequency and intensity. The staining frequency of the protein was semi-

quantitatively scored based on the percentage of positive cells at four quantiles, which was multiplied

by intensity scores as 0, 1, 2, and 3 for no, weak, intermediate and strong staining, respectively, to

obtain immunohistochemical scores.

2.9 Whole exome sequencing and genotype calling

Genomic DNA sample was subjected for library construction using SureSelectXT Human All

Exon + UTRs kit (Agilent Technologies, Santa Clara, CA, USA), followed by sequencing with a

paired-end 2 × 125 bp protocol on a HiSeq instrument (Illumina, San Diego, CA, USA). Sequencing

data was aligned to the human reference genome (GRCh37/hg19) using the BWA software24.

Subsequently, the GATK suite was used for base quality score recalibration, realignment of

insertion/deletions, and variant calling, according to the GATK Best Practices recommendations25.

The genotypes of rs1131636 for the 132 patients from SYSUCC and 10 patients from NCCS were

retrieved from the whole exome data.

2.10 RNA sequencing

RNA sample preparation and RNA sequencing were performed by following the standard

protocols provided in the commercial kits, elsewhere stated. Total RNA was extracted from NPC

tumour tissue (n = 87) by using the RNeasy Mini Kit (Qiagen, Duesseldorf, Germany). Ribosomal

RNAs were depleted using Ribo-Zero Magnetic Kit (Illumina, San Diego, California, USA). Paraffin

embedded (FFPE) primary-recurrent nasopharyngeal tumour pairs were retrieved in the 10 patients

from NCCS and were macro-dissected for RNA extraction. Total RNA was extracted using AllPrep

DNA/RNA FFPE Kit (Qiagen, Duesseldorf, Germany) following the protocol. Library preparation

was performed using TruSeq RNA prep kit (Illumina, San Diego, California, USA). Libraries with

different adaptors were pooled and sequenced using the HiSeq 1500 or Hiseq X Ten instruments,

yielding approximately 25 million pair-ended 125 bp reads or 150 bp reads per sample.

Sequencing data in fastq format was obtained from the instruments, after data conversion from

BCL format using the bcl2fastq software (version 2.18; Illumina, San Diego, California, USA). The

adaptors in the raw fastq data were then trimmed using cutadapt (version 1.11) 26. Read pairs with low

quality were removed, as were reads with high quality but could be aligned to ribosome RNAs using

Bowtie 2 (version 2.3.4) 27. Reads after these filters were realigned to reference human genome (hg19) 28. Finally, the read count for each gene was quantitated using HTseq (version 0.9.1) 29, and expression

level of each gene was normalized as Transcripts per Million reads (TPM) or normalized to log2

transformed transcripts per kilobase million (log2TPM) followed by quantile normalization for

downstream analysis.

2.11 Cell proliferation, migration and invasion assays

For cell proliferation assay, 1 × 103 cells were seeded into each well of 96-well plate. Cell

number and viability were determined at the indicated time using Trypan blue exclusion under a light

microscope. For colony forming assay, 1 × 103 cells were seeded in each well of 6-well plate and

cultured for two weeks. Cells were fixed with methanol and stained with 0.5% crystal violet for 15

mins at room temperature and subsequently the colonies of each well were enumerated under a light

microscope. Wound healing assays were performed using two well silicone inserts (ibidi, Germany)

according to the manufacturer' s instructions. Images were captured by using an inverted microscope

(IX73; Olympus, Tokyo, Japan) and the migration ability of the cells was calculated as the area

measured in pixels between the edges of the scratching at the indicated time point in relative to that of

the starting time point using ImageJ software (version 1.32, https://imagej.nih.gov/ij/). Cell migration

and invasion assays were performed using Transwell chambers (8μm pores; Corning, NY, USA) pre-

coated with or without matrigel (Corning, NY, USA). 5 × 104 cells in serum-free medium were plated

into the top chamber of each insert, whereas 600μL DMEM containing 10% FBS was added to the

lower chambers. Cells were then stained with 0.5% crystal violet after incubation for 20 hours and

subsequently stained cells were enumerated under a light microscope. Each test was performed in

triplicate.

2.12 In vivo xenograft tumour model

Four-week-old male BALB/c nude mice were purchased from Beijing Vital River Laboratory

Animal Technology (Beijing, China). Mice were randomly divided into experimental groups. Cell

suspension containing 1 × 106 of S18 cells per 100μL were subcutaneously injected into dorsal flank

of the nude mice (10 mice per group). Tumour size was measured every three days. Finally, mice were

sacrificed, and tumours were collected for further analysis. Tumour volume was calculated by the

formula as V = Length × Width2 × 0.5230. All animal experiments were performed under protocols

approved by the Institutional Animal Care and Use Committee of Sun Yat-sen University.

2.13 Clonogenic cell survival assay

The clonogenic assays were performed according to a modified protocol as previously

described31. Briefly, 1 × 103 of S18 cells were seeded in each well of a 6-well plate and cultured

overnight, followed by irradiation at 0, 1, 2, 3, 4, and 5 Gy using a Rad Source R2000 X-ray irradiator

(1.1 Gy/mins, 160 kV, 25 mA, 0.3 mm copper filters; Rad Source Tech, USA). After 8~10 days of

incubation, cells were rinsed twice with Phosphate-buffered saline (PBS), and then fixed with

methanol and stained with 0.5% crystal violet. The stained colonies were enumerated. The plating

efficiency (PE) was calculated as the ratio of the number of colonies counted to the number of cells

seeded, multiplying by 100. The surviving fractions (SF) were then determined by dividing the PE of

the treated cells by the PE of the controls (with irradiation dose of 0 Gy) and multiplying by 100.

2.14 Statistical analysis

Patient clinical and pathological characteristics were summarized either as numbers and

percentages of patients for categorical data or as median and range of values for continuous data. The

primary endpoint for this study is overall survival (OS), defined as time from the first treatment after

primary diagnosis to death of any cause. Other secondary endpoints included and their definitions are

as following: disease-free survival (DFS) defined as the time from first treatment to the date of first

recurrence or death of any cause; distant metastasis-free survival (DMFS) defined as the time from

first treatment to the date of first distant metastasis or death of any cause, and local recurrence-free

survival (LRFS) defined as the time from first treatment to the date of first loco/regional recurrence or

death of any cause. Patients who were alive and/or not recorded progression were censored at the date

of last follow-up.

The association between genetic variants and survival was performed using Cox proportional

hazards regression analyses under additive genetic model, with adjustment for known prognostic

covariates, including age, sex, clinical stage, treatment regimen, and population. Genotypes for each

SNP were treated as variables of 0, 1 and 2 representing the homozygote of minor allele, heterozygote,

and the homozygote of major allele, respectively. Population structure was estimated by using PCA,

and the first five principal components were included in the regression analyses. The Cox model

estimated hazard ratio (HR) and 95% confidence interval (CI) for individual sample and meta-analysis

between samples was performed under fixed-effect model using R package survival (version 2.44-1.1) 32 and Metafor (version 2.0-0) 33. Bonferroni correction was used to adjust for multiple comparisons.

Survival curves were estimated using the Kaplan-Meier method. The associations between the variants

of interest and the clinical and pathological characteristics were determined using 2-tailed Chi-square

test for a categorical variable and 2-tailed Student' s t-test for a continuous variable. Linear regression

models were applied to regress rank-transformed IgA titres against VCA and EA, separately, on the

SNP genotypes under additive model, jointly fitted with age, sex, stage and first five genetic principle

components as covariates. The expression quantitative trait loci (eQTL) surveys were conducted

against GTEx portal34 (release V7, http://www.gtexportal.org/home/) and Blood eQTL browser35

(http://genenetwork.nl/bloodeqtlbrowser/). Student' s t-test was applied to compare the difference

between two groups, using GraphPad Prism 7 (version 7.00; GraphPad Software, USA). Wilcoxon

signed-rank test (paired samples) implicated in R program was used to compare gene expression

between the paired pre-radiotherapy tumours and post-radiotherapy recurrences, with a two-sided P

<0.05 for statistical significance.

The clonogenic survival curves were fitted to a linear-quadratic model and compared using the

extra sum-of-squares F-test. Gene set enrichment analysis was performed using the GSEA desktop

application (version 3.0; Broad Institute at MIT, USA). The gene sets implemented were derived from

the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database

(c2.cp.kegg.v6.2.symbols.gmt, 186 gene sets), which was collected in the Molecular Signatures

Database (MSigDB; version 6.2) 36,37. One thousand permutations were taken in GSEA. The input

phenotypes were the high or low expression levels of RPA1 normalized to the mean of RPA1

expression among all the tested samples. For the mRNA sequencing data retrieved from Gene

Expression Omnibus database (series GSE102349; n = 113), same pipeline was applied for above data

processing. $P < 0.05$ was considered to indicate a significant difference, unless otherwise stated. The

key raw data have been deposited in the Research Data Deposit (RDDB2020000816;

rdd.sysucc.org.cn/).

3. Results

3.1 Patient characteristics and survival status

Clinical characteristics of patients with NPC were summarized for all sample collections (Table 1

and Table S2, Supporting Information). At the time of last follow-up, we recorded a total of 777

(14.0%) deaths. Median follow-up duration of patients ranged from 48.8~94.5 months for the cohorts.

We observed superior overall survival in patients who underwent IMRT (Hazard Ratio or HR = 0.49,

95% Confidence Interval or CI = 0.42-0.57, P < 0.0001) and concurrent chemoradiotherapy (CCRT;

HR = 0.76, 95% CI = 0.66-0.88, P = 0.0002), as compared with those did not (Table 1), which is

consistent with the previous findings that utilization of IMRT or CCRT led to superior survival in

patients with NPC3,4.

3.2 Genetic variants associated with NPC prognosis

After stringent quality control filters in the discovery phase, 1,471 patients in the SYSUCC-1 and

1,786 patients in the SYSUCC-2 with genotypes of 31,870 common SNPs were remained for analyses

(Figure S1, Supporting Information). To identify the candidate prognostic markers, survival analyses

were first carried out independently in each cohort using Cox proportional hazards regression model,

followed by meta-analysis of the two cohorts under the fixed-effect model. The quantile-quantile plot

revealed a good match between the distributions of the observed P values and the expected ones by

chance; and a small genomic control factor indicated a minimal inflation of genome-wide association

significance due to the cryptic population stratification (λgc = 1.102; Figure 1B). The meta-analysis

revealed that only one SNP rs1131636, located at the 3′-UTR of RPA1 on chromosome 17, was

significantly associated with overall survival in patients with NPC (HR = 1.35, 95% CI = 1.20-1.51, P

= 7.21 × 10-7; Pheterogeneity = 0.94, I2 = 0%; Figure 1A and 2A), surpassing the exome-wide

significance (P < 1.56 × 10-6 as corrected for multiple tests). Moreover, rs1131636 was significantly

associated with disease-free survival (HR = 1.23, 95% CI = 1.11-1.36, P = 5.09 × 10-5, I2 = 0%,

Pheterogeneity = 0.968) and distant-metastasis-free survival (HR = 1.23, 95% CI = 1.10-1.38, P =

0.0004, I2 = 0%, Pheterogeneity = 0.713), as shown in the Figure S2 and Table S3, Supporting

Information. In addition, rs1131636 was also associated with local-recurrence-free survival of patients

with NPC, with a borderline significance (HR = 1.20, 95% CI = 1.00-1.44, P = 0.0561, I2 = 0%,

Pheterogeneity = 0.878; Figure S2 and Table S3, Supporting Information). Patients carrying the risk

rs1131636-T allele tended to have inferior outcomes as compared to those with rs1131636-[CC]

genotype (Figure 2B). The 5-year overall survival was estimated at 90.3% (95% CI = 88.0%-92.5%)

for patients carrying rs1131636-[CC] genotype, and 83.4% (95% CI = 82.0%-84.9%) for patients with

other genotypes. No statistically significant associations between rs1131636 and the baseline clinical

and pathological characteristics were consistently observed in both sample collections (Table S4,

Supporting Information). In addition, we investigated the relationship between RPA1 genotypes and

EBV IgA titres in a subset of patients with available data (n = 1,266). We observed that rs1131636-T

was significantly associated with higher anti-viral-capsid-antigen (VCA) and anti-early-antigen IgA

titres (P = 0.0187 and P = 0.032, respectively). Consistently, rs11078676 at an intronic region of

RPA1 was reportedly associated with elevated anti-VCA IgA titre in healthy Southern Chinese38.

rs1131636 and rs11078676 are within 32 kb distance in RPA1 and share considerable linkage

disequilibrium (r2 = 0.11 and D' = 0.75), suggesting that these two SNPs are potentially tagging for a

same genetic variant of RPA1.

3.3 Replication of the genetic effect of rs1131636 on NPC prognosis

To replicate the association between rs1131636 and NPC prognosis, we performed survival

analyses in two additional independent samples from Southern China (SYSUCC-3, n = 1,751) and

Singapore (NCCS, n = 545), which revealed consistent trend of association of rs1131636 with overall

survival (Figure 2A). Furthermore, meta-analysis of the two replication cohorts (n = 2,296) revealed a

significant association between rs1131636 and overall survival (HR = 1.27, 95% CI = 1.03-1.58, P =

0.0269), consistent with that observed in the discovery (Figure 2A and Table S3, Supporting

Information). Moreover, meta-analysis of all cohorts in the discovery and replication stages (n = 5,553)

revealed a strong association between rs1131636 and overall survival (HR = 1.33, 95% CI = 1.20-1.47,

P = 6.31 × 10-8; Figure 2A and Table S3, Supporting Information), with the best survival estimates for

the rs1131636-[CC] carriers (Figure 2B). Similar unfavourable effects of rs1131636-T were also

observed on other prognosis measurements (disease-free, distant-metastasis-free, and local-

recurrence-free survivals) in NPC (Figure S2 and Table S3, Supporting Information). Collectively,

these observations suggest that patient of the CT/TT genotype might harbour an aggressive tumour

subtype leading to poor survival. In addition, we observed similar allele frequency of rs1131636-T

between NPC cohorts and Eastern Asian populations with close geography and ancestry; and

interestingly, the highest and lowest frequencies are observed in the European and African/South

Asian populations, respectively, which might suggest the allelic heterogeneity and various linkage

disequilibrium structures in RPA1 locus among the populations (Table S5, Supporting Information).

3.4 Regulatory effect of rs1131636 on gene expression

As rs1131636 is located at the 3′-UTR of RPA1, we assumed its regulatory effect on gene

expression. We performed luciferase reporter assays with a human normal cell line (293T) and NPC

cell lines (5-8F and S18). We observed that cells transfected with the rs1131636-[C] construct had

substantially lower luciferase activity than the cells with the rs1131636-[T] construct, indicating that

the variants exhibited regulatory effects on its upstream gene (Figure S3A and S3B, Supporting

Information). In support, eQTL signals were observed at rs1131636 with a cis effect on RPA1

expression (Table S6 and Table S7, Supporting Information). Immunohistochemistry analysis

revealed higher expression of RPA1 in the samples derived from patients carrying T allele at

rs1131636 (genotypes CT or TT) compared to the samples with CC genotypes (Figure S3C,

Supporting Information).

3.5 RPA1 modulates the proliferation, migration, and invasion of NPC cells

Next, to investigate the functional relevance of RPA1, we manipulated the expression of RPA1 in

NPC cells and assess their abilities of proliferation, wound-healing, migration and invasion (Figure 3A

and Figure S4, Supporting Information). The knockdown of RPA1 significantly decreased cell

proliferation (Figure 3B and Figure S4, Supporting Information) and cell transformation (Figure 3C

and Figure S4, Supporting Information). RPA1-knockdown cells also demonstrated significantly

delayed wound healing and reduced abilities to migrate and invade (Figure 3D, 3E and Figure S4,

Supporting Information). In vivo xenograft model revealed that the shRPA1 xenografts with RPA1

knockdown cells had reduced tumour volume and weight, as compared to the control group, indicating

that the knockdown of RPA1 significantly inhibited orthotopic tumour formation of NPC cells in mice

(Figure 3F). To verify these findings, we introduced exogenous RPA1 in NPC cells and observed that

the overexpression of RPA1 conferred aggressive phenotypes to NPC cells (Figure S5 and Figure S6,

Supporting Information), in contrast to the effects of the knockdown of RPA1 on NPC cells as

abovementioned.

3.6 RPA1 modulates radiation sensitivity of NPC cells

Given the observations that radiosensitization through incremental DNA damage on tumour cells

is highly related to NPC remission39, we performed clonogenic assays to investigate if manipulating

RPA1 has any effect on the radiosensitivity of NPC cells. We observed increased radiation sensitivity

with knockdown of RPA1, and conversely, radiation resistance with overexpression of RPA1 in NPC

cells (Figure 4A). To further validate this finding, we examined the mRNA expression of RPA1 in a

subset of primary-recurrent tumour pairs. The recurrent samples were tumour biopsies following a

course of high-dose radiotherapy upon primary treatment, and thus were truly radioresistant. We

observed that the transcription level of RPA1 mRNA was significantly increased in the recurrent

samples compared to the paired primary tumours before radiotherapy (Figure 4B). Gene Set

Enrichment Analyses (GSEA) in two independent RNA sequencing datasets also revealed a

significant correlation between RPA1 and transcriptional expression of genes involved in repairing

DNA damage including homologous recombination (in-house data: n = 87, NES = 1.93, P < 0.0001;

GSE102349: n = 113, NES = 1.63, P = 0.0299) and nucleotide excision repair (in-house data: NES =

1.99, P < 0.0001; GSE102349: NES = 1.96, P < 0.0001) pathways (Figure S7, Supporting

Information), thereby supporting this functional role of RPA1.

3.7 miR-1253 targeted rs1131636 and suppressed the proliferation and migration of NPC cells

To further investigate how rs1131636 regulates upstream RPA1, we identified that miR-1253

might target a miRNA binding site at the 3' -UTR of RPA1 (717-723bp starting from the stop codon;

Figure 5A), using TargetScan. Luciferase reporter assays showed that the luciferase activity of RPA1

3′-UTR was significantly reduced in the 293T cells transfected with miR-1253 mimics compared to

the control group in a dose-dependent manner (Figure 5B), and the inhibitory effect of miR-1253 was

abolished when mutations were introduced to the targeting site (Figure 5C). Moreover, luciferase

reporter assays also revealed that the transfection of miR-1253 mimics reduced the luciferase activity

of RPA1 3′-UTR with greater inhibitory effect on the construct of rs1131636-C allele (0.57-fold) than

rs1131636-T allele (0.77-fold; Figure 5D). These results suggested that miR-1253 binds to RPA1 3' -

UTR, and the binding affinity is regulated by different alleles of rs1131636. Furthermore, the

transfection of miR-1253 led to down-regulation of RPA1 at the protein level but not mRNA level in

NPC cells (Figure 5E and 5F), suggesting that miR-1253 binds to RPA1 3'-UTR to hinder its protein

translation. This is corroborated by the observations that miR-1253 inhibited the proliferation and

migration of NPC cells (Figure S8, Supporting Information), which was consistent with the

phenotypes observed in the cells with knock-down of RPA1 (Figure 3 and Figure S4, Supporting

Information). In addition, we observed the presence of miR-1253 in NPC tissues and human cell lines

including 5-8F, S18 and 293T (Figure S9, Supporting Information).

4. Discussion & Conclusion

With the bioinformatic analyses and biological evidence in current study, we identified a novel

germline polymorphism of RPA1 gene (rs1131636) conferring tumour progression and therapeutic

resistance in NPC, which ultimately affects the survival of patients with NPC. Patients with CC

genotype at rs1131636 (account for 20% of all NPC patients) had a better survival due to superior

disease control, as compared to patients with CT or TT genotype who might receive further intensified

systematic therapy. We thus propose that rs1131636 at RPA1 could be a promising germline

biomarker to predict therapeutic efficacy and outcome for patients with NPC. Our findings, together

with other similar studies40,41 support a provocative notion that germline polymorphisms may serve as

potential biomarkers for predicting tumour biology and precision medicine.

Given that rs1131636 is located at the 3' -UTR of RPA1, we provided a series of experimental

evidence showing that rs1131636 has regulatory effect on the expression of upstream RPA1, which

thereby might contribute to the progression of NPC. RPA1 (or p70), together with RPA2 (p34) and

RPA3 (p14), forms the heterotrimeric replication protein A (RPA) complex42, which is essential for

maintenance of telomeres43 and regulating cell cycle44, as well as for DNA replication, repair, and

recombination45,46. It has been reported that mutations in RPA1 induced defective DNA double-strand

break repair, chromosomal instability and development of tumour in mice47. Our study revealed that

RPA1 could promote cell proliferation, migration and invasion in NPC cells with both in vitro and in

vivo experiments, which is consistent with previous observations in other cancers48-50. Given that EBV

infection is a well-known hallmark of NPC in endemic regions, we and others observed correlations

between RPA1 polymorphisms (rs1131636 and rs11078676) and IgA titres against EBV antigens in

patients with NPC and healthy individuals, respectively38. These findings implicate that RPA1 might

play an important role in regulating the activation of EBV and host immune response against its

infection51,52, while the underlying mechanisms are yet to be addressed. In addition, previous studies

have demonstrated the expression of RPA1 and RPA2 are potential prognostic factors in multiple

cancers50,53-56, and likewise, RPA3 has been associated with survival of patients with gastric cancer57,

hepatocellular cancer58, and NPC59. These findings suggest that RPA1 and other subunits of RPA

complex might be important modulators for the progression of several human cancers.

Radiotherapy is the primary and curative treatment for NPC60. Our data confirm that RPA1 plays

a role in modulating the radiosensitivity of NPC cells, whereby overexpression of RPA1 leads to

radioresistance. This is consistent with a very recent finding in another NPC cell line61, as well as the

finding in oesophageal cancer62,63. The correlation of RPA1 to the genes involved in crucial DNA

repair pathways as revealed by transcriptome analysis would suggest that NPC tumours with higher

expression or activation of RPA1 might result in a higher repair fidelity of radiation-induced DNA

damage leading to radioresistance64. Next, we identified that miR-1253 can target the seeding region

spanning rs1131636 at the 3' -UTR of RPA1, which suppresses the protein expression of RPA1, but

not transcription. Higher inhibitory effect of C allele of rs1131636 was also observed as compared

with its counterpart T allele, suggesting that the accessibility of miRNA to the locus harbouring

rs1131636-C is higher and thus exerts stronger inhibition on RPA1 translation. Therefore, it is

plausible that the favourable outcomes of NPC patients with CC-genotype may be linked to increased

tumour radiosensitivity and less aggressiveness due to reduced RPA1 expression, as opposed to

patients with the CT or TT genotype. In addition, miR-1253 has been reported as a biomarker for

certain types of cancers, and a regulatory miRNA binding to 3' -UTR of gene targets to promote

tumorigenesis65,66. Nevertheless, it is unclear that how miR-1263 is regulated, although sponging of

miR-1253 by circular RNAs has been demonstrated67.

Our study has a number of strengths. Foremost, to the best of our knowledge, this is the first

study to investigate associations between genetic variants and NPC prognosis, with a substantially

large sample size and maximal coverage of genes by far. Second, the distributions of clinical and

pathological characteristics are consistent with previous reports, revealing unbiased selection. Third,

our discovery withstood a strict threshold for statistical significance that is corrected for multiple tests,

and the clinical associations remained significant after adjustment for known prognostic factors. We

note that the hazard ratios on survival estimates appear to be less significant in the replication stage

compared to the discovery stage, likely reflecting the effect of Winner' s curse68,69, which could be

attributed to different genomic structures of linkage disequilibrium underlying causal variants,

heterogenous factors contributing to disease control and survival outcomes, and smaller sample sizes

in the replication cohorts. Taken together with our consistent observations among all cohorts and the

experimental evidence, these assure that the likelihood of a potential false discovery is low.

Our study also has some limitations. First, our findings are based on a retrospective study in

population of Southern Chinese descendants, further replications in more population and with

prospective design are needed to confirm the significance of rs1131636 as a germline biomarker to

predict NPC prognosis, aiding in stratifying patient participants with more aggressive behaviours and

poor survival for clinical trials. Second, we acknowledge that further studies should be performed to

fine map the genetic variations within RPA1 gene locus and explore their regulatory potentials on

RPA1 expression, as well as the exact role of RPA1 as a potential target of radiosensitizer for

treatment of NPC.

ACKNOWLEDGEMENTS

Y.-M. G., J.-R. C., Y.-C. F., M. L. K. C., and Y. Z. contributed equally to this work. J.-X. B., Y.-X. Z.,

J. L., H.-Q. M. jointly supervised this work. This study is supported by the national Key research and

development program of China (2016YFC0902001), the Sci-Tech Project Foundation of Guangzhou

City (201707020039), Guangdong Innovative and Entrepreneurial Research Team Program

(2016ZT06S638), the National High Technology Research and Development Program of China

(2012AA02A206), the National Science Foundation for Excellent Young Scholars (81222035),

National Program for Support of Top-Notch Young Professionals (J.-X.B.), Chang Jiang Scholars

Program (J.-X.B.), Special Support Program of Guangdong (J.-X.B.) and Sun Yat-sen University

Young Teacher Key Cultivate Project (17ykzd29). M.L.K.C. is supported by the National Medical

Research Council Singapore Clinician-Scientist Award (NMRC/CSA/0027/2018), Ministry of

Education Tier 3 grant (MOE2016-T3-1-004), National Research Foundation Competitive Research

Programme (NRF-CRP17-2017-05) and the Duke-NUS Oncology Academic Program Proton

Research Program. D.Y.M.L. and A.K.C.C. were supported by the Research Grants Council of the

Hong Kong SAR Government under the Theme-based research scheme (T12-401/16-W). M.L.L. was

supported by the Area of Excellence grant (AoE/M-06/08). We thank all the participants in the study,

and staffs at the biobank of SYSUCC for processing sample preparation and staffs at the High-

Throughput Analysis Platform (HTAP) of SYSUCC for data generation and processing. The authors

also acknowledge Hai-Tao Wang for his assistance with result presentation. M.L.K.C. reports personal

fees from Astellas, personal fees from Janssen, grants and personal fees from Ferring, non-financial

support from Astrazeneca, personal fees and non-financial support from Varian, grants from Sanofi

Canada, grants from GenomeDx Biosciences, non-financial support from Medlever, non-financial

support from PVMed, outside the submitted work. D.Y.M.L. is the director of Take2, DRA and

consultant of Grail, and Decheng Capital, and holds equity of Grail, Take2, DRA. E.P.H. reports

Speaker 's honoraria from Merck Sharp & Dohme and Merck Serono, participates as

Consultant/Advisory Boards for Merck Sharp & Dohme, and received Research funding (institution)

from Merck Sharp & Dohme, Pfizer. A.K.C.C. is the director of Take2, DRA and consultant of Grail,

and holds equity of Grail, Take2, DRA. All remaining authors have declared no conflicts of interest.

References

1. Torre LA, Bray F, Siegel RL, et al. Global cancer statistics, 2012. CA: a cancer journal for clinicians 2015;

65(2): 87-108.

2. Yang L, Hong S, Wang Y, et al. Development and External Validation of Nomograms for Predicting Survival

in Nasopharyngeal Carcinoma Patients after Definitive Radiotherapy. Sci Rep 2015; 5: 15638.

3. Lai SZ, Li WF, Chen L, et al. How does intensity-modulated radiotherapy versus conventional two-

dimensional radiotherapy influence the treatment results in nasopharyngeal carcinoma patients? Int J Radiat Oncol Biol

Phys 2011; 80(3): 661-8.

4. Baujat B, Audry H, Bourhis J, et al. Chemotherapy in locally advanced nasopharyngeal carcinoma: an

individual patient data meta-analysis of eight randomized trials and 1753 patients. Int J Radiat Oncol Biol Phys 2006;

64(1): 47-56.

5. Sun Y, Li WF, Chen NY, et al. Induction chemotherapy plus concurrent chemoradiotherapy versus

concurrent chemoradiotherapy alone in locoregionally advanced nasopharyngeal carcinoma: a phase 3, multicentre,

randomised controlled trial. Lancet Oncol 2016; 17(11): 1509-20.

6. Chan AT, Gregoire V, Lefebvre JL, et al. Nasopharyngeal cancer: EHNS-ESMO-ESTRO Clinical Practice

Guidelines for diagnosis, treatment and follow-up. Ann Oncol 2012; 23 Suppl 7: vii83-5.

7. Wang HY, Sun BY, Zhu ZH, et al. Eight-signature classifier for prediction of nasopharyngeal [corrected]

carcinoma survival. J Clin Oncol 2011; 29(34): 4516-25.

8. Lu YF, Goldstein DB, Angrist M, et al. Personalized medicine and human genetic diversity. Cold Spring

Harb Perspect Med 2014; 4(9): a008581.

9. Bei JX, Li Y, Jia WH, et al. A genome-wide association study of nasopharyngeal carcinoma identifies three

new susceptibility loci. Nature genetics 2010; 42(7): 599-603.

10. Tse KP, Tsang NM, Chen KD, et al. MCP-1 Promoter Polymorphism at 2518 is associated with metastasis of

nasopharyngeal carcinoma after treatment. Clinical cancer research : an official journal of the American Association for

Cancer Research 2007; 13(21): 6320-6.

11. Ghandri N, Gabbouj S, Farhat K, et al. Association of HLA-G polymorphisms with nasopharyngeal

carcinoma risk and clinical outcome. Human immunology 2011; 72(2): 150-8.

12. Li ML, Dong Y, Hao YZ, et al. Association between p53 codon 72 polymorphisms and clinical outcome of

nasopharyngeal carcinoma. Genetics and molecular research : GMR 2014; 13(4): 10883-90.

13. Guo YM, Sun MX, Li J, et al. Association of CELF2 polymorphism and the prognosis of nasopharyngeal

carcinoma in southern Chinese population. Oncotarget 2015; 6(29): 27176-86.

14. Chen R, Xu Y, Du X, et al. CXCL12 genetic variants as prognostic markers in nasopharyngeal carcinoma.

OncoTargets and therapy 2015; 8: 2835-42.

15. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer

staging manual and the future of TNM. Annals of surgical oncology 2010; 17(6): 1471-4.

16. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next

generation of genome-wide association studies. PLoS Genet 2009; 5(6): e1000529.

17. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-

based linkage analyses. American journal of human genetics 2007; 81(3): 559-75.

18. Song LB, Yan J, Jian SW, et al. [Molecular mechanisms of tumorgenesis and metastasis in nasopharyngeal

carcinoma cell sublines]. Ai Zheng 2002; 21(2): 158-62.

19. Sizhong Z, Xiukung G, Yi Z. Cytogenetic studies on an epithelial cell line derived from poorly differentiated

nasopharyngeal carcinoma. International journal of cancer 1983; 31(5): 587-90.

20. Qian CN, Berghuis B, Tsarfaty G, et al. Preparing the "soil": the primary tumor induces vasculature

reorganization in the sentinel lymph node before the arrival of metastatic cancer cells. Cancer research 2006; 66(21):

10365-76.

21. Shannon-Lowe C, Adland E, Bell AI, et al. Features distinguishing Epstein-Barr virus infections of epithelial

cells and B cells: viral genome expression, genome maintenance, and genome amplification. Journal of virology 2009;

83(15): 7749-60.

22. Agarwal V, Bell GW, Nam JW, et al. Predicting effective microRNA target sites in mammalian mRNAs.

Elife 2015; 4: e05005.

23. Feng Y, Duan T, Du Y, et al. LRRC25 Functions as an Inhibitor of NF-kappaB Signaling Pathway by

Promoting p65/RelA for Autophagic Degradation. Scientific reports 2017; 7(1): 13448.

24. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics

2009; 25(14): 1754-60.

25. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-

generation DNA sequencing data. Nature genetics 2011; 43(5): 491-8.

26. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. Embnet Journal

2011; 17(1).

27. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012; 9(4): 357-9.

28. Pertea M, Kim D, Pertea GM, et al. Transcript-level expression analysis of RNA-seq experiments with

HISAT, StringTie and Ballgown. Nat Protoc 2016; 11(9): 1650-67.

29. Anders S, Pyl PT, Huber W. HTSeq–a Python framework to work with high-throughput sequencing data.

Bioinformatics 2015; 31(2): 166-9.

30. Zhang S, Zhao BS, Zhou A, et al. m(6)A Demethylase ALKBH5 Maintains Tumorigenicity of Glioblastoma

Stem-like Cells by Sustaining FOXM1 Expression and Cell Proliferation Program. Cancer Cell 2017; 31(4): 591-606

e6.

31. Franken NA, Rodermond HM, Stap J, et al. Clonogenic assay of cells in vitro. Nat Protoc 2006; 1(5): 2315-9.

32. Lin H, Zelterman D. Modeling Survival Data: Extending the Cox Model. Technometrics 2002; 44(1): 85-6.

33. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. Journal of Statistical Software

34.

35. Consortium GT, Laboratory DA, Coordinating Center -Analysis Working G, et al. Genetic effects on gene

expression across human tissues. Nature 2017; 550(7675): 204-13.

35. Westra HJ, Peters MJ, Esko T, et al. Systematic identification of trans eQTLs as putative drivers of known

disease associations. Nat Genet 2013; 45(10): 1238-43.

36. Mootha VK, Lindgren CM, Eriksson KF, et al. PGC-1alpha-responsive genes involved in oxidative

phosphorylation are coordinately downregulated in human diabetes. Nat Genet 2003; 34(3): 267-73.

37. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach

for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005; 102(43): 15545-50.

38. Shen GP, Pan QH, Hong MH, et al. Human genetic variants of homologous recombination repair genes first

found to be associated with Epstein-Barr virus antibody titers in healthy Cantonese. International journal of cancer

2011; 129(6): 1459-66.

39. Lin JC, Jan JS, Hsu CY, et al. Phase III study of concurrent chemoradiotherapy versus radiotherapy alone for

advanced nasopharyngeal carcinoma: positive effect on overall and progression-free survival. Journal of clinical

oncology : official journal of the American Society of Clinical Oncology 2003; 21(4): 631-7.

40. Schutz FA, Pomerantz MM, Gray KP, et al. Single nucleotide polymorphisms and risk of recurrence of renal-

cell carcinoma: a cohort study. Lancet Oncol 2013; 14(1): 81-7.

41. Hua JT, Ahmed M, Guo H, et al. Risk SNP-Mediated Promoter-Enhancer Switching Drives Prostate Cancer

through lncRNA PCAT19. Cell 2018; 174(3): 564-75 e18.

42. Liu T, Huang J. Replication protein A and more: single-stranded DNA-binding proteins in eukaryotic cells.

Acta Biochim Biophys Sin (Shanghai) 2016; 48(7): 665-70.

43. Schramke V, Luciano P, Brevet V, et al. RPA regulates telomerase action by providing Est1p access to

chromosome ends. Nat Genet 2004; 36(1): 46-54.

44. Dodson GE, Shi Y, Tibbetts RS. DNA replication defects, spontaneous DNA damage, and ATM-dependent

checkpoint activation in replication protein A-deficient cells. J Biol Chem 2004; 279(32): 34010-4.

45. Wold MS. Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for

eukaryotic DNA metabolism. Annu Rev Biochem 1997; 66: 61-92.

46. Fanning E, Klimovich V, Nager AR. A dynamic model for replication protein A (RPA) function in DNA

processing pathways. Nucleic acids research 2006; 34(15): 4126-37.

47. Wang Y, Putnam CD, Kane MF, et al. Mutation in Rpa1 results in defective DNA double-strand break repair,

chromosomal instability and cancer in mice. Nat Genet 2005; 37(7): 750-5.

48. Zou Z, Ni M, Zhang J, et al. miR-30a can inhibit DNA replication by targeting RPA1 thus slowing cancer

cell proliferation. Biochem J 2016; 473(14): 2131-9.

49. Ni Z, Yao C, Zhu X, et al. Ailanthone inhibits non-small cell lung cancer cell growth through repressing

DNA replication via downregulating RPA1. Br J Cancer 2017; 117(11): 1621-30.

50. Wang J, Yang T, Chen H, et al. Oncogene RPA1 promotes proliferation of hepatocellular carcinoma via

CDK4/Cyclin-D pathway. Biochem Biophys Res Commun 2018; 498(3): 424-30.

51. Shao JY, Li YH, Gao HY, et al. Comparison of plasma Epstein-Barr virus (EBV) DNA levels and serum

EBV immunoglobulin A/virus capsid antigen antibody titers in patients with nasopharyngeal carcinoma. Cancer 2004;

100(6): 1162-70.

52. Leung SF, Tam JS, Chan AT, et al. Improved accuracy of detection of nasopharyngeal carcinoma by

combined application of circulating Epstein-Barr virus DNA and anti-Epstein-Barr viral capsid antigen IgA antibody.

Clin Chem 2004; 50(2): 339-45.

53. Dahai Y, Sanyuan S, Hong L, et al. A relationship between replication protein A and occurrence and

prognosis of esophageal carcinoma. Cell biochemistry and biophysics 2013; 67(1): 175-80.

54. Givalos N, Gakiopoulou H, Skliri M, et al. Replication protein A is an independent prognostic indicator with

potential therapeutic implications in colon cancer. Modern pathology : an official journal of the United States and

Canadian Academy of Pathology, Inc 2007; 20(2): 159-66.

55. Kanakis D, Levidou G, Gakiopoulou H, et al. Replication protein A: a reliable biologic marker of prognostic

and therapeutic value in human astrocytic tumors. Hum Pathol 2011; 42(10): 1545-53.

56. Levidou G, Gakiopoulou H, Kavantzas N, et al. Prognostic significance of replication protein A (RPA)

expression levels in bladder urothelial carcinoma. BJU Int 2011; 108(2 Pt 2): E59-65.

57. Dai Z, Wang S, Zhang W, et al. Elevated Expression of RPA3 Is Involved in Gastric Cancer Tumorigenesis

and Associated with Poor Patient Survival. Dig Dis Sci 2017; 62(9): 2369-75.

58. Xiao W, Zheng J, Zhou B, et al. Replication Protein A 3 Is Associated with Hepatocellular Carcinoma

Tumorigenesis and Poor Patient Survival. Dig Dis 2018; 36(1): 26-32.

59. Qu C, Zhao Y, Feng G, et al. RPA3 is a potential marker of prognosis and radioresistance for nasopharyngeal

carcinoma. J Cell Mol Med 2017; 21(11): 2872-83.

60. Chua MLK, Wee JTS, Hui EP, et al. Nasopharyngeal carcinoma. Lancet 2016; 387(10022): 1012-24.

61. Zhang Z, Huo H, Liao K, et al. RPA1 downregulation enhances nasopharyngeal cancer radiosensitivity via

blocking RAD51 to the DNA damage site. Exp Cell Res 2018; 371: 330-341.

62. Di Z, Sanyuan S, Hong L, et al. Enhanced radiosensitivity and G2/M arrest were observed in radioresistant

esophageal cancer cells by knocking down RPA expression. Cell Biochem Biophys 2014; 70(2): 887-91.

63. Zhang DJ, Xiang J, Wang X, et al. RPA1 expression in esophageal carcinoma and its influence on

radiosensitivity of esophageal carcinoma TE-1 cells. Panminerva Med 2015; 57(4): 183-9.

64. Sancar A. DNA excision repair. Annu Rev Biochem 1996; 65: 43-81.

65. Gasparini P, Cascione L, Landi L, et al. microRNA classifiers are powerful diagnostic/prognostic tools in

ALK-, EGFR-, and KRAS-driven lung cancers. Proceedings of the National Academy of Sciences of the United States

of America 2015; 112(48): 14924-9.

66. Liu M, Zhang Y, Zhang J, et al. MicroRNA-1253 suppresses cell proliferation and invasion of non-small-cell

lung carcinoma by targeting WNT5A. Cell Death Dis 2018; 9(2): 189.

67. Xu Y, Yao Y, Gao P, et al. Upregulated circular RNA circ_0030235 predicts unfavorable prognosis in

pancreatic ductal adenocarcinoma and facilitates cell progression by sponging miR-1253 and miR-1294. Biochemical

and biophysical research communications 2019; 509(1): 138-42.

68. Lohmueller KE, Pearce CL, Pike M, et al. Meta-analysis of genetic association studies supports a

contribution of common variants to susceptibility to common disease. Nature genetics 2003; 33(2): 177-82.

69. Xiao R, Boehnke M. Quantifying and correcting for the winner' s curse in genetic association studies. Genetic

epidemiology 2009; 33(5): 453-62.

Figure Legends

Figure 1. Results of meta-analysis for 31,870 autosomal SNPs and overall survival time of patients

with NPC. (A) Manhattan plot of P values derived from the meta-analysis in two cohorts, where

survival analyses were conducted with Cox proportional hazards regression under an additive model

adjusted for covariates including age, sex, clinical stage, treatment regimens, and the top five principal

components of population structure. The red line represents the significance level of P value with

correction of multiple comparisons (P =1.56 × 10-6). (B) Quantile-Quantile plot of observed versus

expected P values. No evidence of inflation was observed (inflation factor λ =1.102).

Figure 2. Survival analysis results for rs1131626. (A) Forest plot for results from cox proportional

hazards regression and meta-analysis. CI, confidence interval; P_het: P value from heterogeneity test.

(B) Kaplan–Meier estimates of overall survival curves in patients with NPC grouped by genotypes of

rs1131636 in combined discovery and replication samples. Colours represent patients of different

genotypes and shades represent confidence interval for point estimates of survival curves.

Figure 3. Knockdown of RPA1 inhibited the proliferation, migration, and invasion of S18 NPC cells.

(A) The down-regulated expression of RPA1 in S18 cells transduced with lentivirus carrying pLKO.1-

shRPA1 (shRPA1 1# and shRPA1 2#) was verified by using immunoblotting, as compared to that of

pLKO.1-Scrambled-shRNA (scrm). β-actin served as a loading control. The mRNA expression of

RPA1 was measured by real-time PCR analysis and normalized to β-actin (right). (B) Numbers of S18

cells transfected with respective lentivirus-construct were determined by averaging the cell numbers in

triplicate wells at the indicated time points. (C) Colony formation assay was assessed in cell lines with

indicated lentivirus-construct by crystal violet staining method and the representative images were

shown at top. Histogram showed the quantification of colony in three independent experiments

(bottom). (D) Representative brightfield images for wound-healing assay in S18 cell lines with

indicated lentivirus-constructs as being monitored at 0- or 48-hour (h) timepoint (top). Histogram

showed the relative wound closure rate of three independent experiments (bottom). Scale bar, 100μm.

(E) The migration (left top) and invasion (left bottom) abilities of S18 cells with indicated lentivirus-

constructs were measured by transwell assays without or with Matrigel. Representative images were

shown. Scale bar, 100μm. Histograms showed the fold changes relative to the scrambled cells in three

independent experiments. (F) Xenograft tumours grown in BALB/c nude mice (n = 8 per group) were

shown at left panel, which were subcutaneously injected with S18 cells carrying respective lentivirus-

construct as indicated. The volumes (middle) and the weight (right) of xenograft tumours in nude mice

were also measured. All data are shown as mean ± S.D. from at least three independent experiments.

P < 0.05, ***P < 0.01,*** P < 0.001.

Figure 4. Correlation between RPA1 expression and the radiosensitivity of NPC cells. (A) Clonogenic

cell survival assay showed the fractions of cells survived from irradiation at indicated doses for S18

cells transduced with lentivirus carrying shRNAs knocking down RPA1 or constructs with exogenous

expression of RPA1 or GFP and their respective control constructs. Data points were mean ± S.D. of

the surviving fractions derived from at least three separate experiments. P value derived from two-

tailed student' s t test. (B) Left: Longitudinal change in RPA1 gene expression in primary tumours

before radiotherapy (pre-RT; n = 10) and recurrent tumours after primary radiotherapy (post-RT; n =

10) from the same patients. Coloured circles represent patients with different genotypes of rs1131636

(CT, blue; TT, red; black circle, one patient was not profiled). Right: Box and whisker plot show the

median mRNA abundance of both subgroups. Data of longitudinal recurrent samples were available

for three patients. *P-value derived from Wilcoxon signed-rank test.

Figure 5. miR-1253 targeted rs1131636 and suppressed expression of RPA1. (A) Sequences were

shown for miR-1253 (middle), the predicted targeting site of miR-1253 on the 3′-UTR of RPA1 (top;

RPA1-3UTR*C; # indicates position of rs1131636) and a negative control with four mutations as

indicated (bottom; RPA1-3UTR*MUT). The miR-1253 seed region was boxed. (B) Luciferase

activities in HEK293T cells transfected with increasing amounts of synthetic miR-1253 mimics or

scrambled miRNAs negative control (miR-c). (C) Luciferase activities in HEK293T cells co-

transfected with RPA1-3UTR*C or* MUT reporters and miR-1253 mimics or scrambled miRNAs as

negative control (miR-c). (D) Luciferase activities in HEK293T cells transfected with the psiCHECK2

constructs with either C or T allele at rs1131636 in RPA1 3' -UTR fragment, together with miR-1253

mimics or scrambled miRNAs negative control (miR-c). (E) Western blotting showed the expression

of RPA1 and β-actin as control in the S26, 5-8F and S18 cells transfected with either miR-1253

mimics or scrambled miRNAs negative control (miR-c). (F) The mRNA expression of RPA1 was

measured by real-time PCR analysis in S26, 5-8F and S18 transfected with miR-1253 mimics and

scrambled miRNAs as negative control (miR-c) and normalized to β-actin. All data (B-D and F) are

shown as mean $\pm$ S.D. from at least three independent experiments. $P < 0.05$, ***P < 0.01,*** P <

0.001.

Figure 1. Results of meta-analysis for 31,870 autosomal SNPs and overall survival time of patients with NPC. (A) Manhattan plot of P values derived from the metaanalysis in two cohorts, where survival analyses were conducted with Cox proportional hazards regression under an additive model adjusted for covariates including age, sex, clinical stage, treatment regimens, and the top five principal components of population structure. The red line represents the significance level of P value with correction of multiple comparisons (P =1.56 × 10- 6). (B) Quantile-Quantile plot of observed versus expected P values. No evidence of inflation was observed (inflation factor λ =1.102).

Figure 2. Survival analysis results for rs1131626. (A) Forest plot for results from cox proportional hazards regression and meta-analysis. CI, confidence interval; P_het: P value from heterogeneity test. (B) Kaplan–Meier estimates of overall survival curves in patients with NPC grouped by genotypes of rs1131636 in combined discovery and replication samples. Colours represent patients of different genotypes and shades represent confidence interval for point estimates of survival curves.

Figure 3. Knockdown of RPA1 inhibited the proliferation, migration, and invasion of S18 NPC cells. (A) The down-regulated expression of RPA1 in S18 cells

transduced with lentivirus carrying pLKO.1-shRPA1 (shRPA1 1# and shRPA1 2#) was verified by using immunoblotting, as compared to that of pLKO.1-Scrambled-shRNA (scrm). β-actin served as a loading control. The mRNA expression of RPA1 was measured by real-time PCR analysis and normalized to β-actin (right). (B) Numbers of S18 cells transfected with respective lentivirus-construct were determined by averaging the cell numbers in triplicate wells at the indicated time points. (C) Colony formation assay was assessed in cell lines with indicated lentivirus-construct by crystal violet staining method and the representative images were shown at top. Histogram showed the quantification of colony in three independent experiments (bottom). (D) Representative bright-field images for wound-healing assay in S18 cell lines with indicated lentivirus-constructs as being monitored at 0- or 48-hour (h) time-point (top). Histogram showed the relative wound closure rate of three independent experiments (bottom). Scale bar,

100μm. (E) The migration (left top) and invasion (left bottom) abilities of S18 cells with indicated lentivirusconstructs were measured by transwell assays without or with Matrigel. Representative images were shown. Scale bar, 100μm. Histograms showed the fold changes relative to the scrambled cells in three independent experiments. (F) Xenograft tumours grown in BALB/c nude mice (n = 8 per group) were shown at left panel, which were subcutaneously injected with S18 cells carrying respective lentivirus-construct as indicated. The volumes (middle) and the weight (right) of xenograft tumours in nude mice were also measured. All data are shown as mean ±S.D. from at least three independent experiments. $P < 0.05$, $\boldsymbol{P < 0.01}$, $P < 0.001$.

Figure 4. Correlation between RPA1 expression and the radiosensitivity of NPC cells. (A) Clonogenic cell survival assay showed the fractions of cells survived from irradiation at indicated doses for S18 cells transduced with lentivirus carrying shRNAs knocking down RPA1 or constructs with exogenous expression of RPA1 or GFP and their respective control constructs. Data points were mean ±S.D. of the surviving fractions derived from at least three separate experiments. P value derived from two-tailed student' s t test. (B) Left: Longitudinal change in RPA1 gene expression in primary tumours before radiotherapy (pre-RT; n = 10) and recurrent tumours after primary radiotherapy (post-RT; n = 10) from the same patients. Coloured circles represent patients with different genotypes of rs1131636 (CT, blue; TT, red; black circle, one patient was not profiled). Right: Box and whisker plot show the median mRNA abundance of both subgroups. Data of longitudinal recurrent samples were available for three patients. *P-value derived from Wilcoxon signed-rank test.

Figure 5. miR-1253 targeted rs1131636 and suppressed expression of RPA1. (A) Sequences were shown for miR-1253 (middle), the predicted targeting site of miR-1253 on the 3′-UTR of RPA1 (top; RPA1-3UTR*C; # indicates position of rs1131636) and a negative control with four mutations as indicated (bottom; RPA1-3UTR*MUT). The miR-1253 seed region was boxed. (B) Luciferase activities in HEK293T cells transfected with increasing amounts of synthetic miR-

1253 mimics or scrambled miRNAs negative control (miR-c). (C) Luciferase activities in HEK293T cells co-transfected with RPA1-3UTR*C or* MUT reporters and miR-1253 mimics or scrambled miRNAs as negative control (miR-c). (D) Luciferase activities in HEK293T cells transfected with the psiCHECK2 constructs with either C or T allele at rs1131636 in RPA1 3' -UTR fragment, together with miR-1253 mimics or scrambled miRNAs negative control (miR-c). (E) Western blotting showed the expression of RPA1 and β-actin as control in the S26, 5-8F and S18 cells transfected with either miR-1253 mimics or scrambled miRNAs negative control (miR-c). (F) The mRNA expression of RPA1 was measured by real-time PCR analysis in S26, 5-8F and S18 transfected with miR-1253 mimics and scrambled miRNAs as negative control (miR-c) and normalized to β-actin. All data (B-D and F) are shown as mean ±S.D. from at least three independent experiments. $P < 0.05$, $\boldsymbol{P < 0.01}$, $P < 0.001$.

Table 1. Clinical characteristics and overall survival of patients with NPC.

Discovery cohorts Replication cohorts Combined Characteristics HR (95% CI) P SYSUCC-1 SYSUCC-2 SYSUCC-3 NCCS samples Total 1,471 1,786 1,751 545 5,553 Death 346 (23.5%) 256 (14.3%) 106 (6.1%) 69 (12.7%) 777 (14%) Distant metastasis 197 (13.4%) 186 (10.4%) 161 (9.2%) 72 (13.2%) 616 (11.1%) Locoregional relapse 143 (9.7%) 95 (5.3%) 58 (3.3%) 71 (13%) 367 (6.6%) Duration 2003.3-2007.12 2008.1-2012.4 2008.4-2015.6 2008.1-2018.6 2003.3-2018.6 MST, months (IQR) 94.5 (61.6-102) 48.8 (40.6-58) 53.6 (44.9-62.8) 64.4 (29.5-94.8) 55.6 (43.2-73.5) Gender Male 1074 (73%) 1345 (75.3%) 1278 (73%) 413 (75.8%) 4110 (74%) 0.60 (0.50-0.72) <0.0001 Female 397 (27%) 441 (24.7%) 473 (27%) 132 (24.2%) 1443 (26%) Age (Mean±SD, years) 52.3 (±)11.2 48 (±)11.7 47.2 (±)11.8 51.2 (±)10.9 49.2 (±)11.7 1.82 (1.58-2.11) <0.0001 Tumour classification T1 198 (13.5%) 170 (9.5%) 117 (6.7%) 162 (29.7%) 647 (11.7%) T2 295 (20.1%) 333 (18.6%) 306 (17.5%) 126 (23.1%) 1060 (19.1%) 1.45 (1.33-1.57) <0.0001 T3 700 (47.6%) 871 (48.8%) 924 (52.8%) 156 (28.6%) 2651 (47.7%) T4 278 (18.9%) 412 (23.1%) 404 (23.1%) 101 (18.5%) 1195 (21.5%) Lymph node metastasis N0 360 (24.5%) 233 (13%) 228 (13%) 72 (13.2%) 893 (16.1%) N1 446 (30.3%) 725 (40.6%) 702 (40.1%) 182 (33.4%) 2055 (37%) 1.54 (1.41-1.67) <0.0001 N2 579 (39.4%) 637 (35.7%) 629 (35.9%) 214 (39.3%) 2059 (37.1%) N3 86 (5.8%) 191 (10.7%) 192 (11%) 77 (14.1%) 546 (9.8%) Distant metastasis M0 1441 (98%) 1684 (94.3%) 1712 (97.8%) 540 (99.1%) 5377 (96.8%) 6.13 (4.84-7.75) <0.0001 M1 30 (2%) 102 (5.7%) 39 (2.2%) 5 (0.9%) 176 (3.2%) Clinical stage 80 (5.4%) 59 (3.3%) 44 (2.5%) 41 (7.5%) 224 (4%) 222 (15.1%) 222 (12.4%) 189 (10.8%) 116 (21.3%) 749 (13.5%) 803 (54.6%) 899 (50.3%) 943 (53.9%) 225 (41.3%) 2870 (51.7%) 1.75 (1.64-1.87) <0.0001 A 258 (17.5%) 352 (19.7%) 354 (20.2%) 83 (15.2%) 1047 (18.9%) B 78 (5.3%) 152 (8.5%) 182 (10.4%) 75 (13.8%) 487 (8.8%) C 30 (2%) 102 (5.7%) 39 (2.2%) 5 (0.9%) 176 (3.2%) IMRT No 1234 (83.9%) 747 (41.8%) 0 (0%) 0 (0%) 1981 (35.7%) 0.49 (0.42-0.57) <0.0001 Yes 237 (16.1%) 1039 (58.2%) 1751 (100%) 545 (100%) 3572 (64.3%) CCRT No 903 (61.4%) 554 (31%) 261 (14.9%) 156 (28.6%) 1874 (33.7%) 0.76 (0.66-0.88) 0.0002 Yes 568 (38.6%) 1232 (69%) 1490 (85.1%) 389 (71.4%) 3679 (66.3%) MST, median survival time; IMRT, intensity modulated radiation

therapy; CCRT, concurrent chemoradiotherapy; ICT, induction

chemotherapy; ACT, adjuvant chemotherapy. HR and P values were derived from univariate Cox proportional hazards regression

analyses.

SUPPORTING INFORMASTION

Figure S1. Quality control for study collections. SYSUCC-1/2/3, patient recruitment at Sun Yat-sen University Cancer Centre; NCCS, patient recruitment at National Cancer Centre of Singapore; MAF, minor allele frequency.

Figure S2. Forest plots of the association results for survival status. (A) disease-free survival, (B) distant metastasis-free survival, and (C) local recurrence-free survival other than overall survival. CI, confidence interval; P_het: P value from heterogeneity test.

Figure S3. Luciferase reporter assay showing regulatory potential of the variants at rs1131636. (A) Layout of luciferase reporter constructs; the fragment of 3' -UTR of RPA1 spanning either C or T allele at rs1131636 was inserted downstream of luciferase coding region at psiCHECK2. (B) Luciferase reporter assays were performed in cell lines as indicated, which were transduced with either [C] or [T] constructs as shown above. The Y-axis was relative luciferase activity normalized to that of C allele. Data are shown as mean ± S.D. from three independent experiments. (C) The protein expression of RPA1 in NPC tumour specimens. Box and whisker plot showing the difference in PRA1 protein expression level between the two groups with different genotypes (CC or CT/TT). After the section was immunohistochemistry (IHC) with antibody against RPA1, the expression level of RPA1 was determined based on staining intensity and the percentage of positive cells. Individual genotypes for the matching blood samples were determined using whole-exome sequencing. P-value was generated using two-tailed student' s t test. $\mathbf{P < 0.01}$, *$P < 0.001$.

Figure S4. Knockdown of RPA1 inhibited the proliferation, migration, and invasion of 5-8F cells. (A) The expression of RPA1 in 5-8F cells transduced with lentivirus carrying pLKO.1-shRPA1 (shRPA1 1# and shRPA1 2#) was determined by using immunoblotting, as compared to that of pLKO.1-Scrambled-shRNA (scrm). β-actin was served as a loading control. mRNA expression of RPA1 was measured by real-time PCR analysis and normalized to that of β-actin (right panel). (B) Numbers of 5-8F cells transfected with respective lentivirus-construct were determined by averaging the cell numbers in triplicate wells at the indicated time points. (C) Colony forming assay was assessed in 5-8F cells with indicated lentivirus-construct by crystal violet staining method and the representative images were shown at top. Histogram showed the quantification of colony in three independent experiments (bottom). (D) Wound healing assay of 5-8F cells with indicated lentivirus-constructs as being monitored at 0- or 48-hour (h) time-point (top). Scale bar, 100μm. Histogram at bottom represented the relative wound closure rate of three independent ex-

periments. (E) The migration (left top) and invasion (left bottom) abilities of 5-8F cells with indicated lentivirus-constructs were measured by transwell assays without or with Matrigel. Representative images were shown. Scale bar, 100μm. Histograms showed the fold changes relative to the scrambled cells in three independent experiments. All data are shown as mean ± S.D. from at least three independent experiments. $P < 0.05$, $\boldsymbol{P < 0.01}$, $P < 0.001$.

Figure S5. Exogenous expression of RPA1 promoted the proliferation, migration, and invasion of S18 cells. (A) Cell lysates of S18 transduced with lentivirus carrying construct either pCDH-Flag-RPA1 (RPA1) or pCDH-GFP (GFP) as control were subjected to immunoblot with anti-Flag antibody; and β-actin was used as a loading control. (B) Cell proliferation assay showed the number of S18 cells transduced with indicated lentivirus-construct as measured at each time point. (C) Colony formation assay was assessed in cell lines transduced with indicated lentivirus-construct by crystal violet staining method and the representative images were shown at top. Histogram showed the average number of colonies in three independent experiments (bottom). (D) Representative images of wound healing assay in S18 cells transduced with lentivirus-construct as shown. Scale bar, 100μm. Histogram represents the relative wound closure rate of three independent experiments (right). (E) The migration (left top) and invasion (left bottom) abilities of S18 cells with indicated lentivirus-constructs were measured by transwell assays without or with Matrigel. Representative images were shown. Scale bar, 100μm. Histograms showed fold changes relative to pCDH-GFP cells in three independent experiments. (F) Xenograft tumours grown in BALB/c nude mice (n = 9 per group) were shown at left panel, which were subcutaneously injected with S18 cells carrying respective lentivirus-construct as indicated (RPA1 or GFP). The volumes (middle) and the weight (right) of xenograft tumours in nude mice were also measured, respectively. Data are shown as mean ±S.D. from at least three independent experiments. $P < 0.05$, $\boldsymbol{P < 0.01}$, $P < 0.001$.

Figure S6. Exogenous expression of RPA1 promoted the proliferation, migration, and invasion of 5-8F cells. (A) Immunoblotting revealed the expression of RPA1 in 5-8F cells transduced with lentivirus carrying pCDH-Flag-RPA1(RPA1) or pCDH-GFP(GFP) as indicated. (B) Cell proliferation assay showed the number of 5-8F cells transduced with indicated lentivirus-construct as measured at each time point. (C) Colony forming assay was assessed in cell lines transduced with indicated lentivirus-construct by crystal violet staining method and the representative images were shown at top. Histogram showed the average number of colonies in three independent experiments (bottom). (D) Representative images of wound healing assay in 5-8F cells transduced with lentivirus-construct as shown. Histogram represented the relative wound closure rate of three independent experiments (right). Scale bar, 100μm. (E) The migration (left top) and invasion (left bottom) abilities of 5-8F cells transduced with lentivirus carrying pCDH-Flag-RPA1 or pCDH-GFP were measured by transwell assays without or with Matrigel. Representative images were shown. Scale bar, 100μm. Histograms showed fold changes relative to the pCDH-GFP cells of three inde-

pendent experiments. Data are shown as mean ± S.D. from at least three independent experiments. $P < 0.05$, $\boldsymbol{P < 0.01}$, $P < 0.001$.

Figure S7. Enrichment plots of results from Gene Set Enrichment Analyses (GSEA). Genes in the homologous recombination pathway were analysed in an in-house mRNA sequencing data (A) and GSE102349 dataset (B), as were genes in the excision repair pathway (C: in-house mRNA sequencing data, D: GSE102349 dataset). ES, normalized enrichment score. Genes are ranked by signal/noise ratio according to their differential expression of RPA1. Genes in the selected gene sets are marked with vertical bars, and the enrichment score is shown as green curve.

Figure S8. miR-1253 inhibited the proliferation and migration of NPC cells. (A) Numbers of S26, 5-8F and S18 cells transfected with miR-1253 mimics and miR-c were determined by averaging the cell numbers in triplicate wells at the indicated time points. (B) Colony forming assay was assessed in S26, 5-8F and S18 cells by crystal violet staining method and the representative images were shown. (C) Histograms represented the relative wound closure rate of three independent experiments. (D) The migration abilities of S26, 5-8F and S18 cells transfected with miR-1253 were measured and histograms showed the fold changes relative to the scrambled cells in three independent experiments. Scale bar, 100 μm. Data are shown as mean ± S.D. from three independent experiments. $P < 0.05$, $\boldsymbol{P < 0.01}$, $P < 0.001$.

Figure S9. Figure 4. Sequence alignment of miR-1253 from NPC tumours and human cell

lines. miR-1253 was amplified and was cloned into pMD™19-T Vector (TAKARA, Japan). Plasmids contained

sequences of miR-1253 was determined using Sanger sequencing.

Table S1. Sequences of oligonucleotides used in this study. Experiment Oligo name Sequence RPA1-q-F 5' -AAGGAGCCCGAGTCTCTGAT-3' RPA1-q-R 5' -TGGTGTTACTCCCTCCGACT-3' Primers for qPCR -actin-q-F 5' -CCCACACTGTGCCCATCTAC-3' -actin-q-R 5'-GGAACCGCTCATTGCCAATG-3' RPA1-3' UTR-F 5' -GAGGAGCAGTGCCAATCGGGC-3' RPA1-3' UTR-R 5' -AGAGATTAGCAAGGTTTTAAT-3' Primers for cloning Flag-RPA1-FL-F 5' -ATGGACTACAAAGACGATGACGACAAGCTTAACTGGAAA ACCTTGTATGAGGTC-3'Flag-RPA1-FL-R 5'-TCACATCAATGCACTTCTCCTGATGC-3' scramble 5' -TTCTCCGAACGTGTCACGA-3' Targeting sequences for shRNA shRPA1 1# 5' -GCAATCCAGTGCCCTATAA-3' shRPA1 2# 5' -TTGTTAGCAATCTTCAGGG-3'

Table S2. Clinical characteristics and overall survival in the discovery and replication cohorts. Discovery cohorts Replication cohorts Characteristics SYSUCC-1 SYSUCC-2 SYSUCC-3 NCCS n Death (%) HR (95% CI) P n Death (%) HR (95% CI) P n Death (%) HR (95% CI) P n Death (%) HR (95% CI) P Total 1,471 346 (23.5) 1,786 256 (14.3) 1751 106 (6.1) 545 69 (12.7) Gender Male

1,074 283 (26.4) 0.56 (0.42-0.73) <0.0001 212 (15.8) 0.60 (0.43-0.83) 0.0019 1278 78 (6.1) 0.96 (0.62-1.48) 0.8518 413 62 (15) 0.35 (0.16-0.77) 0.0092 Female 397 63 (15.9) 44 (10) 473 28 (5.9) 132 7 (5.3) Age, years <50 617 101 (16.4) 1.88 (1.49-2.37) <0.0001 1,033 117 (11.3) 1.71 (1.33-2.18) <0.0001 991 45 (4.5) 2.02 (1.38-2.98) 0.0003 249 30 (12) 1.29 (0.80-2.08) 0.2960 >=50 854 245 (28.7) 753 139 (18.5) 760 61 (8) 296 39 (13.2) Tumour classification T1 198 24 (12.1) 170 12 (7.1) 117 4 (3.4) 162 9 (5.6) T2 295 58 (19.7) 1.54 (1.35-1.74) <0.0001 333 39 (11.7) 1.40 (1.20-1.63) <0.0001 306 13 (4.2) 1.56 (1.20-2.02) 0.0009 126 20 (15.9) 1.35 (1.08-1.69) 0.0094 T3 700 158 (22.6) 871 124 (14.2) 924 51 (5.5) 156 25 (16) T4 278 106 (38.1) 412 81 (19.7) 404 38 (9.4) 101 15 (14.9) Lymph node metastasis N0 360 56 (15.6) 233 17 (7.3) 228 3 (1.3) 72 7 (9.7) N1 446 114 (25.6) 1.36 (1.21-1.54) <0.0001 725 67 (9.2) 1.90 (1.64-2.21) <0.0001 702 34 (4.8) 1.87 (1.49-2.36) <0.0001 182 17 (9.3) 1.38 (1.05-1.83) 0.0216 N2 579 139 (24) 637 114 (17.9) 629 45 (7.2) 214 29 (13.6) N3 86 37 (43) 191 58 (30.4) 192 24 (12.5) 77 16 (20.8) Distant metastasis M0 1,441 333 (23.1) 2.76 (1.58-4.80) 0.0003 1,684 208 (12.4) 6.35 (4.63-8.71) <0.0001 1712 91 (5.3) 12.7 (7.33-22.0) <0.0001 540 67 (12.4) 4.70 (1.15-19.2) 0.0314 M1 30 13 (43.3) 102 48 (47.1) 39 15 (38.5) 5 2 (40) Clinical stage 80 4 (5) 59 0 (0) 44 1 (2.3) 41 3 (7.3) 222 34 (15.3) 222 11 (5) 189 3 (1.6) 116 9 (7.8) 803 168 (20.9) 1.62 (1.46-1.79) <0.0001 899 100 (11.1) 1.90 (1.72-2.11) <0.0001 943 38 (4) 2.23 (1.86-2.67) <0.0001 225 28 (12.4) 1.41 (1.15-1.75) 0.0012 A 258 93 (36) 352 59 (16.8) 354 28 (7.9) 83 12 (14.5) B 78 34 (43.6) 152 38 (25) 182 21 (11.5) 75 15 (20) C 30 13 (43.3) 102 48 (47.1) 39 15 (38.5) 5 2 (40) IMRT No 1,234 295 (23.9) 0.84 (0.63-1.13) 0.2574 747 161 (21.6) 0.44 (0.34-0.56) <0.0001 0 0 (0) N/A N/A 0 0 (0) N/A N/A Yes 237 51 (21.5) 1,039 95 (9.1) 1751 106 (6.1) 545 69 (12.7) ICT No 850 186 (21.9) 1.26 (1.02-1.56) 0.0311 913 124 (13.6) 1.12 (0.88-1.43) 0.3595 986 56 (5.7) 1.22 (0.83-1.78) 0.3150 475 57 (12) 0.96 (0.51-1.79) 0.8930 Yes 621 160 (25.8) 873 132 (15.1) 765 50 (6.5) 70 12 (17.1) CCRT No 903 221 (24.5) 0.87 (0.70-1.08) 0.2051 554 91 (16.4) 0.81 (0.62-1.04) 0.0999 261 15 (5.7) 1.09 (0.63-1.88) 0.7536 156 15 (9.6) 1.55 (0.88-2.75) 0.1320 Yes 568 125 (22) 1,232 165 (13.4) 1490 91 (6.1) 389 54 (13.9) ACT No 1,403 327 (23.3) 1.24 (0.78-1.97) 0.3581 1,736 239 (13.8) 2.67 (1.63-4.37) <0.0001 1748 104 (5.9) 18.1 (4.46-73.6) <0.0001 404 52 (12.9) 0.92 (0.54-1.58) 0.7760 Yes 68 19 (27.9) 50 17 (34) 3 2 (66.7) 143 18 (12.6) Chemotherapeutic regimens RT alone 493 110 (22.3) Reference 272 43 (15.8) Reference 165 11 (6.7) Reference 151 15 (9.9) Reference CCRT alone 309 62 (20.1) 0.89 (0.65-1.22) 0.4642 622 76 (12.2) 0.75 (0.52-1.09) 0.1338 819 44 (5.4) 0.79 (0.41-1.54) 0.4940 181 24 (13.3) 1.91 (1.00-3.66) 0.0512 ICT+RT 399 107 (26.8) 1.15 (1.01-1.32) 0.0349 259 41 (15.8) 0.99 (0.80-1.23) 0.9300 95 4 (4.2) 0.78 (0.44-1.39) 0.4070 5 0 (0) 0.00 (0.00-Inf) 0.9980 CCRT+ACT 42 13 (31) 1.15 (0.95-1.39) 0.1519 16 5 (31.3) 1.25 (0.92-1.70) 0.1600 1 1 (100) 2.63 (1.19-4.70) 0.0144 143 18 (12.6) 1.08 (0.86-1.36) 0.5080 RT+ACT 6 1 (16.7) 0.94 (0.57-1.53) 0.7894 3 0 (0) 0.47 (0.01-16.6) 0.6783 1 0 (0) 0.02 (0.00-Inf) 0.9990 0 0 (0) N/A N/A ICT+RT+ACT 5 3 (60) 1.28 (1.01-1.60) 0.0381 20 7 (35) 1.18 (1.01-1.39) 0.0384 0 0 (0) N/A N/A 0 0 (0) N/A N/A ICT+CCRT 202 48 (23.8) 1.01 (0.96-1.07) 0.6277 583 79 (13.6) 0.97 (0.91-1.03) 0.3545 669 45 (6.7) 1.01 (0.91-1.13) 0.8450 65 12 (18.5) 1.05 (0.92-1.19) 0.4510 ICT+CCRT+ACT

15 2 (13.3) 0.92 (0.75-1.12) 0.4156 11 5 (45.5) 1.17 (1.03-1.34) 0.0188 1 1 (100) 1.76 (1.27-2.43) 0.0006 0 0 (0) N/A N/A HR, hazard ratio; CI, confidence interval; RT, radiotherapy; IMRT, intensity modulated radiation therapy; CCRT, concurrent chemoradiotherapy; ICT, induction chemotherapy; ACT, adjuvant chemotherapy. HR and P values were derived from univariate Cox proportional hazards regression analyses.

Table S3. Associations results of rs1131636 with overall survival, disease-free survival, distant metastasis-free survival, and local recurrence-free survival in patients with NPC. Discovery (SYSUCC-1, SYSUCC-2) Replication (SYSUCC-3, NCCS) Meta analysis Survival Characteristics Univariate Multivariate Univariate Multivariate Univariate Multivariate HR (95% CI) P HR (95% CI) P HR (95% CI) P HR (95% CI) P HR (95% CI) P I P Het HR (95% CI) P I2 P Het -7 -5 -8 -6 Gender (Female vs. Male) 0.57 (0.46-0.71) 1.66 × 10 0.63 (0.51-0.77) 1.20 × 10 0.70 (0.48-1.01) 0.059 0.76 (0.52-1.10) 0.1406 0.60 (0.50-0.72) 4.11 × 10 0.00% 0.354 0.65 (0.54-0.79) 5.85 × 10 0.00% 0.384 Age (>=50 vs. <50 years) 1.77 (1.50-2.09) 1.70 × 10-11 1.61 (1.36-1.90) 3.24 × 10-8 1.13 (0.87-1.46) 0.353 1.73 (1.28-2.34) 0.0004 1.76 (1.53-2.04) 1.95 × 10-14 0.00% 0.938 1.63 (1.41-1.89) 5.34 × 10-11 0.00% 0.673 Clinical stage (Advanced vs. Early) 1.74 (1.62-1.87) 2.94 × 10-52 1.76 (1.63-1.90) 2.12 × 10-50 1.84 (1.60-2.11) 1.70 × 10-17 2.00 (1.72-2.32) 5.41 × 10-20 1.76 (1.65-1.87) 6.00 × 10-68 0.00% 0.494 1.80 (1.69-1.93) 3.57 × 10-68 56.2% 0.131 OS IMRT (Yes vs. No) 0.62 (0.51-0.75) 7.27 × 10 0.70 (0.57-0.85) 0.0003 N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A ICT (Yes vs. No) 1.20 (1.02-1.41) 0.0254 0.79 (0.66-0.93) 0.0054 1.05 (0.77-1.44) 0.7418 0.76 (0.54-1.06) 0.1036 1.17 (1.01-1.35) 0.0325 0.00% 0.466 0.78 (0.67-0.91) 0.0013 0.00% 0.849 CCRT (Yes vs. No) 0.86 (0.73-1.01) 0.0582 0.81 (0.69-0.96) 0.0172 1.24 (0.84-1.84) 0.2849 0.64 (0.42-0.98) 0.0411 0.90 (0.78-1.05) 0.1762 65.5% 0.088 0.79 (0.67-0.92) 0.0030 1.18% 0.314 ACT (Yes vs. No) 1.68 (1.20-2.35) 0.0027 1.52 (1.08-2.14) 0.0175 1.29 (0.80-2.08) 0.3045 0.95 (0.57-1.59) 0.8582 1.54 (1.17-2.02) 0.0023 0.00% 0.377 1.31 (0.99-1.75) 0.0614 54.1% 0.140 -8 -7 -8 -8 rs1131636 (Additive model) 1.38 (1.23-1.55) 6.28 × 10 1.35 (1.20-1.51) 7.21 × 10 1.22 (0.98-1.51) 0.0741 1.27 (1.03-1.58) 0.0269 1.34 (1.21-1.49) 2.07 × 10 3.57% 0.309 1.33 (1.20-1.47) 6.31 × 10 0.00% 0.970 Gender (Female vs. Male) 0.64 (0.54-0.77) 9.25 × 10-7 0.69 (0.58-0.82) 3.60 × 10-5 0.74 (0.59-0.95) 0.0153 0.77 (0.60-0.98) 0.0314 0.68 (0.59-0.78) 7.07 × 10-8 0.00% 0.340 0.71 (0.62-0.82) 4.15 × 10-6 0.00% 0.467 Age (>=50 vs. <50 years) 1.40 (1.21-1.61) 3.00 × 10-6 1.26 (1.09-1.45) 0.0015 1.26 (1.03-1.53) 0.0221 1.22 (1.00-1.49) 0.0503 1.35 (1.20-1.51) 3.29 × 10-7 0.00% 0.406 1.25 (1.11-1.40) 0.0002 0.00% 0.797 -40 -41 -18 -21 -57 -61 Clinical stage (Advanced vs. Early) 1.54 (1.45-1.64) 2.30 × 10 1.57 (1.47-1.68) 1.60 × 10 1.53 (1.39-1.68) 4.30 × 10 1.64 (1.48-1.82) 4.44 × 10 1.54 (1.46-1.62) 9.23 × 10 0.00% 0.884 1.59 (1.50-1.68) 8.70 × 10 0.00% 0.505 -10 -6 DFS IMRT (Yes vs. No) 0.61 (0.52-0.71) 6.00× 10 0.67 (0.57-0.79) 3.00× 10 N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A ICT (Yes vs. No) 1.11 (0.97-1.28) 0.1363 0.81 (0.70-0.94) 0.0043 1.16 (0.95-1.42) 0.1532 0.90 (0.72-1.12) 0.3472 1.13 (1.00-1.26) 0.0416 0.00% 0.741 0.83 (0.74-0.94) 0.0038 0.00% 0.418 CCRT (Yes vs. No) 0.80 (0.69-0.92) 0.0015 0.78 (0.68-0.91) 0.0014 0.95 (0.74-

1.23) 0.7164 0.59 (0.45-0.77) 0.0001 0.83 (0.74-0.94) 0.0031 33.1% 0.221 0.73 (0.64-0.84) 3.35×10-6 70.1% 0.068 ACT (Yes vs. No) 1.57 (1.15-2.14) 0.0047 1.43 (1.04-1.96) 0.027 1.23 (0.86-1.77) 0.255 1.10 (0.75-1.61) 0.6194 1.41 (1.12-1.79) 0.0039 0.00% 0.328 1.28 (1.01-1.64) 0.0436 6.14% 0.302 rs1131636 (Additive model) 1.26 (1.14-1.39) 1.00 × 10-5 1.23 (1.11-1.36) 5.09 × 10-5 1.02 (0.88-1.17) 0.8379 1.04 (0.90-1.20) 0.604 1.17 (1.08-1.27) 0.0002 82.7% 0.016 1.16 (1.07-1.26) 0.0003 20.4% 0.290 Gender (Female vs. Male) 0.62 (0.51-0.76) 5.00 × 10-6 0.69 (0.56-0.84) 0.0003 0.71 (0.52-0.97) 0.0329 0.73 (0.53-1.01) 0.054 0.65 (0.54-0.77) 5.42 × 10-7 0.00% 0.504 0.70 (0.59-0.83) 4.89 × 10-5 0.00% 0.750 Age (>=50 vs. <50 years) 1.54 (1.31-1.81) 1.60×10-7 1.38 (1.17-1.63) 0.0001 0.98 (0.76-1.27) 0.8671 0.96 (0.74-1.25) 0.7549 1.36 (1.18-1.55) 1.29×10-5 88.2% 0.004 1.25 (1.09-1.43) 0.0017 81.5% 0.020 -51 -50 -13 -12 Clinical stage (Advanced vs. Early) 1.71 (1.60-1.84) 6.58 × 10 1.73 (1.61-1.86) 2.27 × 10 1.57 (1.39-1.78) 8.49 × 10 1.61 (1.41-1.85) 3.08 × 10 1.68 (1.58-1.78) 9.55 × 10-62 30.8% 0.229 1.71 (1.60-1.82) 8.67×10-61 0.00% 0.355 DMFS IMRT (Yes vs. No) 0.55 (0.46-0.66) 1.92×10-10 0.61 (0.51-0.74) 6.19×10-7 N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A ICT (Yes vs. No) 1.16 (0.99-1.35) 0.0726 0.80 (0.67-0.94) 0.0074 1.20 (0.93-1.56) 0.1654 0.98 (0.73-1.31) 0.8835 1.17 (1.02-1.34) 0.0242 0.00% 0.795 0.84 (0.73-0.97) 0.0167 31.9% 0.226 CCRT (Yes vs. No) 0.83 (0.71-0.97) 0.0188 0.82 (0.70-0.97) 0.0221 1.20 (0.84-1.70) 0.3182 0.69 (0.47-1.01) 0.0535 0.88 (0.76-1.02) 0.0832 71.5% 0.061 0.80 (0.69-0.93) 0.0040 0.00% 0.403 ACT (Yes vs. No) 1.85 (1.33-2.57) 0.0002 1.63 (1.16-2.27) 0.0044 1.83 (1.21-2.76) 0.004 1.65 (1.07-2.57) 0.0247 1.84 (1.42-2.38) 3.17 × 10-6 0.00% 0.966 1.64 (1.25-2.13) 0.0003 0.00% 0.950 rs1131636 (Additive model) 1.26 (1.13-1.42) 6.40 × 10-5 1.23 (1.10-1.38) 0.0004 0.97 (0.81-1.17) 0.7717 0.99 (0.82-1.19) 0.891 1.17 (1.07-1.29) 0.0012 82.2% 0.018 1.16 (1.05-1.28) 0.0033 39.4% 0.180 Gender (Female vs. Male) 0.72 (0.52-0.98) 0.0362 0.72 (0.53-0.99) 0.0436 0.73 (0.48-1.11) 0.1432 0.74 (0.48-1.13) 0.1655 0.72 (0.56-0.93) 0.0106 0.00% 0.951 0.73 (0.57-0.94) 0.0144 0.00% 0.935 Age (>=50 vs. <50 years) 1.13 (0.88-1.46) 0.3535 1.06 (0.82-1.37) 0.6583 1.34 (0.95-1.90) 0.0952 1.25 (0.89-1.78) 0.2035 1.20 (0.98-1.47) 0.0824 0.00% 0.428 1.12 (0.91-1.38) 0.2660 0.00% 0.448 Clinical stage (Advanced vs. Early) 1.09 (0.96-1.24) 0.1706 1.10 (0.96-1.25) 0.1837 1.11 (0.93-1.32) 0.2504 1.32 (1.09-1.59) 0.0039 1.10 (0.99-1.22) 0.0744 0.00% 0.902 1.17 (1.05-1.30) 0.0058 59.4% 0.116 IMRT (Yes vs. No) 0.80 (0.61-1.06) 0.1138 0.87 (0.65-1.17) 0.3551 N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A LRFS ICT (Yes vs. No) 1.14 (0.89-1.47) 0.3059 1.04 (0.79-1.37) 0.7712 0.79 (0.54-1.15) 0.209 0.74 (0.49-1.11) 0.1404 1.02 (0.82-1.25) 0.8843 61.6% 0.107 0.93 (0.74-1.17) 0.5626 48.0% 0.166 CCRT (Yes vs. No) 0.75 (0.58-0.97) 0.0269 0.76 (0.57-0.99) 0.0464 0.59 (0.40-0.87) 0.0076 0.45 (0.29-0.70) 0.0004 0.70 (0.56-0.86) 0.0009 1.16% 0.314 0.65 (0.52-0.82) 0.0003 74.1% 0.050 ACT (Yes vs. No) 1.07 (0.55-2.09) 0.8373 1.10 (0.56-2.15) 0.7893 1.21 (0.66-2.23) 0.5449 1.17 (0.61-2.23) 0.6358 1.14 (0.73-1.79) 0.5585 0.00% 0.796 1.13 (0.71-1.80) 0.5983 0.00% 0.893 rs1131636 (Additive model) 1.20 (0.99-1.44) 0.0529 1.20 (1.00-1.44) 0.0561 1.16 (0.91-1.49) 0.24 1.14 (0.89-1.46) 0.311 1.19 (1.02-1.37) 0.0241 0.00% 0.838 1.17 (1.01-1.36) 0.0325 0.00% 0.920 HR, hazard ratio; CI, confidence interval; OS, overall survival; DFS, disease-free survival; DMFS, distant metastasis-free survival; LRFS,

localregional recurrence-free survival. HR and P values for discovery and replication cohorts were derived from the Cox proportional hazards regression analyses; Meta-analysis was performed under fixed-effect model.

Table S4. Distribution of rs1131636 genotypes stratified by clinical characteristics in the discovery cohorts, validation cohorts and combined samples. Discovery cohorts Replication cohorts Combined samples SYSUCC-1 SYSUCC-2 SYSUCC-3 NCCS Characteristics rs1131636, n (%) rs1131636, n (%) rs1131636, n (%) rs1131636, n (%) rs1131636, n (%) n P n P n P n P n P CC CT TT CC CT TT CC CT TT CC CT TT CC CT TT Total 1471 293 (19.9) 741 (50.4) 437 (29.7) 1786 364 (20.4) 914 (51.2) 508 (28.4) 1751 356 (20.3) 882 (50.4) 513 (29.3) 545 101 (18.5) 264 (48.4) 180 (33.0) 5553 1114 (20.1) 2801 (50.4) 1638 (29.5) Gender Male 1074 212 (19.7) 539 (50.2) 323 (30.1) 0.8728 1345 262 (19.5) 688 (51.2) 395 (29.4) 0.149 1278 251 (19.6) 645 (50.5) 382 (29.9) 0.4299 413 78 (18.9) 197 (47.7) 138 (33.4) 0.8247 4110 803 (19.5) 2069 (50.3) 1238 (30.1) 0.1176 Female 397 81 (20.4) 202 (50.9) 114 (28.7) 441 102 (23.1) 226 (51.2) 113 (25.6) 473 105 (22.2) 237 (50.1) 131 (27.7) 132 23 (17.4) 67 (50.8) 42 (31.8) 1443 311 (21.6) 732 (50.7) 400 (27.7) Age, years <50 617 122 (19.8) 317 (51.4) 178 (28.8) 0.7837 1033 219 (21.2) 518 (50.1) 296 (28.7) 0.508 991 202 (20.4) 507 (51.2) 282 (28.5) 0.658 249 48 (19.3) 122 (49) 79 (31.7) 0.8182 2890 591 (20.4) 1464 (50.7) 835 (28.9) 0.5334 >=50 854 171 (20) 424 (49.6) 259 (30.3) 753 145 (19.3) 396 (52.6) 212 (28.2) 760 154 (20.3) 375 (49.3) 231 (30.4) 296 53 (17.9) 142 (48) 101 (34.1) 2663 523 (19.6) 1337 (50.2) 803 (30.2) Tumour classification T1 198 46 (23.2) 108 (54.5) 44 (22.2) 170 33 (19.4) 87 (51.2) 50 (29.4) 117 24 (20.5) 54 (46.2) 39 (33.3) 162 30 (18.5) 77 (47.5) 55 (34) 647 133 (20.6) 326 (50.4) 188 (29.1) T2 295 51 (17.3) 150 (50.8) 94 (31.9) 0.1842 333 68 (20.4) 175 (52.6) 90 (27) 0.2315 306 69 (22.5) 148 (48.4) 89 (29.1) 0.7236 126 26 (20.6) 57 (45.2) 43 (34.1) 0.5856 1060 214 (20.2) 530 (50) 316 (29.8) 0.2451 T3 700 140 (20) 353 (50.4) 207 (29.6) 871 195 (22.4) 441 (50.6) 235 (27) 924 183 (19.8) 481 (52.1) 260 (28.1) 156 33 (21.2) 76 (48.7) 47 (30.1) 2651 551 (20.8) 1351 (51) 749 (28.3) T4 278 56 (20.1) 130 (46.8) 92 (33.1) 412 68 (16.5) 211 (51.2) 133 (32.3) 404 80 (19.8) 199 (49.3) 125 (30.9) 101 12 (11.9) 54 (53.5) 35 (34.7) 1195 216 (18.1) 594 (49.7) 385 (32.2) Lymph node metastasis N0 360 69 (19.2) 192 (53.3) 99 (27.5) 233 37 (15.9) 125 (53.6) 71 (30.5) 228 45 (19.7) 110 (48.2) 73 (32.1) 72 15 (20.8) 34 (47.2) 23 (31.9) 893 166 (18.6) 461 (51.6) 266 (29.8) N1 446 95 (21.3) 227 (50.9) 124 (27.8) 0.0949 725 167 (23) 364 (50.2) 194 (26.8) 0.2876 702 142 (20.2) 353 (50.3) 207 (29.5) 0.9071 182 37 (20.3) 80 (44) 65 (35.7) 0.7522 2055 441 (21.5) 1024 (49.8) 590 (28.7) 0.4151 N2 579 106 (18.3) 291 (50.3) 182 (31.4) 637 126 (19.8) 326 (51.2) 185 (29) 629 125 (19.9) 322 (51.2) 182 (28.9) 214 34 (15.9) 111 (51.9) 69 (32.2) 2059 391 (19) 1050 (51) 618 (30) N3 86 23 (26.7) 31 (36) 32 (37.2) 191 34 (17.8) 99 (51.8) 58 (30.4) 192 44 (22.9) 97 (50.5) 51 (26.6) 77 15 (19.5) 39 (50.6) 23 (29.9) 546 116 (21.2) 266 (48.7) 164 (30.1) Distant metastasis M0 1441 288 (20) 725 (50.3) 428 (29.7) 0.8973 1684 355 (21.1) 858 (51) 471 (28) 0.0078 1712 345 (20.2) 861 (50.3) 506 (29.5) 0.2159 540 99 (18.3) 263 (48.7) 178 (33) 0.3376 5377 1087 (20.2) 2707 (50.3) 1583 (29.5) 0.2829 M1 30 5 (16.7) 16 (53.3) 9 (30) 102 9 (8.8) 56 (54.9) 37 (36.3) 39 11 (28.3) 21 (53.8) 7 (17.9) 5 2 (40) 1 (20) 2 (40) 176

27 (15.3) 94 (53.4) 55 (31.3) Clinical stage 80 19 (23.8) 41 (51.3) 20 (25) 59 7 (11.9) 31 (52.5) 21 (35.6) 44 9 (20.5) 17 (38.6) 18 (40.9) 41 10 (24.4) 21 (51.2) 10 (24.4) 224 45 (20.1) 110 (49.1) 69 (30.8) 222 43 (19.4) 120 (54.1) 59 (26.6) 222 43 (19.4) 118 (53.2) 61 (27.5) 189 45 (23.8) 87 (46) 57 (30.2) 116 22 (19) 51 (44) 43 (37.1) 749 153 (20.4) 376 (50.2) 220 (29.4) 803 153 (19.1) 417 (51.9) 233 (29) 0.1645 899 211 (23.5) 457 (50.8) 231 (25.7) 0.0093 943 182 (19.3) 489 (51.9) 272 (28.8) 0.3345 225 43 (19.1) 110 (48.9) 72 (32) 0.626 2870 589 (20.5) 1473 (51.3) 808 (28.2) 0.2386 A 258 52 (20.2) 121 (46.9) 85 (32.9) 352 62 (17.6) 174 (49.4) 116 (33) 354 68 (19.2) 174 (49.2) 112 (31.6) 83 10 (12) 42 (50.6) 31 (37.4) 1047 192 (18.3) 511 (48.8) 344 (32.9) B 78 21 (26.9) 26 (33.3) 31 (39.7) 152 32 (21.1) 78 (51.3) 42 (27.6) 182 41 (22.5) 94 (51.6) 47 (25.9) 75 14 (18.7) 39 (52) 22 (29.3) 487 108 (22.2) 237 (48.7) 142 (29.1) C 30 5 (16.7) 16 (53.3) 9 (30) 102 9 (8.8) 56 (54.9) 37 (36.3) 39 11 (28.2) 21 (53.8) 7 (18) 5 2 (40) 1 (20) 2 (40) 176 27 (15.3) 94 (53.4) 55 (31.3) IMRT No 1234 243 (19.7) 623 (50.5) 368 (29.8) 0.8826 747 145 (19.4) 390 (52.2) 212 (28.4) 0.6498 0 0 (0) 0 (0) 0 (0) N/A 0 0 (0) 0 (0) 0 (0) N/A 1981 388 (19.6) 1013 (51.1) 580 (29.3) 0.7077 Yes 237 50 (21.1) 118 (49.8) 69 (29.1) 1039 219 (21.1) 524 (50.4) 296 (28.5) 1751 356 (20.3) 882 (50.4) 513 (29.3) 545 101 (18.5) 264 (48.4) 180 (33) 3572 726 (20.3) 1788 (50.1) 1058 (29.6) ICT No 850 169 (19.9) 437 (51.4) 244 (28.7) 0.5735 913 191 (20.9) 471 (51.6) 251 (27.5) 0.6302 986 200 (20.3) 484 (49.1) 302 (30.6) 0.3488 475 91 (19.2) 229 (48.2) 155 (32.6) 0.6073 3224 651 (20.2) 1621 (50.3) 952 (29.5) 0.9473 Yes 621 124 (20) 304 (49) 193 (31.1) 873 173 (19.8) 443 (50.7) 257 (29.4) 765 156 (20.4) 398 (52) 211 (27.8) 70 10 (14.3) 35 (50) 25 (35.7) 2329 463 (19.9) 1180 (50.7) 686 (29.4) CCRT No 903 177 (19.6) 458 (50.7) 268 (29.7) 0.9166 554 100 (18.1) 293 (52.9) 161 (29.1) 0.2583 261 61 (23.4) 136 (52.1) 64 (24.5) 0.1376 156 37 (23.7) 68 (43.6) 51 (32.7) 0.1199 1874 375 (20) 955 (51) 544 (29) 0.833 Yes 568 116 (20.4) 283 (49.8) 169 (29.8) 1232 264 (21.4) 621 (50.4) 347 (28.2) 1490 295 (19.8) 746 (50.1) 449 (30.1) 389 64 (16.5) 196 (50.4) 129 (33.2) 3679 739 (20.1) 1846 (50.2) 1094 (29.7) ACT No 1403 277 (19.7) 706 (50.3) 420 (29.9) 0.6018 1736 352 (20.3) 891 (51.3) 493 (28.4) 0.7248 1748 355 (20.3) 881 (50.4) 512 (29.3) 0.8026 402 76 (18.9) 191 (47.5) 135 (33.6) 0.7671 5289 1060 (20) 2669 (50.5) 1560 (29.5) 0.984 Yes 68 16 (23.5) 35 (51.5) 17 (25) 50 12 (24) 23 (46) 15 (30) 3 1 (33.3) 1 (33.3) 1 (33.3) 143 25 (17.5) 73 (51) 45 (31.5) 264 54 (20.5) 132 (50) 78 (29.5) IMRT, intensity modulated radiation therapy;CCRT, concurrent chemoradiotherapy; ICT, induction chemotherapy; ACT, adjuvant chemotherapy. P values were derived from Pearson' s Chi-squared test.

Table S5. Allele frequencies of rs1131636 in NPC cohorts and general populations from 1000 Genomes Project (https://www.internationalgenome.org/). Allele frequency rs1131636, n (%) Population n C T CC CT TT NPC cohorts ALL 5553 0.453 0.547 1114 (20.1) 2801 (50.4) 1638 (29.5) SYSUCC-1 1471 0.451 0.549 293 (19.9) 741 (50.4) 437 (29.7) SYSUCC-2 1786 0.46 0.54 364 (20.4) 914 (51.2) 508 (28.4) SYSUCC-3 1751 0.455 0.545 356 (20.3) 882 (50.4) 513 (29.3) NCCS 545 0.428 0.572 101 (18.5) 264 (48.4) 180 (33.0) 1000 Genomes Project (Phase 3) ALL 2504 0.535 0.465 753 (30.1) 1175 (46.9) 576 (23) African 661 0.643 0.357 278 (42.1) 294 (44.5) 89 (13.5) African Caribbeans in Barbados 96

0.625 0.375 36 (37.5) 48 (50) 12 (12.5) Americans of African Ancestry in SW USA 61 0.615 0.385 23 (37.7) 29 (47.5) 9 (14.8) Esan in Nigeria 99 0.672 0.328 42 (42.4) 49 (49.5) 8 (8.1) Gambian in Western Divisions in the Gambia 113 0.571 0.429 38 (33.6) 53 (46.9) 22 (19.5) Luhya in Webuye, Kenya 99 0.672 0.328 46 (46.5) 41 (41.4) 12 (12.1) Mende in Sierra Leone 85 0.676 0.324 40 (47.1) 35 (41.2) 10 (11.8) Yoruba in Ibadan, Nigeria 108 0.671 0.329 53 (49.1) 39 (36.1) 16 (14.8) Ad Mixed American 347 0.465 0.535 75 (21.6) 173 (49.9) 99 (28.5) Colombians from Medellin, Colombia 94 0.5 0.5 21 (22.3) 52 (55.3) 21 (22.3) Mexican Ancestry from Los Angeles USA 64 0.406 0.594 11 (17.2) 30 (46.9) 23 (35.9) Peruvians from Lima, Peru 85 0.476 0.524 21 (24.7) 39 (45.9) 25 (29.4) Puerto Ricans from Puerto Rico 104 0.462 0.538 22 (21.2) 52 (50) 30 (28.8) East Asian 504 0.476 0.524 115 (22.8) 250 (49.6) 139 (27.6) Chinese Dai in Xishuangbanna, China 93 0.484 0.516 18 (19.4) 54 (58.1) 21 (22.6) Han Chinese in Beijing, China 103 0.447 0.553 23 (22.3) 46 (44.7) 34 (33) Southern Han Chinese 105 0.419 0.581 19 (18.1) 50 (47.6) 36 (34.3) Japanese in Tokyo, Japan 104 0.476 0.524 25 (24) 49 (47.1) 30 (28.8) Kinh in Ho Chi Minh City, Vietnam 99 0.561 0.439 30 (30.3) 51 (51.5) 18 (18.2) European 503 0.374 0.626 66 (13.1) 244 (48.5) 193 (38.4) Utah Residents (CEPH) with Northern and 99 0.348 0.652 13 (13.1) 43 (43.4) 43 (43.4) Western European Ancestry Finnish in Finland 99 0.348 0.652 9 (9.1) 51 (51.5) 39 (39.4) British in England and Scotland 91 0.374 0.626 12 (13.2) 44 (48.4) 35 (38.5) Iberian Population in Spain 107 0.421 0.579 18 (16.8) 54 (50.5) 35 (32.7) Toscani in Italia 107 0.374 0.626 14 (13.1) 52 (48.6) 41 (38.3) South Asian 489 0.667 0.333 219 (44.8) 214 (43.8) 56 (11.5) Bengali from Bangladesh 86 0.663 0.337 39 (45.3) 36 (41.9) 11 (12.8) Gujarati Indian from Houston, Texas 103 0.684 0.316 47 (45.6) 47 (45.6) 9 (8.7) Indian Telugu from the UK 102 0.632 0.368 44 (43.1) 41 (40.2) 17 (16.7) Punjabi from Lahore, Pakistan 96 0.63 0.37 36 (37.5) 49 (51) 11 (11.5) Sri Lankan Tamil from the UK 102 0.721 0.279 53 (52) 41 (40.2) 8 (7.8)

Table S6. Cis -eQTL effects of rs1131636 on RPA1 expression from GTEx portal. SNP Chromosome Position Gene Symbol Normalized effect size P -value Tissue -0.23 3.6×10-9 Tibial nerve -0.16 4.1×10-7 Skeletal muscle rs1131636 17 1801189 RPA1 -0.35 4.1×10-6 Brain cortex -0.24 1.0×10-5 Adrenal gland -0.1 2.5×10-5 Oesophagus mucosa * https://gtexportal.org/home/

Table S7. Cis -eQTL effects of rs1131636 on RPA1 expression from Blood eQTL browser. SNP Chromosome Position Gene Symbol Z-score P -value Tissue rs1131636 17 1801189 RPA1 -3.610 0.0003 Peripheral blood * https://genenetwork.nl/bloodeqtlbrowser/

Note: Figure translations are in progress. See original paper for figures.