European Journal of Cancer 107 (2019) 68–78

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejcancer.com

Original Research

Common genetic variants contribute to incomplete penetrance: evidence from cancer-free *BRCA1* mutation carriers

Bradley Downs [a], Simon Sherman [b], Jian Cui [a], Yeong C. Kim [a], Carrie Snyder [d], Maria Christensen [d], Jiangtao Luo [c], Henry Lynch [d,**], San Ming Wang [a,e,*]

[a] *Department of Genetics, Cell Biology and Anatomy, College of Medicine, USA*
[b] *Eppley Institute for Research in Cancer and Allied Diseases, USA*
[c] *Department of Biostatistics, College of Public Health, University of Nebraska Medical Center, Omaha, NE 68198, USA*
[d] *Hereditary Cancer Center, Department of Preventive Medicine, Creighton University School of Medicine, Omaha, NE 68178, USA*
[e] *Cancer Centre and Institute of Translational Medicine, Faculty of Health Sciences, University of Macau, Taipa, Macau, SAR, China*

Received 28 July 2018; received in revised form 25 October 2018; accepted 31 October 2018
Available online 11 December 2018

KEYWORDS
BRCA1;
Breast cancer;
Cancer resistance;
Exome;
Variation

Abstract ***Purpose:*** The presence of pathogenic germline mutation in *BRCA1* gene is considered as the most penetrant genetic predisposition for breast cancer. However, a portion of *BRCA1* mutation carriers never develops breast cancer throughout their lifetime. This phenomenon is called incomplete penetrance. Genetic factor is proposed to contribute to this phenomenon, but the details regarding the genetic factor remain elusive. *BRCA1* mutations were inherited from the ancestors of the mutation carrier families during human evolution, and their presence is a consistent threat to the survival of the mutation carrier population. In the present study, we hypothesize that evolution could positively select genetic components in the mutation carrier population to suppress the oncogenesis imposed by the predisposition. ***Experimental design:*** To test our hypothesis, we used whole exome sequencing to compare germline variation of all genes in pairs of breast cancer–unaffected and breast cancer–affected *BRCA1* mutation carriers, each pair was from the same family carrying the same *BRCA1* mutation. ***Results:*** We identified a group of 'beneficial' variants enriched in the breast cancer–unaffected carrier group. These were the common variants in human population distributed in multiple genes

* *Corresponding author*: Department of Preventive Medicine, Creighton University School of Medicine, 2500 California Plaza, Omaha, NE 68178, USA. Fax + 1 402 280 1734.
** *Corresponding author*. Faculty of Health Sciences, University of Macau, Macau, Taipa, SRA, China. Fax + 853 88222314.
E-mail addresses: htlynch@creighton.edu (H. Lynch), sanmingwang@unmc.edu (S.M. Wang).

https://doi.org/10.1016/j.ejca.2018.10.022

involved in multiple functionally important pathways. We found a single-nucleotide polymorphism, rs3735400 located in *ANLN* gene, which plays an essential role in controlling cytokinesis and is often found to be overexpressed in cancer. The carriers of this variant had lower cumulative risk of developing breast cancer; overexpression of the variant-containing *ANLN* decreased *ANLN* nuclear localization suppressed expression of the variant-containing *ANLN,* and decreased the cellular proliferation respectively.

Conclusion: Our findings support our hypothesis that common genetic variants can be evolutionarily selected in *BRCA1* mutation carrier population to counterpart the oncogenic effects imposed by mutation predisposition in *BRCA1*, contributing to the incomplete penetrance.

1. Introduction

Penetrance refers to the rate of individuals with a given genotype to express the phenotype associated with the genotype. Not all genotypes, no matter how strong, can lead to full phenotype, this is known as incomplete penetration [1]. Incomplete penetration has been well observed in hereditary cancer caused by multiple genes including *BRCA1* and *BRCA2* and so on. Although predispositions in each of these genes are well known to contribute to cancer, the incidence of cancer development in the mutation carrier population rarely reaches 100%. For example, predisposition in *BRCA1* is the most penetrant mutation in causing breast cancer [2–4]. Even in this case, the highest estimate for cancer risk in mutation carriers was 80% by the age of 70 years [5–7], whereas the remaining portion will never develop breast cancer in their lifetime. Understanding the causes for incomplete penetration is important as it can explain the genetic basis of disease development in the mutation carriers, estimate likelihood of mutation carriers to develop the disease and identify potential targets to prevent disease development.

Genetic factor has been considered to contribute to incomplete penetrance; however, the details remain elusive. The presence of incomplete penetrance in the mutation carriers highlights the possibility of the presence of disease-resistance mechanisms in the disease-free mutation carrier population. Taking cancer as an example: cancer resistance has been considered as a common phenomenon in Metazoans [8–10]. For instance, different mouse strains have different cancer susceptibilities [11], and certain human populations have extremely low cancer rates [12–16]. In the case of *BRCA1*, most of the mutations were generated during human evolution process, and some were generated even before human diversified from other hominins [7]. Due to its severity in disrupting homologous repair and genome stability leading to increased risk of cancer, the mutation in *BRCA1* is a permanent threat to the survival of the mutation carrier population. To counterpart this lethal threat, protective mechanisms in the mutation carrier population can be developed to suppress *BRCA1* mutation imposed oncogenic effects. It is likely that whether a mutation carrier develops or escapes cancer depends on the interaction between the oncogenic effects imposed by *BRCA1* mutation and the resistant effects in the mutation carrier. The severity, high rate of cancer incidence and deep understanding of the oncogenic mechanism make *BRCA1* mutations as an ideal model to study incomplete penetrance.

We hypothesize that evolution could positively select certain genetic components in the mutation carrier population to suppress the oncogenesis imposed by the predisposition. In this study, we searched for the evidence for the presence of genetic variants differentially present between the cancer-unaffected and cancer-affected mutation carrier populations using whole exome sequencing and identified a set of 'beneficial' variants enriched in the cancer-unaffected group. Furthermore, functional test of a variant in *ANLN*, a nuclear protein gene essential for cytokinesis, showed that the variant altered its nuclear localization and delayed its cell proliferation. Data from our study support our hypothesis for the presence of common genetic variants to suppress the oncogenic effects imposed by genetic predisposition.

2. Materials and methods

2.1. Samples

The samples used for this study were obtained from the Creighton University Breast Cancer Family Registry. All the participants provided a written informed consent to be a part of the cancer genetic study. The Institutional Review Board of Creighton University and the Institutional Review Board of the University of Nebraska Medical Center approved the study (CU #00–12265, UNMC #718-11-EP).

2.2. Exome sequencing, mapping, variant calling and damaging variant identification

Genomic DNA was extracted from the blood mononuclear cells of *BRCA1*+ carriers. The whole exome sequencing was performed as following: exome libraries were constructed using the SureSelectXT2 kit following

the manufacturer's protocol (Agilent, Santa Clara, CA) and sequenced on an Illumina HiSeq2500 sequencer with paired-end reads (2 × 150). The adaptor sequences were trimmed and mapped to the reference human genome (hg19) using BWA [17]. Duplicates were removed using Picard (http://broadinstitute.github.io/picard/). Single-nucleotide variants and indels were called using GATK (Broad Institute, Cambridge, MA 02142) according to the GATK best practices protocol [18], under the conditions of quality score ≥40, detected in ≥10 individual sequences and present in ≥20% of sequences mapped to the same region. The variants causing non-synonymous coding change, affecting the consensus dinucleotide splice junction sequence and introducing stop codon were identified and annotated using ANNOVAR [19], damaging variants were predicted using both Polymorphism Phenotyping v2 (Polyphen-2) and Sorts Intolerant From Tolerant (SIFT) programs [20,21]. PolyPhen-2 is a computational program designed to predict the impact of amino acid substitutions on the structure and function of the human proteins based on physical and evolutionary conservation. It predicts the consequences using the predicted scores (probably damaging, 0.909–1, possibly damaging 0.447–0.908 and benign 0–0.446; http://genetics.bwh.harvard.edu/pph2/dokuwiki/start); SIFT is a computational program designed to predict whether amino acid substitutions in proteins can have a phenotypic effect based on sequence homology and physical properties of the replaced amino acids (http://sift.bii.a-star.edu.sg). SIFT predicts the consequences using the predicted scores (damaging < 0.05 and tolerant ≥ 0.05). Sanger sequencing was used to validate selected variants. The exome data were deposited in NCBI dbGaP database with accession number phs001243.v1.p1 (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001243.v1.p1).

2.3. Small hairpin RNA expression and cell growth

Packaging plasmid psPAX2, envelope plasmid pMD2.G and scramble small hairpin RNA (shRNA) on the vector pLKO.1 were obtained from Addgene (Cambridge, MA). ANLN shRNA clone TRCN0000117257 was purchased from Sigma–Aldrich (St. Louis, MO). Lentiviral particles were produced using the pLKO.1-TRC cloning vector from Addgene (Cambridge, MA). Transfected cells were selected under Puromycin (1 μg/mL). MCF7 cells were grown in Eagle's Minimum Essential Medium from American Type Culture Collection (Manassas, VA) supplemented with 10 μg/mL insulin, 10% fetal bovine serum (FBS) and 1X streptomycin/penicillin at 37 °C with 5% CO_2. Stable shRNA-transfected cells were seeded at 3.0×10^4 in 6-well plates with Eagle's Minimum Essential Medium. Cell viability was determined by Trypan Blue staining. Cells were counted with a hemocytometer.

2.4. Nuclear expression

Site-directed mutagenesis was used to introduce the variant rs3735400 (C/G) into the eGFP-anillin plasmid [23] using a QuickChange 2 XL Site-Directed Mutagenesis Kit from Agilent (Santa Clara, CA). The oligos used to induce the variant in the *ANLN* construct were forward 5′- gagaaatcttgtacaaaaccatGgccatcaaaaaaacg ctg-3′ and reverse 5′-cagcgttttttgatggcCatggttttgtacaa-gatttctc-3′. Sanger sequencing was used to verify the resulting variant using primer 5′-ctgcacctgaggagacacag-3′. Constructs were transfected into HeLa cells with Lipofectamine 2000 following the manufacturer's protocol of Thermo Fisher (Waltham, MA). The transfected cells were grown in DMEM HyClone (SH30022.01) supplemented with 10% FBS at 37 °C with 5% CO_2. After 2 days, the cells were transferred and grown on a glass coverslip for 24 h. The cells were fixed with 4% paraformaldehyde in phosphate-buffered saline. DAPI Fluoromount-G was used to mount and stain the nucleus. Images from the transfected cells were captured in a Nikon Eclipse TI-E microscope with a Nikon Digital sight DS-QiMc camera and overlaid in NIS Elements of Nikon (Melville, NY). Fluorescence intensity was measured using the ImageJ software.

2.5. Statistical data analysis

Variants differentially present between breast cancer–unaffected carrier group and breast cancer–affected carrier group were determined using the paired test. Binomial exact test was used to determine discordant or a lack of agreement on each variant. *P* values <0.05 were set as significant [24], Odds ratio (OR) and their 95% confidence intervals (CI) were calculated using the unconditional maximum likelihood with the R package EpiTools made by Ausvet Company (Bruce ACT 2617, Australia) software. The variants that passed the biological filters and OR statistics were then reanalysed in the cohort study using the Cox regression model [25]. Hazard ratio (HR), their 95% CI and Wald's *P* value were calculated using the R Package Survival software.

3. Results

3.1. Sequence collection and analysis

Twenty-seven pairs of breast cancer–unaffected and breast cancer–affected members (four had both breast cancer and ovarian cancer), of which each pair was from the same *BRCA1* mutation carrier family, were selected for exome sequencing. Each pair had inherited the same mutation in *BRCA1*; 25 mutations were classified as pathogenic by Breast Cancer Information Core and/or ClinVar databases, and two mutations were large deletions in *BRCA1* (Table 1, Supplementary Table 1). In

Table 1
Pairs of *BRCA1* mutation carriers used for exome sequencing.

Pair	HGVS cDNA	HGVS protein	Mutation type	Clinvar database	Age		
					Affected	Unaffected	Difference
1	c.66_67delAG	p.Leu22_Glu23LeuValfs	Frameshift	Pathogenic	37	64	27
2	c.66_67delAG	p.Leu22_Glu23LeuValfs	Frameshift	Pathogenic	38	82	44
3	c.66_67delAG	p.Leu22_Glu23LeuValfs	Frameshift	Pathogenic	36	59	23
4	c.69_70insAG	p.Glu23_Cys24?fs	Frameshift	Pathogenic	31	73	42
5	c.135-1G>T	–	IVS	Pathogenic	21	59	38
6	c.181 T>G	p.Cys61Gly	Missense	Pathogenic	27	65	38
7	c.213-11 T>G	–	IVS	Pathogenic	35	80	45
8	c.302-2_302-2delA	–	IVS	Pathogenic	33	60	27
9	c.427G>T	p.Glu143Ter	Nonsense	Pathogenic	27	55	28
10	c.763G>T	p.Glu255Ter	Nonsense	Pathogenic	34	56	22
11	c.797_798delTT	p.Val266=fs	Frameshift	Pathogenic, likely pathogenic	23	59	36
12	c.1121_1121delC	p.Thr374Asnfs	Frameshift	Pathogenic	38	70	32
13	c.1504_1508delTTAAA	p.Leu502_Lys503?fs	Frameshift	Pathogenic	34	64	30
14	c.2722G>T	p.Glu908Ter	Nonsense	Pathogenic	27	58	31
15	c.3710_3710delT	p.Ile1237Asnfs	Frameshift	Pathogenic	59	58	−1
16	c.4065_4068delTCAA	p.Asn1355_Gln1356?fs	Frameshift	Pathogenic, likely pathogenic	31	57	26
17	c.4186−1643_4357+2020del	–	IVS	Pathogenic	28	75	47
18	c.4689C>G	p.Tyr1563Ter	Nonsense	Pathogenic	27	74	47
19	c.5153−2_5153-2delA	–	IVS	Pathogenic	38	66	28
20	c.5263_5264insC	p.Ser1755?fs	Frameshift	Pathogenic	33	64	31
21	c.5263_5264insC	p.Ser1755?fs	Frameshift	Pathogenic	32	64	32
22	c.5263_5264insC	p.Ser1755?fs	Frameshift	Pathogenic	44	51	7
23	c.5503C>T	p.Arg1835Ter	Nonsense	Pathogenic, likely pathogenic	31	63	32
24	exons 1−2 del37kb	–	Large deletion	–	36	82	46
25	exon 13 ins6kb	–	Large insertion	Pathogenic	39	60	21
26	exon 13 ins6kb	–	Large insertion	Pathogenic	43	70	27
27	exon 17 del1008bp		Large deletion	–	30	58	28
Average age					34	65	31

HGVS, human genome variation society.

the case where multiple breast cancer−unaffected carriers were available in a given family, the oldest carrier was selected; in the case of breast cancer−affected carriers, the carrier diagnosed at the youngest age was selected for the study. The mean age was 65 years for the breast cancer−unaffected group and 34 years for the breast cancer−affected group. The exome coverage was at 230*x* on an average for breast cancer−unaffected group, and 184× on an average for the breast cancer−affected group (Supplementary Table 2). In the 27 pairs, 6,509 variants in 4,447 genes were identified. Of these variants, 1,813 named as 'beneficial' variants were present only in the breast cancer−unaffected carrier group, 1,941 named as 'deleterious' variants were present only in the breast cancer−affected carrier group, and the rests were present in both groups (Fig. 1, Supplementary Table 3).

We tested if the sample size of 54 cases had the sufficient power to detect meaningful variation data. We searched in the variant list for the presence of the two variants located in coding genes identified by previous breast cancer genome-wide association studies (GWAS). One variant was rs8100241 in exon 1 of *ANKLE1*, which decreases the risk of breast cancer [26–29], and the other variant was rs2295190 located in exon 146 of *SYNE1*,

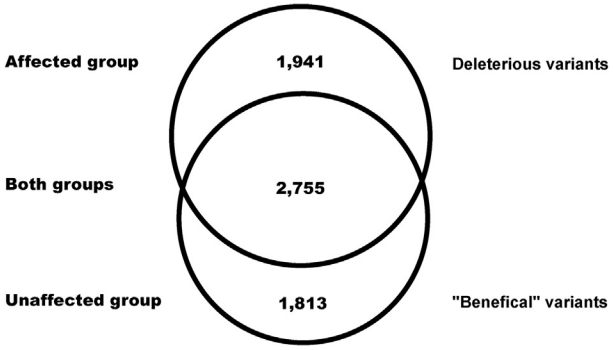

Fig. 1. Venn diagram for the distribution of damaging variants identified in breast cancer−unaffected and cancer-affected *BRCA1* mutation carrier groups. Variants from exome sequences of the 27-pair *BRCA1* mutation carriers were identified, and damaging variants were annotated by Polyhphen-2 and SIFT programs. Affected: the variants detected in the cancer-affected *BRCA1* mutation carrier group; unaffected: the variants detected in the cancer-unaffected *BRCA1* mutation carrier group; both groups: the variants detected in both groups. Polyhphen-2, Polymorphism Phenotyping v2; SIFT, Sorts Intolerant From Tolerant.

Table 2
Variants significantly different between breast cancer–unaffected and breast cancer–affected *BRCA1* mutation carriers.

Distribution					Chr.	Position	Gene	Exon	hg19	Variant	Variation		dbSNP ID	Frequency (1000G)	Damaging score	
Unaffected	Affected	Both	None	*P* value							Nucleotide	Amino acid			Polyphen2	SIFT
Higher in breast cancer-unaffected carriers																
9	0	2	16	0.004	7	36438709	*ANLN*	3	C	G	c.C194G	p.S65W	rs3735400	0.113	1	0
8	0	0	19	0.008	1	232568041	*SIPA1L2*	14	G	A	c.C4208T	p.S1403L	rs1547742	0.097	0.991	0.01
10	1	9	7	0.012	1	226019633	*EPHX1*	3	T	C	c.T337C	p.Y113H	rs1051740	0.300	0.998	0
7	0	17	3	0.016	11	59132798	*OR5AN1*	1	G	C	c.G867C	p.L289F	rs7941190	0.695	1	0.03
7	0	5	15	0.016	3	122354792	*PARP15*	8	G	A	c.G1180A	p.G394R	rs12489170	0.132	1	0.03
14	1	7	5	0.001	19	17392894	*ANKLE1*	2	G	A	c.G220A	p.A74T	rs8100241	0.572	1	0.1
8	1	2	16	0.039	1	241951329	*WDR64*	23	C	T	c.C2854T	p.R952W	rs12074374	0.169	1	0
6	0	1	20	0.031	1	161495040	*HSPA6*	1	C	T	c.C592T	p.L198F	rs1079109	0.115	1	0.01
6	0	2	19	0.031	X	140994407	*MAGEC1*	4	C	G	c.C1217G	p.S406C	rs62611965	0.094	0.991	0.01
6	0	0	21	0.031	1	35227362	*GJB4*	2	C	G	c.C507G	p.C169W	rs79193415	0.037	1	0
6	0	0	21	0.031	6	26093141	*HFE*	2	G	A	c.G305A	p.C102Y	rs1800562	0.064	1	0
10	2	9	6	0.039	19	49232226	*RASIP1*	5	G	A	c.C1801T	p.R601C	rs2287922	0.440	0.999	0.02
Higher in breast cancer–affected carriers																
1	9	1	16	0.021	18	47429022	*MYO5B*	21	C	T	c.G2753A	p.R918H	rs2298624	0.134	0.987	0.01
1	8	1	17	0.039	1	16532498	*ARHGEF19*	8	G	A	c.C1379T	p.P460L	rs41269185	0.120	0.999	0.02

dbSNP = single-nucleotide polymorphism database; Polyphen2 = Polymorphism Phenotyping v2; SIFT = Sorts Intolerant From Tolerant.

Fig. 2. Cumulative risk of variant carriers and non-carriers. A total of 161 *BRCA1* mutation carriers were genotyped by Sanger sequencing, and the Cox regression model was used to calculate the cumulative risk. The results show that three variants significantly decreased the risk of breast cancer in these *BRCA1* mutation carriers.

which increases the risk of ovarian cancer [30]. Both the variants were identified in our study. rs8100241 was present in 14 unaffected, 1 affected and 7 of both the unaffected and affected; rs2295190 was present in eight unaffected, three affected and one of both the unaffected and affected (Supplementary Table 3), respectively. The presence of the two known variants provided a useful internal control to ensure that the sample size used in our study allowed to identify potential variants differentially present between the cancer-unaffected and cancer-affected groups although larger sample size should allow more comprehensive detection.

3.2. Variant identification

By statistical analysis, 14 single-base variants were identified significantly different between the breast cancer–unaffected carrier group and the breast cancer–affected carrier group, of which 12 showed high significance in breast cancer–unaffected carrier group and two in the breast cancer–affected carrier group (Table 2). All the 14 variants were validated by Sanger sequencing of the original DNA samples used in exome sequencing. These variants were present in the dbSNP database, also all, but one, were common polymorphisms with frequencies >5% in the human population. Each of these variants was predicted as 'deleterious' by both Polyphen-2 and SIFT programs.

The 12 variants enriched in breast cancer–unaffected group were located in the genes with different functional categories. SIPA1L2 is involved in the Ras signalling pathway; EPHX1 hydrolyses arene and aliphatic epoxides to less reactive and more water-soluble dihydrodiols, and mutations in EPHX1 cause preeclampsia, epoxide hydrolase deficiency and increased epoxide hydrolase activity; PARP15 has poly adenosine diphosphate ribose polymerase activity and is associated with urethral intrinsic sphincter deficiency; ANKLE1 has endonuclease activity, is involved in nuclear envelope formation and the variant rs8100241 in ANKLE1 has been determined by multiple breast cancer GWAS studies as one of the 71 traits associated with breast cancer and, in particular, with decreased risk for breast cancer [6]; HSPA6 is a chaperone protein essential in maintaining protein folding; MAGEC1 is a member of the melanoma antigen gene family; GJB4 is a transmembrane connexin protein involved in gap junction. Mutation in GJB4 is associated with erythrokeratodermia variabilis, progressive

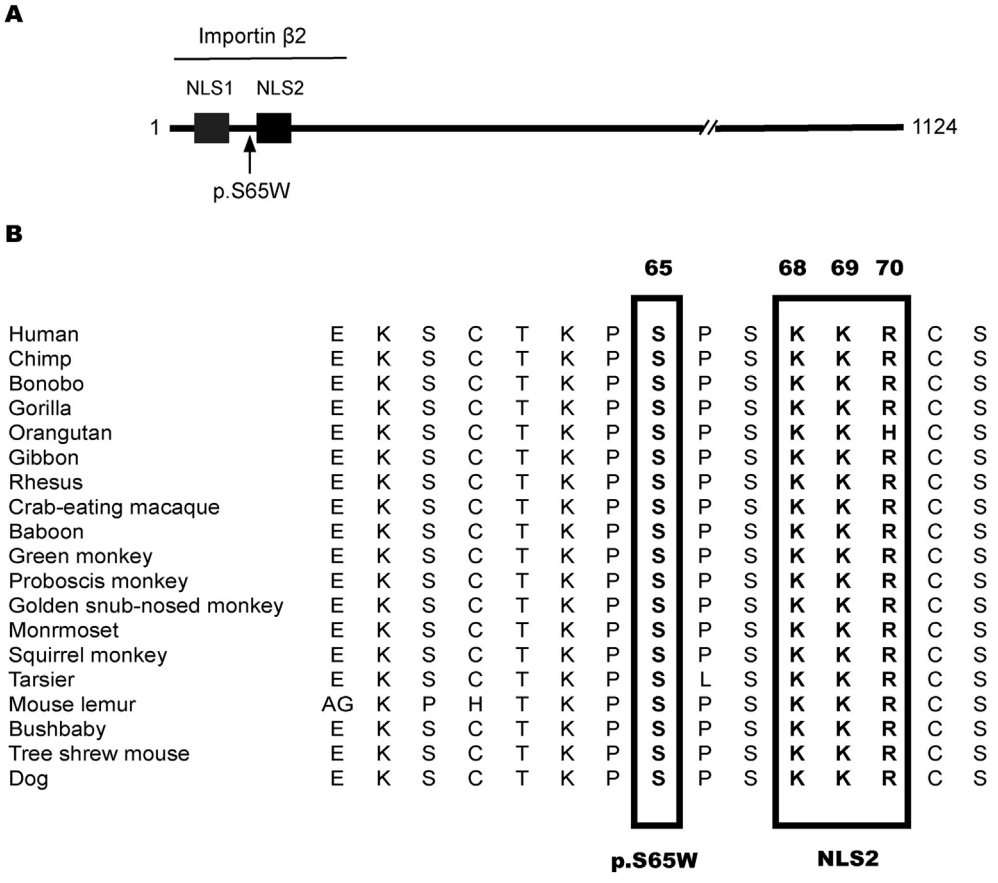

Fig. 3. Variant rs3735400 in *ANLN*. (A) Location of rs3735400 in ANLN. The variant causes the coding change (p.S65W) upstream of the second nuclear localization signal within the importin β2 interaction domain; (B) Conservation of rs3735400-affected residue in ANLN. Exon 3 (58–72) of human ANLN was aligned to ANLN sequences of 18 vertebrates, showing that the Serine residue is highly conserved across all these animals.

symmetric erythrokeratoderma and hearing impairment; HFE regulates iron absorption by regulating the interaction between transferrin receptor and transferrin. Mutations in HFE cause hereditary haemochromatosis.

We analysed the cumulative risks of the mutation carriers in a cohort of 161 *BRCA1* mutation carriers. Each was genotyped by Sanger sequencing, and Cox regression model was used to calculate the cumulative risk for integrated tumorigenesis. We observed that three variants modified the risk of breast cancer in the *BRCA1* mutation carriers, including rs8100241 in *ANKLE1* (HR 0.64, CI 0.5–0.099, $P = 0.046$) consistent with GWAS results [6] (Fig. 2A); rs3735400 in *ANLN* (HR 0.56, CI 0.34–0.93, $P = 0.025$; Fig. 2B); rs1547742 in *SIPA1L2* (HR 0.464, CI 0.28–0.77, $P = 0.003$; Fig. 2C). We also conducted a multinomial distribution analysis for the frequency distributions across our sample of variant calls [31]. The results show that there was a significant difference of frequency distributions for the variants with P-value <0.0001 (Supplementary Table 4).

3.3. Effects of variant-affected ANLN

We further selected the variant-affected *ANLN* to test the effects of the variants on the behaviours of *BRCA1* mutated cells. ANLN is essential in cytokinesis. It provides structural integrity of the cleavage furrow by coordinating the assembly of actin, myosin, septins, Rho and microtubules into functional cytokinesis machinery to facilitate the cleavage of a single cell into two daughter cells [32,33]. Therefore, its alteration can affect cellular proliferation. In the functional studies, the mutated ANLN and suppressed expression of ANLN inhibited growth, colony formation, cell migration and cell cycle progression [34–37]; in clinical studies, ANLN was observed to be overexpressed in many types of cancer including breast cancer [22,38–41]. rs3735400 in *ANLN* was enriched in breast cancer–unaffected group: of the 11 carriers, nine were in the unaffected carrier group, two were shared between the two groups, and none was present in the affected carrier group ($P = 0.004$). rs3735400 is a common polymorphism in the human population (0.113 in the 1000 Genomes data; Table 2).

ANLN has two nuclear localization signals (NLS). rs3735400 (C>G) causes a non-synonymous change (p.S65W), two residues upstream of the second NLS. This region is within the interaction domain of ANLN with importin β2, a nuclear membrane receptor that transports ANLN into nucleus [37]. The serine residue at this position is evolutionarily highly conserved across

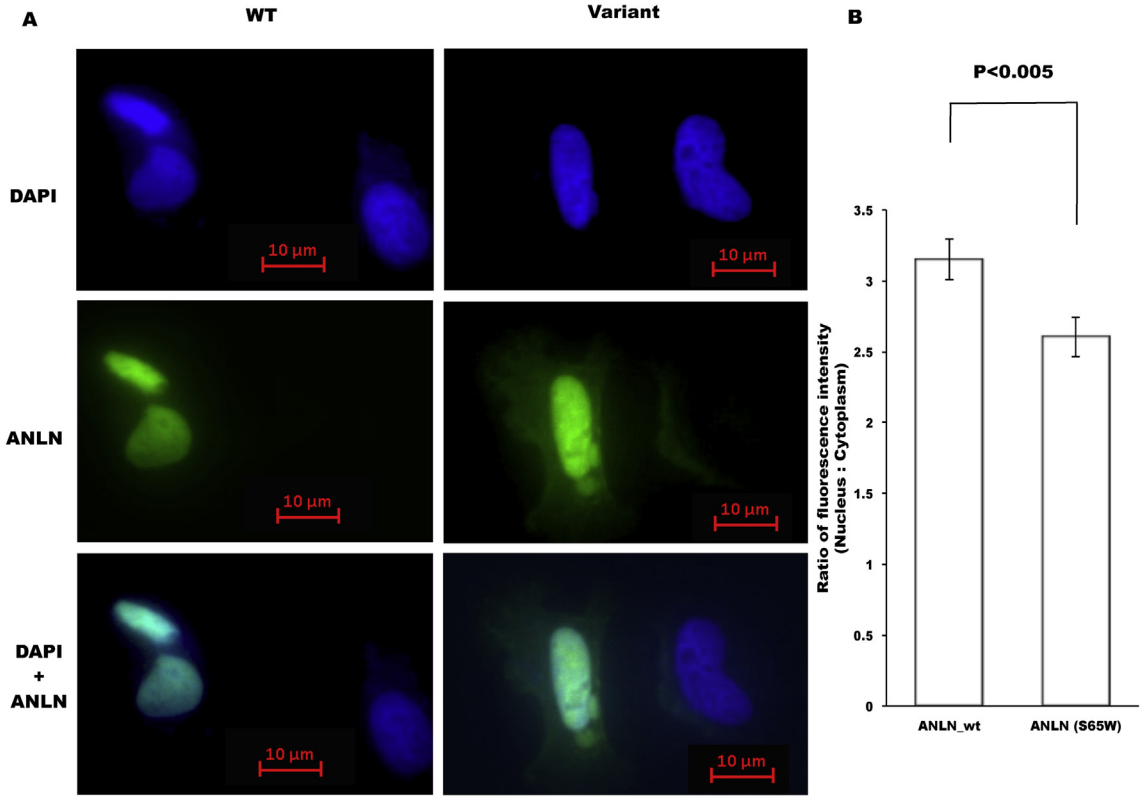

Fig. 4. Variant rs3735400 decreased ANLN nuclear localization in MCF7 cells. (A) Expression of GFP-tagged wild-type and rs3735400-containing *ANLN* constructs in HeLa cells. (B) Different ratios between nucleus and cytoplasm localization in wild-type and rs3735400-containing ANLN. It shows that the ratio in rs3735400-containing ANLN was significantly lower than that in the wild type ANLN ($P < 0.005$).

mammals (Fig. 3A and B). We tested the effects of rs3735400 on ANLN nuclear localization. We transfected wild-type and the rs3735400-containing *ANLN* using GFP as the marker in Hela cells. Compared to the wild-type *ANLN*, the rs3735400-containing *ANLN* showed significantly decreased nuclear localization and increased cytoplasmic accumulation (Fig. 4A and B). The results show that rs3735400 can influence ANLN nuclear localization likely through interfering the NLS recognition and interaction with importin β2.

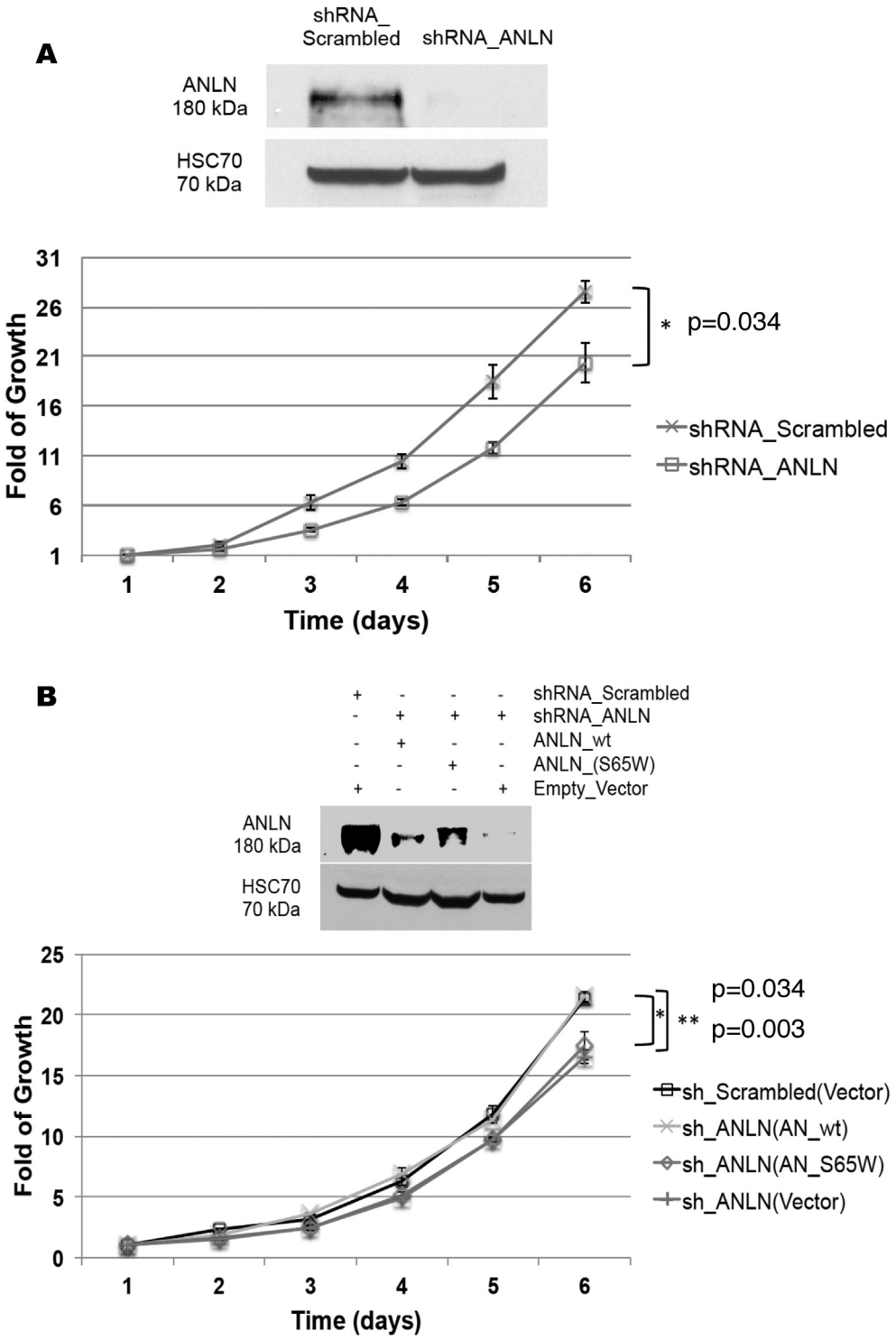

Fig. 5. Variant rs3735400 in *ANLN* led delayed cell growth. (A) shRNA knockdown expression of endogenous *ANLN* caused delayed cell growth. (B) The delayed cell growth was rescued by introducing wild-type ANLN but not rs3735400-containing ANLN. Each test was repeated three times with mean ± standard deviation.

We further tested the effects of rs3735400 on cell growth. We used shRNA to suppress the expression of endogenous *ANLN* in two breast cancer cell lines: HCC1937, in which both copies of *BRCA1* were mutated, and MCF7, in which one copy of *BRCA1* was mutated [42,43]. In HCC1937 cells, suppressed *ANLN* expression caused cells to stop the growth and eventually die; in MCF7 cells, suppressed *ANLN* expression significantly decreased cell growth (Fig. 5A). The decreased growth of MCF7 cells was rescued by transfecting the wild-type *ANLN* and not the rs3735400-containing *ANLN* (Fig. 5B).

4. Discussion

Data from the study support our hypothesis that evolution could select common genetic variants to counterpart the oncogenic effects imposed by genetic predisposition, contributing to the incomplete penetration. By comparing breast cancer–unaffected *BRCA1* mutation carrier group and breast cancer–affected *BRCA1* mutation carrier group, we observed the enriched presence of common genetic variants in the breast cancer–unaffected group. Incomplete penetration can be acquired during a long evolutionary selection process. The high penetrant mutation in *BRCA1* is a constant threat to the survival of the mutation carrier population due to the genome instability the mutation caused towards cancer. To diminish this oncologic pressure, protective mechanism(s) can be evolved to protect the *BRCA1* mutation carrier population from cancer [44]. Our study suggests that common variants can be selected to play significant roles for this purpose. It is important to note that while the contribution of each variant can be limited, the combined effects of multiple variants can contribute significantly to incomplete penetrance. This situation is similar to the common, low-risk genetic variants contributing to cancer risk [44], in which multiple such variants can contribute up to 14% of risk in hereditary breast cancer [45]. Therefore, common genetic variants can influence both cancer promotion and incomplete penetrance-generated cancer resistance [46].

In our study, we identified the 'beneficial' variants in the unaffected mutation carriers. However, it is worth to note that the risk of cancer development can still remain for the unaffected members after the tested aging time points, and the presence of the 'beneficial' variants can either fully prevent or delay the penetrance of the mutation predisposition.

Our study addresses the penetrance influenced by evolution selection of genetic variation. However, it is important to note that other factors may also contribute to the penetrance. For example, eenvironmental factors including nutrition, smoking, pregnancy, breastfeeding etc. can also contribute to penetrance. Furthermore, polygenic and epigenetic factors can also contribute to penetrance as observed in certain genetic diseases.

In summary, further study is commissioned to reveal the roles of common genetic variants in incomplete penetrance.

Acknowledgments

The authors thank Dr. Andrew Wilde (University of Toronto) for providing eGFP-anillin plasmid. This study was supported by an NIH grant CA180008 and three pilot grants from Fred & Pamela Buffett Cancer Center, University of Nebraska Medical Center (SMW), University of Macau Faculty of Health Sciences startup fund (SMW), American Cancer Society institutional grant (YCK), by revenue from Nebraska's excise tax on cigarettes awarded to Creighton Univeristy through the Nebraska Department of Health & Human Services (DHHS), the Charles F. and Mary C. Heider Chair in Cancer Research and Liz's Legacy fund through Kicks for a Cure (HL). The funders have no roles in study design, data collection and analysis, decision to publish or preparation of the article. The authors thank Ms. Khyati Chandratre for article editing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejca.2018.10.022.

Conflict of interest statement

The authors declare no competing financial interest.

Authors' contributions

BD collected exome data; BD, SS and YK performed genomic data analysis and interpretation; BD, JL and JC performed statistic data analysis; CS, MC, JC and HL collected and analysed patients' samples, and SMW conceived the study and wrote the article.

References

[1] Penetrance Ott J. In: Analysis of human genetic linkage. Revised Edition. Baltimore: The John Hopkins University Press; 1991. 146–164p.
[2] Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, et al. Linkage of early-onset familial breast cancer to chromosome 17q21. Science 1990;250:1684–9.
[3] Narod SA, Feunteun J, Lynch HT, Watson P, Conway T, Lynch J, et al. Familial breast-ovarian cancer locus on chromosome 17q12-q23. Lancet 1991;388:82–3.
[4] Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, et al. A strong candidate for the breast

and ovarian cancer susceptibility gene BRCA1. Science 1994;266:66–71.
[5] Chen S, Parmigiani G. Meta-analysis of BRCA1 and BRCA2 penetrance. J Clin Oncol 2007;25:1329–33.
[6] Antoniou A, Pharoah PD, Narod S, Risch HA, Eyfjord JE, Hopper JL, et al. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case series unselected for family history: a combined analysis of 22 studies. Am J Hum Genet 2003;72:1117–30.
[7] Pavlicek A, Noskov VN, Kouprina N, Barrett JC, Jurka J, Larionov V. Evolution of the tumor suppressor BRCA1 locus in primates: implications for cancer predisposition. Hum Mol Genet 2004;13:2737–51.
[8] Caulin AF, Maley CC. Peto's Paradox: evolution's prescription for cancer prevention. Trends Ecol Evol 2011;26:175–82.
[9] Roche B, Hochberg ME, Caulin AF, Maley CC, Gatenby RA, Misse D, et al. Natural resistance to cancers: a Darwinian hypothesis to explain Peto's paradox. BMC Canc 2012;12:387–91.
[10] Seluanov A, Hine C, Bozzella M. Distinct tumor suppressor mechanisms evolve in rodent species that differ in size and lifespan. Aging Cell 2008;7:813–23.
[11] Nagase H, Bryson S, Cordell H, Kemp CJ, Fee F, Balmain A. Distinct genetic loci control development of benign and malignant skin tumours in mice. Nat Genet 1995;10:424–9.
[12] Pharoah PD, Antoniou A, Bobrow M, Zimmern RL, Easton DF, Ponder BA. Polygenic susceptibility to breast cancer and implications for prevention. Nat Genet 2002;31:33–6.
[13] Li GM. Mechanisms and functions of DNA mismatch repair. Cell Res 2008;18:85–98.
[14] Takahashi M, Shimodaira H, Andreutti-Zaugg C, Iggo R, Kolodner RD, Ishioka C. Functional analysis of human MLH1 variants using yeast and in vitro mismatch repair assays. Cancer Res 2007;67:4595–604.
[15] Xie Y, Yang H, Cunanan C, Okamoto K, Shibata D, Pan J, et al. Deficiencies in mouse Myh and Ogg1 result in tumor predisposition and G to T mutations in codon 12 of the K-ras oncogene in lung tumors. Cancer Res 2004;64:3096–102.
[16] Hung RJ, Hall J, Brennan P, Boffetta P. Genetic polymorphisms in the base excision repair pathway and cancer risk: a HuGE review. Am J Epidemiol 2005;162:925–42.
[17] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie2. Nat Methods 2012;9:357–9.
[18] McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297–303.
[19] Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38:e164.
[20] Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods 2010;7:248–9.
[21] Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 2009;4:1073–81.
[22] Suzuki C, Daigo Y, Ishikawa N, Kato T, Hayama S, Ito T, et al. ANLN plays a critical role in human lung carcinogenesis through the activation of RHOA and by involvement in the phosphoinositide 3-kinase/AKT pathway. Cancer Res 2005;65:11314–25.
[23] Breslow NE, Day NE. Statistical methods in cancer Research: volume 1: analysis of case control studies. Lyon, France: International Agency for Research in Cancer; 1980.
[24] Therneau TM, Li H. Computing the cox model for case cohort designs. Lifetime Data Anal 1999;5:99–112.
[25] Antoniou AC, Wang X, Fredericksen ZS, McGuffog L, Tarrell R, Sinilnikova OM, et al. A locus on 19p13 modifies risk of breast cancer in BRCA1 mutation carriers and is associated with hormone receptor-negative breast cancer in the general population. Nat Genet 2010;42:885–92.
[26] Couch FJ, Gaudet MM, Antoniou AC, Ramus SJ, Kuchenbaecker KB, Soucy P, et al. Common variants at the 19p13.1 and ZNF365 loci are associated with ER subtypes of breast cancer and ovarian cancer risk in BRCA1 and BRCA2 mutation carriers. Cancer Epidemiol Biomark Prev 2012;21:645–57.
[27] Stevens KN, Fredericksen Z, Vachon CM, Wang X, Margolin S, Lindblom A, et al. 19p13.1 is a triple-negative-specific breast cancer susceptibility locus. Cancer Res 2012;72:1795–803.
[28] Couch FJ, Wang X, McGuffog L, Lee A, Olswold C, Kuchenbaecker KB, et al. Genome-wide association study in BRCA1 mutation carriers identifies novel loci associated with breast and ovarian cancer risk. PLoS Genet 2013;9:e1003212.
[29] Doherty JA, Rossing MA, Cushing-Haugen KL, Chen C, Van Den Berg DJ, Wu AH, et al. ESR1/SYNE1 polymorphism and invasive epithelial ovarian cancer risk: an ovarian cancer association consortium study. Cancer Epidemiol Biomark Prev 2010;19:245–50.
[30] Field CM, Alberts BM. Anillin, a contractile ring protein that cycles from the nucleus to the cell ccortex. J Cell Biol 1995;131:165–78.
[31] Anderson TW. Introduction to mmultivariate statistical analysis. 3rd ed. Wiley; 2004.
[32] Piekny AJ, Glotzer M. Anillin is a scaffold protein that links RhoA, actin, and myosin during cytokinesis. Curr Biol 2008;18:30–6.
[33] Wang D, Chadha GK, Feygin A, Ivanov AI. F-actin binding protein, anillin, regulates integrity of intercellular junctions in human epithelial cells. Cell Mol Life Sci 2015;72:3185–200.
[34] Oegema K, Savoian MS, Mitchison TJ, Field CM. Functional analysis of a human homologue of the Drosophila actin binding protein anillin suggests a role in cytokinesis. J Cell Biol 2000;150:539–52.
[35] Hickson GR, O'Farrell PH. Anillin: a pivotal organizer of the cytokinetic machinery. Biochem Soc Trans 2008;36(Pt 3):439–41.
[36] D'Avino PP. How to scaffold the contractile ring for a safe cytokinesis - lessons from Anillin-related proteins. J Cell Sci 2009;122(Pt 8):1071–9.
[37] Chen A, Akhshi TK, Lavoie BD, Wilde A. Importin β2 mediates the spatio-temporal regulation of anillin through a noncanonical nuclear localization signal. J Biol Chem 2015;290:13500–9.
[38] Hall PA, Todd CB, Hyland PL, McDade SS, Grabsch H, Dattani M, et al. The septin-binding protein anillin is overexpressed in diverse human tumors. Clin cancer res 2005;11:6780–6.
[39] Gbadegesin RA, Hall G, Adeyemo A, Hanke N, Tossidou I, Burchette J, et al. Mutations in the gene that encodes the F-actin binding protein anillin cause FSGS. J Am Soc Nephrol 2014;25:1991–2002.
[40] Zhou W, Wang Z, Shen N, Pi W, Jiang W, Huang J, et al. Knockdown of ANLN by lentivirus inhibits cell growth and migration in human breast cancer. Mol Cell Biochem 2015;398:11–9.
[41] Tomlinson GE, Chen TT, Stastny VA, Virmani AK, Spillman MA, Tonk V, et al. Characterization of a breast cancer cell line derived from a germ-line BRCA1 mutation carrier. Cancer Res 1998;58:3237–42.
[42] Gazdar AF, Kurvari V, Virmani A, Gollahon L, Sakaguchi M, Westerfield M, et al. Characterization of paired tumor and non-

tumor cell lines established from patients with breast cancer. Int J Canc 1998;78:766–74.
[43] Schwartz MLB, Williams MS, Murray MF. Adding protective genetic variants to clinical reporting of genomic screening results: restoring balance. J Am Med Assoc 2017;317:1527–8.
[44] Antoniou AC, Sinilnikova OM, McGuffog L, Healey S, Nevanlinna H, Heikkinen T, et al. Common variants in LSP1, 2q35 and 8q24 and breast cancer risk for BRCA1 and BRCA2 mutation carriers. Hum Mol Genet 2009;18:4442–56.
[45] Ghoussaini M, Pharoah PD, Easton DF. Inherited genetic susceptibility to breast cancer: the beginning of the end or the end of the beginning? Am J Pathol 2013;183:1038–51.
[46] Tan H. On the protective effects of gene SNPs against human cancer. EBioMedicine 2018;33:4–5.