Single nucleotide variants in KIF14 gene may have prognostic value in breast cancer

Single Nucleotide Variants in KIF14 Gene May Have Prognostic Value in Breast Cancer

Ivona Krus[1], Veronika Brynychová[1], Viktor Hlaváč[1,2], Radka Václavíková[1], Maria Kováčová[3], Renata Koževnikovová[4], Katerina Kopečková[5], Jannis Tornikidis[6], David Vrána[7], Jiří Gatěk[8], Pavel Souček[1,2]

[1]Toxicogenomics Unit, National Institute of Public Health, Srobarova 48, 100 42 Prague 10, Czech Republic

[2]Biomedical Center, Faculty of Medicine in Pilsen, Charles University, Pilsen, Czech Republic

[3]Third Faculty of Medicine, Charles University, Prague, Czech Republic

[4]Department of Oncosurgery, MEDICON, Prague, Czech Republic

[5]Department of Oncology, Second Faculty of Medicine, Charles University and Motol University Hospital, Prague, Czech Republic

[6]Department of Surgery, Second Faculty of Medicine, Charles University and Motol University Hospital, Prague, Czech Republic

[7]Comprehensive Cancer Center Novy Jicin, Hospital Novy Jicin, Novy Jicin, Czech Republic

[8]Department of Surgery, EUC Hospital Zlin and Tomas Bata University in Zlin, Zlin, Czech Republic

[*]Pavel Souček pavel.soucek@szu.cz

Abstract

Introduction Human kinesin 14 (KIF14) is one of the 70 prognostic marker genes (so-called Amsterdam profile) previously identified by the microarray of breast carcinomas, and its high transcript expression in tumor specimens indicates a poor prognosis for patients. We performed a pilot study to explore the prognostic and predictive meaning of KIF14 germline genetic variability in breast cancer patients.

Methods KIF14 coding sequence, including 5' and 3' untranslated regions and overlaps to introns for identification of splicing sites, was analyzed using next-generation sequencing in the testing set of blood DNA samples from 105 breast cancer patients with clinical follow-up. After rigorous evaluation of major allele frequency, haplotype blocks, in silico predicted functional aspects, expression quantitative trait loci, and clinical associations, eight single nucleotide variants were subsequently validated in the evaluation set of 808 patients.

Results Carriers of minor alleles G (rs17448931) or T (rs3806362) had significantly shorter overall survival than wild type homozygotes ($p = 0.010$ and $p = 0.023$, respectively) thus successfully replicating the results of the testing set. Both associations remained significant in the multivariate Cox regression analysis, including molecular subtype and stage as covariates (hazard ratio, HR = 1.7, 95% confidence interval (CI) = 1.1—2.8 for rs17448931 and HR = 1.9, CI 1.2-3.0 for rs3806362).

Discussion In conclusion, our preliminary data suggest that minor alleles in rs17448931 and rs3806362 of KIF14 represent candidate biomarkers of poor prognosis of breast cancer patients. After pending validation in independent populations and eventual functional characterization, these candidates might become useful biomarkers in the clinics

1 Introduction

Breast cancer is the most commonly diagnosed cancer (2.26 million new cases in 2020 worldwide) and the leading cause of cancer death (685 thousand) in women [**1**]. Therefore, robust diagnostic, prognostic, and predictive biomarkers, enabling effective personalized therapy, are a prerequisite for effective clinical decisions.

Cancer evolves over decades through multiple genetic and epigenetic changes in affected cells and tissues formed during an interplay between inherited genetic background and exposome [**2**]. Cancer hallmarks, signaling pathways affected during carcinogenesis, became the focus for development of targeted therapies [**3**]. However, not all patients may benefit from these achievements either due to unknown background mechanisms or to uncertain druggable targets in some tumors. Multidrug resistance is another well-known factor, significantly complicating and regrettably frequently causing failure of both conventional and targeted therapies. Despite the enormous success in the identification of many genetic or biological features of resistant tumors, biomarkers allowing efficient therapy optimization, avoiding or at least limiting resistance potential, are still missing.

Recent studies draw attention to kinesins, a superfamily of microtubule-based motor proteins with diverse functions [**4**]. Kinesins participate in the intracellular transport of various cargos, including organelles, protein complexes, chromosomes, and $mRNAs$, along microtubules in an adenosine triphosphate (ATP)-dependent way [**5**]. Mitotic spindle formation, chromosome segregation, midbody formation, and completion of cytokinesis represent important roles of kinesins during cell division [**6-9**].

Human kinesin 14 (KIF14; OMIM: 611279) was among the 70 prognostic marker genes identified by the microarray of breast cancer samples [**10**]. KIF14 overexpression in breast cancer cells with genomic gain in the 1q chromosome region, frequently also observed in other cancers, has been reported [**11**], and KIF14 transcript overexpression in tumor samples predicted poor survival in breast carcinoma patients [**12**]. Higher risk of metastasis and decreased metastasis-free survival were reported in patients with increased expression of KIF14 at the tips of the torpedo-like structures in breast carcinomas, and in KIF14-positive cells, genes of ether lipid metabolism were highly upregulated [**13**]. A high KIF14 expression was also significantly associated with worse relapse-free survival in breast cancer patients with a triple negative ($TNBC$) subtype [**14**]. Our previous studies have shown that gene expression levels of KIF14 and its interacting partners $PRC1$ (protein regulating cytokinesis 1, OMIM: 603484) and CIT (citron Rho-interacting serine/threonine kinase, OMIM: 605629) are elevated in breast and ovarian tumor tissues and are significantly associated with the survival of patients [**15, 16**]. On the other hand, the expression of these genes has no prognostic meaning in colorectal and pancreatic cancers [**17**]. Silencing KIF14 leads to disruption of the cell cycle—suggesting its potential use as a target for cancer therapy [**18**], reverses acquired resistance to protein kinase inhibitor sorafenib [**19**], and in terms of conventional chemotherapy enhances the sensitivity of breast cancer cells to docetaxel under in vitro conditions [**20**].

Despite the above reviewed evidence relating to clinical associations of KIF14 expression with the clinical progress of breast cancer patients, including sensitivity to conventional and targeted therapy, knowledge about the role of KIF14 genetic variability is virtually non-existent. With reference to recently published data [**21**], this pilot study addressed the question of the prognostic or predictive role of germline KIF14 genetic variability in breast cancer patients from a Czech population.

2 Materials and Methods

2.1 Patients

A total of 808 incident female breast cancer patients, diagnosed in two Prague region hospitals (the Medicon PLC and the Motol University Hospital) and the Hospital Atlas in Zlin (all in the Czech Republic), were included in the study. Patients were recruited between 2001 and 2013.

Diagnosis of all patients was confirmed histologically according to standard diagnostic procedures [**22**]. Hormone receptor expression was evaluated based on a 1% cut-off. Immunohistochemistry was utilized for HER2/ERBB2 (Erb-b2 receptor tyrosine kinase 2) testing; 3+ scores were considered as positive and 1+ as negative. In the case of 2+ scores, fluorescent in situ hybridization ($FISH$) was further used for status confirmation. The cut-off between high and low expression of proliferative marker Ki-67 was 13.25% [**23**]. Intrinsic breast carcinoma subtypes were classified following published recommendations [**24**]. Patients underwent therapy with either neoadjuvant or adjuvant chemotherapy or with hormonal therapy according to the actual treatment guidelines in effect at the time of study. Detailed clinical characteristics of the patients were retrieved from medical records (**Table 1**). Response to the neoadjuvant cytotoxic therapy was evaluated by the Response Evaluation Criteria in Solid Tumors ($RECIST$) v1.1 [**25**].

2.2 Targeted Sequencing in the Testing Phase of the Study

Blood samples were collected during diagnostic procedures using tubes with K_3EDTA anticoagulant. DNA was isolated from lymphocytes using the phenol/chloroform extraction method described earlier [**26**].

Our recent pharmacogenomics study provided data for 509 genes, including KIF14, by next generation sequencing (NGS) of DNA from the blood of 105 breast cancer patients [**21**]. For the present study, we extracted original raw data of KIF14 from the above-mentioned dataset, which further served as a testing set. Briefly, reads were mapped on reference sequence hg19 using the Burrows-Wheeler Alignment (BWA) mem tool [**27**]. Base and indel recalibration, as well as short indels and single nucleotide variants discovery and variant quality score recalibration, was done using the Genome Analysis ToolKit $GATK$ [**28**]. Variant annotation was accomplished using $ANNOVAR$ [**29**]. Details of the library preparation, target enrichment, data processing, and variant calling are provided elsewhere [**21**]. Variants for subsequent validation in the evaluation set were selected using the minor allele frequency ($MAF > 0.05$), construction of haplotype blocks, in silico predicted functional aspects, and clinical associations (clinical data are provided in the Online Supplemental Material (OSM) Resource 1). Based on these criteria, eight variants were assessed in the subsequent evaluation set of patients ($n = 808$).

Table 1 Clinical characteristics of breast cancer patients in the evaluation set

Characteristics	Patients, *N* (%)[a]	PFS[b]	OS[b]
Age at diagnosis, mean ± SD (years)	59.0 ± 12.5	NS	NS
Menopausal status		NS	NS
Premenopausal	202 (25)		
Postmenopausal	606 (75)		
Tumor size (pT)		3.4E−12	2.0E−05
pT1	554 (70)		
pT2	212 (27)		
pT3	19 (2)		
pT4	11 (1)		
pTx	12		
Lymph node metastasis (pN)		1.7E−05	0.001
Absent (pN0)	509 (67)		
Present (pN1–3)	256 (33)		
pNx	43		
Pathological stage		3.6E−10	2.6E−04
I	403 (52)		
II	300 (39)		
III	69 (9)		
Not determined	36		
Histological type		NS	NS
Invasive ductal carcinoma	606 (76)		
Other type	195 (24)		
Unknown	7		
Intrinsic molecular subtype		3.4E−06	0.010
Luminal A	191 (27)		
Luminal B	352 (50)		
HER2-enriched	62 (9)		
Triple-negative/basal-like	96 (14)		
Not determined	107		
Pathological grade (G)		0.002	0.008
G1	181 (23)		
G2	390 (50)		
G3	208 (27)		
Gx	29		
Estrogen receptor status		0.001	NS
Positive	621 (77)		
Negative	184 (23)		
Not determined	3		
Progesterone receptor status		0.004	NS
Positive	584 (73)		
Negative	221 (27)		
Not determined	3		
HER2/ERBB2 status		NS	NS
Positive	193 (24)		
Negative	611 (76)		
Not determined	4		
Expression of Ki-67		NS	NS
Positive	347 (58)		
Negative	253 (42)		

Table 1 (continued)

Characteristics	Patients, *N* (%)[a]	PFS[b]	OS[b]
Not determined	208		
Response to the neoadjuvant cytotoxic therapy		NS	NS
Complete or partial response	125 (75)		
Stable disease or progression	42 (25)		
No neoadjuvant therapy	641		

NS non-significant

[a]*Number of patients with % in parentheses*

[b]*p values according to the Breslow test*

2.3 Functional Predictions and Bioinformatics

All variants identified in the testing set were evaluated using the "Sorting Tolerant From Intolerant" ($SIFT$) algorithm [**30**] and the RegulomeDB database [**31**]. Variants classified as "deleterious" (score 0.0-0.05) by the $SIFT$ or with rank 1 or 2 in the RegulomeDB were selected as potentially functional for the evaluation phase. The distribution of hap-lotypes was tested using HaploView v.4.2 program [**32**] with the algorithm as published [**33**]. All variants in strong local linkage disequilibrium ($r^2 > 0.8$) were analyzed together as haplotype blocks. Expression quantitative trait loci ($eQTL$) were assessed using available data for transcript levels in nonmalignant mammary tissue from a subgroup of patients ($n = 20$) of the testing set reported previously [**16**]. Transcript levels of KIF14 in tumors ($n = 130$) and nonmalignant control tissues ($n = 65$) from available patients of the evaluation set were also used for this comparison. In silico analysis of the candidate variants tested in the evaluation phase was performed with the VarSome software (**https:// varsome.com/**).

2.4 Genotyping of KIF14 Candidate Variants in the Evaluation Phase of the Study

Variants rs12060793, rs74319334, rs17448931, rs12120084, and rs6665951 in KIF14 were assayed by TaqMan SNP Genotyping Assays (C_31277474_10, C_99738324_10, C_33326989_10, C_372014_10 and C_36796_10, respectively) using real-time PCR with ViiA7 Real-Time PCR System (Thermo Fisher Scientific, Waltham, MA, USA). The following cycling conditions were used: initial denaturation at 95 °C for 10 min followed by 40 cycles of denaturation at 95 °C for 15 s and annealing/extension at 60 °C for 60 s. Three additional variants in KIF14 (rs2808244, rs3806361, and rs3806362) were analyzed by direct DNA sequencing in one 653 bp PCR product. Oligonucleotide primers (OSM Resource 2) for sequencing were designed using the Primer3 software [**34**]. PCR products were generated using 50 ng of genomic DNA in 25 μL volume reactions containing 5 x FirePol Master Mix with 7.5 mM $MgCl_2$, 0.2 $mM\ dNTP$, 0.25 μM each primer, and 1.25 U FirePol DNA polymerase (Solis Biodyne, Tallin, Estonia). The PCR cycling conditions were initial denaturation at 94 °C for 2 min followed with 35 cycles of denaturation at 94 °C for 30 s, annealing at 64 °C for 1 min, elongation at 72 °C for 90 s, and final elongation at 72 °C for 5 min. The resulting PCR products were purified by ExoSAP-IT PCR Product Cleanup Reagent (Thermo Fisher Scientific) according to the manufacturer's instructions and sequenced using ABI 3730 DNA Analyzer (Applied Biosystems, Foster City, CA, USA). For sequencing PCR, the forward primer was used. Sequencing chromatograms (OSM Resource 3) were evaluated by

the Sequencing Analyses Software v5.2 (Applied Biosystems) according to the GRCh38.p13 (NC_000001.11) reference sequence.

2.5 Statistical Analysis in the Evaluation Phase of the Study

In the first round of analyses, Hardy-Weinberg equilibrium was assessed. Consequently, associations of variants with clinical data were evaluated for all variants with $MAF > 0.05$. The additive, dominant, and recessive genetic models were used for statistical evaluation. The additive genetic model was assessed by logistic regression and dominant and recessive models by the Pearson chi-square test. Associations between categorized values as genotype and clinical data were analyzed using the Pearson chi-square or two-sided Fisher's exact test. For evaluation of continuous variables such as age and $eQTL$, the Kruskal-Wallis test was used. The tested clinical variables were as follows: menopausal status (pre- vs. post-menopausal), stage, tumor size (pT), regional lymph node involvement (pN), histological type (invasive ductal vs. other invasive carcinomas), grade, expression of estrogen (ER), progesterone (PR), and HER2/ ERBB2 receptors, and Ki67 (negative vs. positive). The response to the neoadjuvant cytotoxic therapy was evaluated as partial or complete response (responders) versus stable or progressive disease (non-responders). Progression-free survival (PFS) was defined as the time elapsed between surgery and relapse. Overall survival (OS) was considered as the time elapsed from surgery to the patient's death. A study follow-up endpoint was set to 120 months (10 years), and therefore PFS and OS data were censored at 120 months and evaluated by the Kaplan-Meier method with the Breslow test. Further, a multivariate Cox regression analysis included variants significant in univariate analyses and factors affecting PFS and OS. Patients lost to follow-up were excluded from PFS analyses. The p-values are departures from twosided tests. The Benjamini-Hochberg false discovery rate (FDR) test was used for correction of multiple testing of all new associations in the evaluation set [**35**] and $q < 0.05$ was considered significant. Analyses were conducted in the statistical program $SPSS$ v16.0 (SPSS, Chicago, IL, USA).

3 Results

3.1 Subjects'Characteristics

Clinical features of the study population are shown in **Table 1**. The median PFS and OS were 78 months and 84 months, respectively. Tumor size, lymph node involvement, grade, stage, and molecular subtype were statistically significantly associated with both PFS and OS of patients in the evaluation set and ER and PR only with PFS. Of these factors, pT and pN represent stage (TNM staging system) and receptors and Ki67, which associates with grade [**36**], are used for stratification of patients into intrinsic subtypes. Considering these facts and the size effects, stage and subtype were included in the multivariate Cox regression as covariates.

3.2 KIF14 Germline Variability in the Testing Set of Patients

For the present study, raw data for KIF14 germline variability were extracted from the previously published dataset [**21**]. Overall, 86 variants in KIF14 were identified in 105 patients with raw sequencing data. Of these, 23 variants were below the tranche sensitivity threshold for indels (99.0%) and thus were considered as false positives. In total, 63 variants passed the filter, 58 of which were known and five that were novel according to dbSNP build 151 (for a list of all variants passing filters,

see OSM Resource 4). A total of 25 variants had $MAF > 0.05$ in the testing set. Two out of these variants had more than 50% missing data and were excluded from further analyses (rs4304619 and rs2794409). Finally, 23 variants were evaluated for $eQTL$, functional predictions, and clinical associations. Among them, three haplotype blocks, each composed of three variants (Block 1: rs12060793-rs74319334-rs12120084, Block 2: rs7543730-rs10753877-rs58696327, and Block 3: rs2808238-rs10577607-rs2794410), were identified.

Taken together, eight variants (OSM Resource 5) had informative and significant $eQTL$, functional predictions, or clinical associations with breast cancer characteristics in the testing set and underwent further evaluation.

3.3 KIF14 Germline Genotyping in the Evaluation Set of Breast Cancer Patients

The genotype distribution of the KIF14 variants in the evaluation set is shown in **Table 2**.

Firstly, associations of rs12060793, rs74319334, rs17448931, rs6665951, rs12120084, rs2808244, rs3806361, and rs3806362 variants with clinical features of breast cancer patients ($n = 808$) were evaluated. Schematic depiction of these variants in KIF14 gene is shown in **Fig. 1**. The linkage disequilibrium analysis showed that these variants can be treated as independent variants ($r^2 < 0.8$) (OSM Resource 6).

Patients carrying the minor allele A in rs12060793 had grade 3 tumors significantly more often than patients with the wild type GG genotype ($p = 0.048$), while patients with the wild C allele in rs3806362 more often had grade 3 or 2 than those bearing the rare TT genotype ($p = 0.049$). The association with grade was also significant for the rs12120084 in the additive model ($p = 0.036$). The positive Ki67 status occurred significantly more frequently in patients with the wild type GG compared with rare AA genotype carriers for rs12060793 ($p = 0.047$) and also patients with the wild type GG in rs12120084 had a Ki67 positive status more frequently than rare CC genotype carriers ($p = 0.034$). Patients with the minor allele T in rs3806362 had larger tumors (pT2-4) significantly more often than wild type CC carriers ($p = 0.040$).

Table 2 Distribution of evaluated candidate variants in KIF14 and their clinical associations in the evaluation set of breast cancer patients

Variant	Genotype	Patients		Minor allele frequencies		Frequency in population[b]	Clinical associations	
		N[a]	%	Testing set	Evaluation set		Testing set	Evaluation set
rs12060793								
G>A	GG	225	28	G=0.44	G = 0.53	G = 0.49	Grade (NS)	Grade ($p = 0.048$)
	GA	403	50				Ki67 (NS)	Ki67 ($p = 0.047$)
	AA	180	22				OS (NS)	OS ($p = 0.014$)
rs74319334								
T>A	TT	609	75	A = 0.19	A = 0.14	A = 0.15	Subtype[e] (NS)	Subtype[e] ($p = 0.004$)
	TA	182	22				TNBC subtype (NS)	TNBC subtype ($p = 0.006$)
	AA	17	2					
rs17448931[c]								
A>G	AA	605	78	G = 0.09	G = 0.12	G = 0.07	HER2/ERBB2 (NS)	HER2/ERBB2 ($p = 0.038$)
	AG	152	20				PFS ($p = 0.029$)	PFS (NS)
	GG	16	2				**OS ($p = 0.036$)**	**OS ($p = 0.010$)**
rs6665951								
T>C	TT	47	6	T = 0.22	T = 0.24	T = 0.23	PR (NS)	PR ($p = 0.023$)
	TC	276	34					
	CC	485	60					
rs12120084								
G>C	GG	313	39	C = 0.32	C = 0.37	C = 0.41	Grade (NS)	Grade ($p = 0.036$)
	CG	390	48				Ki67 (NS)	Ki67 ($p = 0.034$)
	CC	105	13				OS (NS)	OS ($p = 0.033$)
rs2808244[d]								
G>A	GG	626	78	A = 0.12	A = 0.12	A = 0.11	HER2/ERBB2 (NS)	HER2/ERBB2 ($p = 0.049$)
	GA	171	21				TNBC subtype (NS)	TNBC subtype ($p = 0.016$)
	AA	9	1					
rs3806361[d]								
C>A	CC	309	38	A = 0.36	A = 0.39	A = 0.27	ER (NS)	ER ($p = 0.046$)
	CA	372	46				PR (NS)	PR ($p = 0.010$)
	AA	125	16				**HER2/ERBB2 ($p = 0.047$)**	**HER2/ERBB2 ($p = 0.038$)**
							Ki67 ($p = 0.040$)	Ki67 (NS)
							Subtype[e] (NS)	Subtype[e] ($p = 0.017$)

Table 2 (continued)

Variant	Genotype	Patients		Minor allele frequencies		Frequency in population[b]	Clinical associations	
		N^a	%	Testing set	Evaluation set		Testing set	Evaluation set
rs3806362[d]								
C>T	CC	622	77	T = 0.10	T = 0.12	T = 0.11	pT (NS)	pT ($p = 0.040$)
	CT	168	21				Grade (NS)	Grade ($p = 0.049$)
	TT	16	2				**OS ($p = 0.023$)**	**OS ($p = 0.023$)**

Associations seen in the testing set and confirmed in the evaluation shown in bold

[a]*N number of patients, NS non-significant, $TNBC$ triple negative breast cancer*

[b]*MAF in European population according to $ALFA$ (Allele Frequency Aggregator) [**41**] based on genotyping of thousands of subjects.*

[c]*Results for thirty-five patients are missing due to repeated failure of PCR amplification process*

[d]*Results for two patients are missing due to repeated failure of PCR amplification process*

[e]*For subtype analyses, these patients were divided to non-luminal ($HER2$ or $TNBC$) versus luminal (A or B) subgroups.*

Furthermore, patients with the wild type genotypes AA in rs17448931 or GG in rs2808244 had tumors with the positive HER2/ERBB2 status significantly more often compared with minor allele G or A allele carriers ($p = 0.038$ and $p = 0.049$, respectively). The distribution of rs3806361 variants in the additive model was also significantly associated with HER2/ERBB2 status ($p = 0.038$). The negative PR status was more frequent in patients with the rare CC genotype in rs6665951 compared with wild T allele carriers (p =0.023). Moreover, patients with the minor allele A in rs3806361 more often had the PR-negative or ER-negative tumors than wild type CC carriers ($p = 0.010$ and $p = 0.046$, respectively).

The division of breast cancer patients into intrinsic molecular subtypes yielded several significant results. Patients with luminal A and B subtypes compared with those having $TNBC$ or $HER2$ subtypes significantly differed in the distribution of KIF14 rs74319334 and rs3806361 genotypes. Carriers of the minor allele A in rs74319334 significantly prevailed among patients with $TNBC$ or $HER2$ subtypes compared with wild type TT carriers ($p = 0.004$). This association was similarly strong in patients with $TNBC$ subtype only ($p = 0.006$). Carriage of the minor allele A in rs3806361 compared with the wild type CC genotype was associated with these subtypes as well ($p = 0.017$). Additionally, patients bearing the wild type GG genotype in rs2808244 had the $TNBC$ subtype significantly more often than carriers of minor allele A ($p = 0.016$).

Collectively, the association between HER2/ERBB2 and KIF14 rs3806361 observed in the testing set ($p = 0.047$) was replicated in the evaluation set ($p = 0.038$) (**Table 2**). No new associations (observed only in the evaluation set) passed the FDR correction for multiple testing ($q = 0.0006$).

3.4 The Relationship Between KIF14 Variants and Survival of Breast Cancer Patients

In the second round of analyses, associations between the distribution of variants and PFS or OS were analyzed in the whole evaluation set of patients and then patients were also stratified according to therapy type into adjuvant chemotherapy and hormonal therapy only treated subgroups.

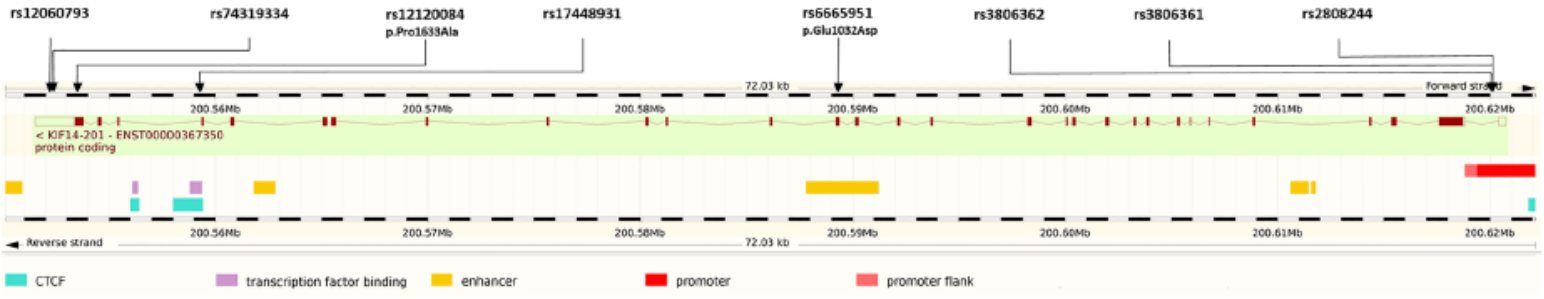

Fig. 1 Scheme of the KIF14 gene and depiction of evaluated candidate variants. Location of $SNPs$ and cú-regulatory regions in KIF14 (ENSG00000118193; Chromosome 1:200,551,500-200,620,751) in comparison with KIF14-201 transcript (ENST00000367350.5) by help of ENSEMBL (**https://www.ensembl.org**) and $dbSNP$ (**https:// www.ncbi.nlm.nih.gov/**) based on hg38 assembly. Amino acid changes for missense $SNPs$ (rs12120084 and rs6665951) are shown using NP_055690.1

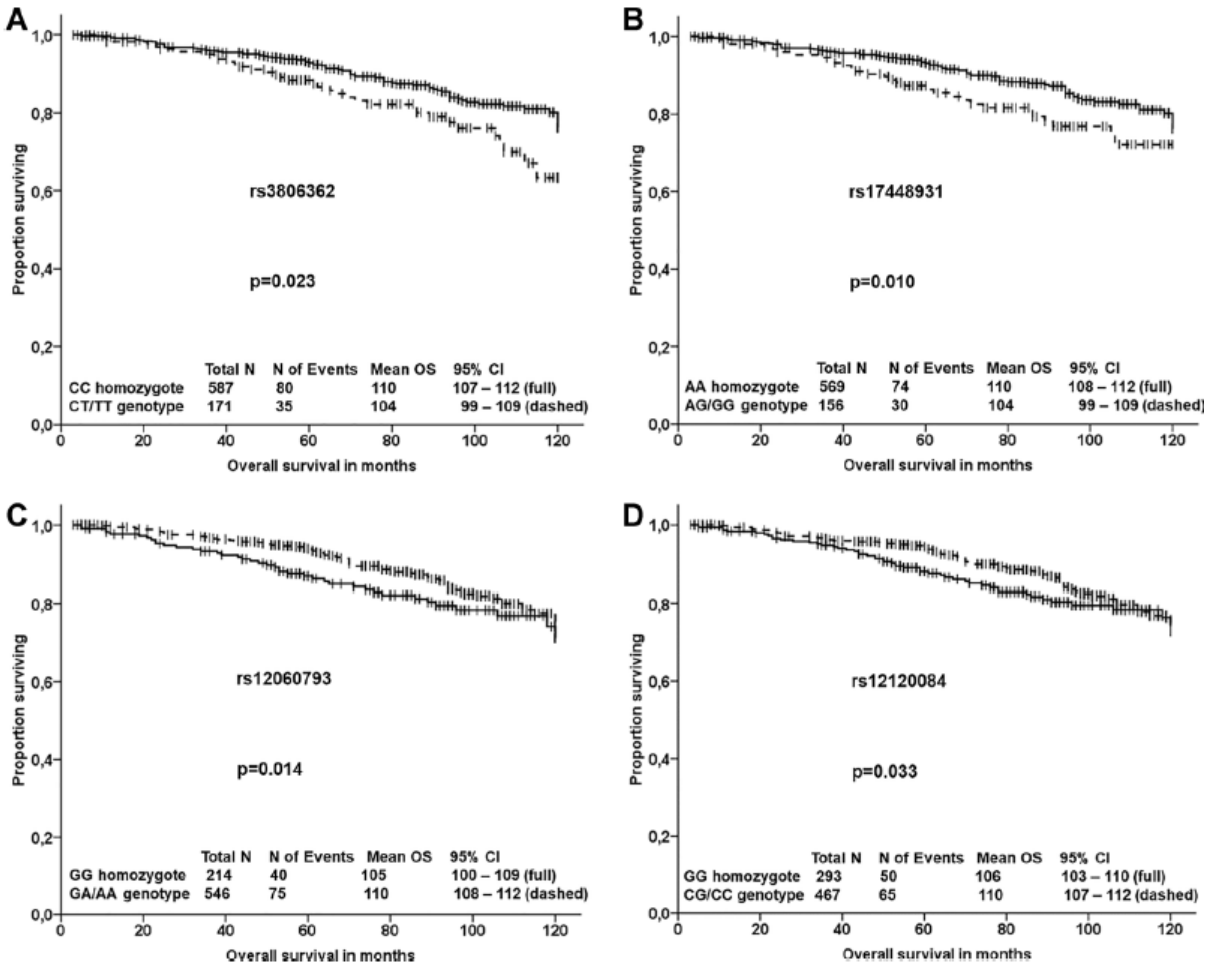

Fig. 2 Kaplan-Meier overall survival plots for KIF14 variants in the whole evaluation set of breast cancer patients. Plot **a** represents the rs3806362 variant, **b** rs17448931, **c** rs12060793, and **d** rs12120084. Patients with the wild type genotype (full line) were compared to minor allele carriers (dashed line)

The intrinsic molecular subtypes (luminal, $TNBC$, and $HER2$ subtypes) were used for further patient stratification. A multivariate Cox regression analysis was employed to evaluate the influence of potential confounding factors affecting PFS and OS (subtype and stage) in the evaluation set.

There were no statistically significant associations between KIF14 variants and PFS.

OS analysis identified that patients bearing the minor allele T in rs3806362 had significantly shorter OS compared with wild type CC genotype carriers ($p = 0.023$, **Fig. 2**a). The Cox regression analysis confirmed these results ($p = 0.006$; hazard ratio (HR) = 1.9; 95% confidence interval (CI) 1.2-3.0). Patients with the minor allele G in rs17448931 had significantly shorter OS than wild type AA genotype carriers (p = 0.010, **Fig. 2**b). This result was also confirmed by the multivariate Cox regression analysis ($p = 0.028$; $HR = 1.7$, 95% CI 1.1-2.8). Both associations were also seen in the testing set of patients (**Table 2**).

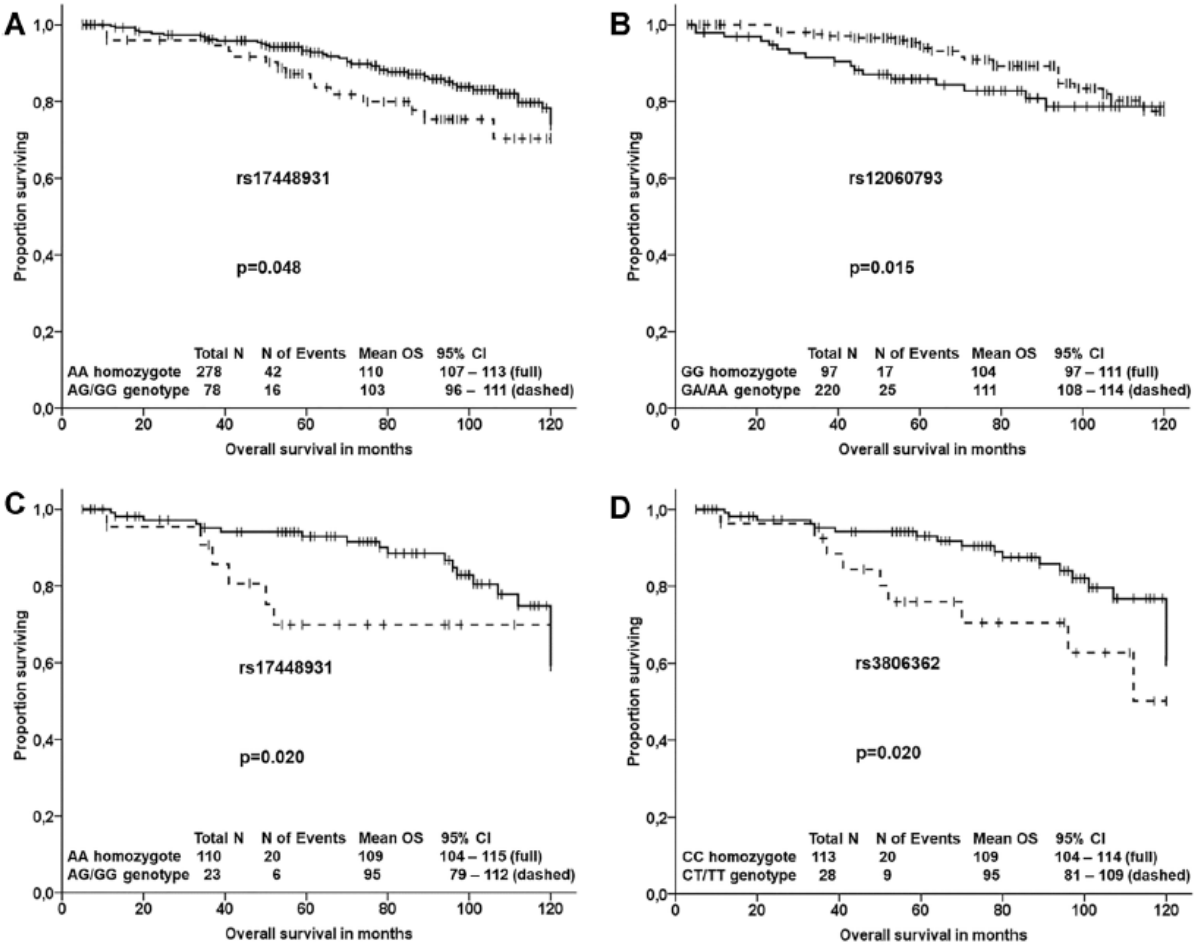

Fig. 3 Kaplan-Meier overall survival plots for KIF14 variants in breast cancer patients stratified by subtype and therapy. Plot a represents the rs17448931 variant in patients treated with adjuvant chemotherapy, b rs12060793 in the subgroup of patients treated only with hormonal therapy, c rs17448931, and d rs3806362 in the subgroup of patients with $TNBC$ or $HER2$ subtype. Patients with the wild type genotype (full line) were compared to minor allele carriers (dashed line)

Additionally, wild type GG carriers had significantly shorter OS than patients with minor alleles A (rs12060793) or C (rs12120084) ($p = 0.014$, **Fig. 2**c, and $p = 0.033$, **Fig. 2**d, respectively). However, these results were not confirmed by the multivariate Cox regression analysis and did not pass the FDR test ($q = 0.006$).

In analyses stratified by therapy, the rs17448931 was associated with OS in the subgroup of patients treated with adjuvant chemotherapy ($p = 0.048$, **Fig. 3**a) and the rs12060793 with OS in the subgroup of patients treated with hormonal therapy only ($p = 0.015$, **Fig. 3**b). The multivariate Cox regression analysis and the FDR test also failed in these cases ($q = 0.003$).

Unlike the subgroup of patients with luminal subtypes, where no association was demonstrated, in the subgroup of patients with $TNBC$ or $HER2$ subtype, a shorter OS was observed in patients carrying minor alleles G in rs17448931 ($p = 0.020$, **Fig. 3**c) or T in rs3806362 ($p = 0.020$, **Fig. 3**d). Furthermore, neither associations passed the FDR test ($q = 0.003$).

We further performed an in silico analysis of the candidate variants using the VarSome software. However, all variants were predicted to be benign according to the American College of Medical Genetics and Genomics ($ACMG$). Thus, these results must be treated with caution. Human Genome Variation Society ($HGVS$) nomenclature, prediction, and conservation scores are shown in **Table 3**.

The last step of statistical evaluation included survival analyses for the rs12060793-rs74319334-rs12120084 haplotype block (**Table 4**). Patients with ATC or ATG haplotypes had longer OS than those with other haplotypes ($p = 0.041$, **Fig. 4**a, and $p = 0.049$, **Fig. 4**b, respectively).

Table 3 In silico evaluation of candidate variants in KIF14

Variant	HGVS[a]	ACMG[b]	Conservation score[c]
rs12060793	c.*1104C>T	Benign	0.914
rs74319334	c.*1074A>T	Benign	−0.383
rs17448931	c.4353+118T>C	Benign	−0.164
rs6665951	c.3096A>G	Benign	0.08
rs12120084	c.4897C>G	Benign	−0.045
rs2808244	c.-408C>A	Benign	−2.014
rs3806361	c.-390G>T	Benign	−0.912
rs3806362	c.-387G>A	Benign	−3.277

[a]Human Genome Variation Society (HGVS) description based on NCBI Reference Sequence: NM_014875.3

[b]Evaluation based on the American College of Medical Genetics and Genomics (ACMG)

[c]PhyloP100 score based on multiple alignments of 99 vertebrate genome sequences to the human genome. The greater the score, the more conserved the site

Table 4 Distribution of the haplotype block in the evaluation set of breast cancer patients

Haplotype[a]	Count	Frequency, %
GTG	609	28.8
ATC	492	23.2
ATG	479	22.6
GAG	197	9.3
Other[b]	342	16.1

Due to heterozygosity some patients fall into more than one haplo-type group

[a]Haplotypes composed of rs12060793, rs74319334, and rs12120084 variants

[b]Haplotypes with frequencies < 5% are not individually presented

Moreover, significantly longer OS was seen in grouped patients with ATC and ATG haplotypes compared with others ($p = 0.014$, **Fig. 4**c) and in the same group of patients limited to those treated only with hormonal therapy ($p = 0.015$, **Fig. 4**d). The multivariate Cox regression analysis and the FDR test failed to confirm these results ($q = 0.002$).

In order to analyze the $eQTL$ in the evaluation set as well, we compared eight candidate $SNPs$ with transcript expression in tumors and control tissues (OSM Resource 7). None of the $SNPs$ were associated with KIF14 expression ($p > 0.05$).

4 Discussion

This two-phase study explored the previously unknown clinical and prognostic significance of KIF14 germline genetic variability in female breast cancer patients. In the in silico testing phase, we used the results of the previous study of 509 pharmacogenes and oncodrivers in 105 patients [**21**]. This analysis identified 23 germline variants in KIF14 with $MAF > 0.05$, which we consequently evaluated for clinical associations, $eQTL$, and functional predictions. As a result, we selected eight variants (rs12060793, rs74319334, rs17448931, rs6665951, rs12120084, rs2808244, rs3806361, and rs3806362) for the subsequent evaluation phase. Three of these variants (rs12060793, rs74319334, and rs12120084) composed a haplotype block with $r^2 > 0.8$.

The main results of the present study show that carriers of minor alleles G in rs17448931 or T in rs3806362 of KIF14 have significantly poorer OS compared with patients carrying the wild type ($p = 0.010$ and $p = 0.023$, respectively). These associations replicated the testing phase results and remained significant in the multivariate Cox regression analysis with molecular subtype and stage as covariates. Both variants represent the non-coding variation (intron variants) and have no established clinical significance (absent in ClinVar, accessed on 26 June 2022). The MAF in Caucasians is 0.07-0.08 for rs17448931 and 0.09-0.11 for rs3806362 in $ALFA$ or gnomAD databases, suggesting that smaller studies could have missed assessing their importance. We selected both variants for evaluation due to their significant associations with the survival of patients in the testing phase, but also for their functional annotations as predicted by the RegulomeDB. These annotations reached a high-rank 2a for rs3806362 and 3a for rs17448931. More specifically, the ChIP-seq in various biosamples identified 454 peaks for the rs3806362 and 73 for the rs17448931. The $FAIRE$-seq or DNase-seq has shown 32 results for the rs3806362 in various cell lines. This analysis indicated 11 results for the rs17448931, of which three hits were in the prototypical MCF7 breast carcinoma cell line, while the rest of the cell types had one or no result. Moreover, chromatin state analysis identified 126 hits on active translation start sites (TSS) for rs3806362, but none for rs17448931, where 66 hits were associated with strong transcription and 61 with even quiescent or low transcription. Based on the above data, we suggest both variants for further clinical and functional validation.

To the best of our knowledge, the present report is the first to show the implication of KIF14 germline variability for the prognosis of breast cancer patients. On the other hand, the association between KIF14 expression and the prognosis of breast cancer patients has been well described in several studies [**11**, **12**, **37**, **38**]. None of the KIF14 variants identified by our study reached significance in the $eQTL$ analysis. Thus, prognostic associations may not be relevant to the expression level of the parent gene, but instead to epigenetic events, including the above “open chromatin” states affecting protein-DNA interactions or KIF14 protein function.

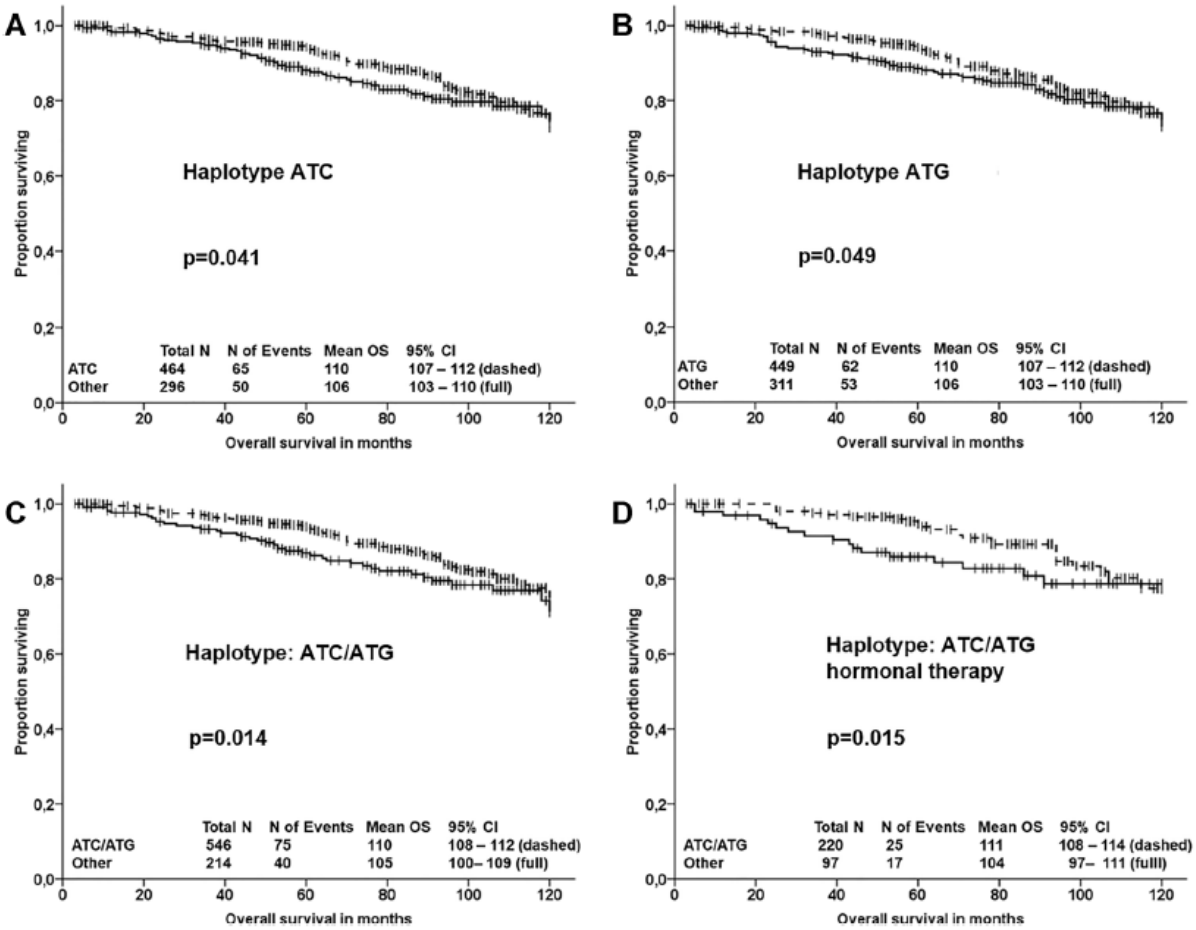

Fig. 4 Kaplan-Meier overall survival plots for breast cancer patients divided according to carriage of the KIF14 haplotype block rs12060793-rs74319334-rs12120084. Whole set of patients carrying ATC (**a**) or ATG (**b**) haplotypes, and both haplotypes combined compared with patients bearing the other haplotypes (**c**). Subgroup analysis of patients treated with hormonal therapy for both haplotypes combined versus the other haplotypes (**d**). Patients with the prevailing haplotype are represented by the dashed line and the rest by the solid line

A recent case-control study with bioinformatics analysis suggested a link between another variant, rs10800708, located within the KIF14 miRNA binding site, and genetic susceptibility to breast cancer [**39**]. The variant allele in the rs10800708 likely disrupts the binding site for miR-892a, miR-4252, and miR-5095, and creates a putative target site for miR-2114-3p. The carriage of the AA genotype or A allele was significantly associated with an increased breast cancer risk (odds ratio 4.8 and $p < 0.0001$ or 3.2 and $p = 0.0003$, respectively), but neither independent validation nor further functional evidence was reported [**39**].

Previous studies have demonstrated KIF14 as a putative therapeutic target in patients with aggressive $TNBC$ subtype [**14**, **20**, **38**]. Our study identified associations between several KIF14 variants and intrinsic molecular subtypes of breast cancer. Carriers of minor alleles in rs74319334 or rs3806361 significantly prevailed among patients with $TNBC$ or $HER2$ subtypes compared with wild type carriers. On the contrary, wild type homozygotes for rs2808244 had the $TNBC$ subtype significantly more often than minor allele carriers. The rs3806361 variant is particularly interesting because we also replicated its association with the $HER2/\ ERBB2$ expression status. This variant is in the 5'-untrans-lated region, which is critical for ribosome recruitment to the $mRNA$ transcript and choice of the start codon. Thus, these regions are recognized as important for the control of translation efficiency and shaping the cellular proteome [**40**]. Indeed, the RegulomeDB analysis identified 126 hits on an active TSS for this

variant, but a lack of $eQTL$. However, the connection between this variant and the $HER2/ERBB2$ expression status, molecular subtype, or tumor aggressiveness of breast carcinomas needs to be further addressed.

Survival analyses also suggested a putative prognostic role of the haplotype block composed of rs12060793, rs74319334, and rs12120084 variants. Patients with ATC or ATG haplotypes had significantly longer OS than patients with other haplotypes despite these findings not being confirmed by the multivariate Cox regression analysis, and did not pass the FDR test, suggesting that individual KIF14 variants provide better prognostic information. Similarly, in analyses stratified by therapy, the association between rs17448931 and OS in the subgroup of patients treated with adjuvant chemotherapy did not pass the above tests. The reasons behind this can be related to the small sample size of the compared subgroups and, thus, these results shall be interpreted with extreme caution, bearing in mind the described drawbacks.

There are several limitations to our study. First, the sample size of analyses stratified by subtype or therapy compared to studies carried out by large consortia or metaanalyses may be considered modest. On the other hand, we collected samples and clinical data from all contributing centers in a unified fashion avoiding heterogeneity in the sample type (e.g., blood or tissue), quality, and analytical approach. Second, the population composition, with the vast majority of patients being of Slavic Caucasian ethnicity, for now, precludes generalization, and independent studies in different populations should follow. Last, we do not provide a functional characterization of the most successful variants for strengthening the evidence and discerning causative from correlative associations. Nevertheless, in vitro or in vivo animal studies are cost- and time-consuming and sometimes lead to biased extrapolation, due to low genetic and phenotypic similarity between isolated cell types or animal models to real-life oncological patients. Instead, we provide cost-effective and detailed in silico predictions as starting points for further research.

To sum up, this pilot study provides the first evidence that KIF14 variants rs17448931 and rs3806362 can, independently of other relevant clinical risk factors, identify patients with worse prognosis from those with more favorable outcomes. Since these data are preliminary, we advocate further research on the prognostic relevance and clinical utility of germline KIF14 variants as revealed by the present study.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA A Cancer J Clin. 2021;71:209-49.

2. Neveu V, Nicolas G, Salek RM, Wishart DS, Scalbert A. Expo-some-Explorer 2.0: an update incorporating candidate dietary biomarkers and dietary associations with cancer risk. Nucleic Acids Res. 2020;48:D908-12.

3. Hanahan D. Hallmarks of cancer: new dimensions. Cancer Discov. 2022;12:31-46.

4. Lucanus AJ, Yip GW. Kinesin superfamily: roles in breast cancer, patient prognosis and therapeutics. Oncogene. 2018;37:833-8.

5. Miki H, Setou M, Kaneshiro K, Hirokawa N. All kinesin superfamily protein, KIF, genes in mouse and human. Proc Natl Acad Sci USA. 2001;98:7004-11.

6. Priyanga J, Guha G, Bhakta-Guha D. Microtubule motors in cen-trosome homeostasis: a target for cancer therapy? Biochim Bio-phys Acta Rev Cancer. 2021;1875: 188524.

7. Sheng L, Hao SL, Yang WX, Sun Y. The multiple functions of kinesin-4 family motor protein KIF4 and its clinical potential. Gene. 2018;678:90-9.

8. Wordeman L. How kinesin motor proteins drive mitotic spindle function: lessons from molecular assays. Semin Cell Dev Biol. 2010;21:260-8.

9. Lee K-Y, Davies T, Mishima M. Cytokinesis microtubule organisers at a glance. J Cell Sci. 2012;125:3495-500.

10. Van' t Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AAM, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature. 2002;415:530-6.

11. Corson TW, Huang A, Tsao M-S, Gallie BL. KIF14 is a candidate oncogene in the 1q minimal region of genomic gain in multiple cancers. Oncogene. 2005;24:4741-53.

12. Corson TW, Gallie BL. KIF14 mRNA expression is a predictor of grade and outcome in breast cancer. Int J Cancer. 2006;119:1088-94.

13. Gerashchenko TS, Zolotaryova SY, Kiselev AM, Tashireva LA, Novikov NM, Krakhmal NV, Cherdyntseva NV, Zavyalova MV, Perelmuter VM, Denisov EV. The activity of KIF14, Mieap, and EZR in a new type of the invasive component, torpedo-like structures, predetermines the metastatic potential of breast cancer. Cancers. 2020;12:1909.

14. Pariyar M, Johns A, Thorne RF, Scott RJ, Avery-Kiejda KA. Copy number variation in triple negative breast cancer samples associated with lymph node metastasis. Neoplasia. 2021;23:743-53.

15. Ehrlichova M, Mohelnikova-Duchonova B, Hrdy J, Brynychova V, Mrhalova M, Kodet R, Rob L, Pluta M, Gut I, Soucek P, et al. The association of taxane resistance genes with the clinical course of ovarian carcinoma. Genomics. 2013;102:96-101.

16. Brynychova V, Ehrlichova M, Hlavac V, Nemcova-Furstova V, Pecha V, Leva J, Trnkova M, Mrhalova M, Kodet R, Vrana D, et al. Genetic and functional analyses do not explain the association of high PRC1 expression with poor survival of breast carcinoma patients. Biomed Pharmacother. 2016;83:857-64.

17. Hanicinec V, Brynychova V, Rosendorf J, Palek R, Liska V, Oli-verius M, Kala Z, Mohelnikova-Duchonova B, Krus I, Soucek P. Gene expression of cytokinesis regulators PRC1, KIF14 and CIT has no prognostic role in colorectal and pancreatic cancer. Oncol Lett. 2021;22:598.

18. Xu H, Choe C, Shin SH, Park SW, Kim HS, Jung SH, Yim SH, Kim TM, Chung YJ. Silencing of KIF14 interferes with cell cycle progression and cytokinesis by blocking the p27(Kip1) ubiquitination pathway in hepatocellular carcinoma. Exp Mol Med. 2014;46: e97.

19. Zhu Q, Ren H, Li X, Qian B, Fan S, Hu F, Xu L, Zhai B. Silencing KIF14 reverses acquired resistance to sorafenib in hepatocellular carcinoma. Aging. 2020;12:22975-3003.

20. Singel SM, Cornelius C, Batten K, Fasciani G, Wright WE, Lum L, Shay JW. A targeted RNAi screen of the breast cancer genome identifies KIF14 and TLN1 as genes that modulate doc-etaxel chemosensitivity in triple-negative breast cancer. Clin Cancer Res. 2013;19:2061-70.

21. Hlavac V, Kovacova M, Elsnerova K, Brynychova V, Kozevnik-ovova R, Raus K, Kopeckova K, Mestakova S, Vrana D, Gatek J, et al. Use of germline genetic variability for prediction of chemoresistance and prognosis of breast cancer patients. Cancers. 2018;10:511.

22. Tavassoli F, Devilee P. Pathology and genetics of tumours of the breast and female genital organs. Lyon: IARC Press; 2003.

23. Cheang MCU, Chia SK, Voduc D, Gao D, Leung S, Snider J, Watson M, Davies S, Bernard PS, Parker JS, et al. Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. JNCI J Natl Cancer Inst. 2009;101:736-50.

24. Coates AS, Winer EP, Goldhirsch A, Gelber RD, Gnant M, Pic-cart-Gebhart M, Thurlimann B, Senn H-J, André F, Baselga J, et al. Tailoring therapies—improving the management of early breast cancer: St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2015. Ann Oncol. 2015;26:1533-46.

25. Schwartz LH, Litiere S, de Vries E, Ford R, Gwyther S, Man-drekar S, Shankar L, Bogaerts J, Chen A, Dancey J, et al. RECIST 1.1—update and clarification: from the RECIST committee. Eur J Cancer. 2016;62:132-7.

26. Topic E, Gluhak J. Isolation of restrictible DNA. Eur J Clin Chem Clin Biochem. 1991;29:327-30.

27. Li H. Toward better understanding of artifacts in variant calling from high-coverage samples. Bioinformatics. 2014;30:2843-51.

28. Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al. From FastQ data to high-confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinform. 2013;43:11101-111033.

29. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38:e164-e164.

30. 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.

31. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karczewski KJ, Park J, Hitz BC, Weng S, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. 2012;22:1790-7.

32. Barrett JC. 2009. Haploview: visualization and analysis of SNP genotype data. Cold Spring Harb Protoc. 2009:pdb.ip71.

33. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blu-menstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, et al. The structure of haplotype blocks in the human genome. Science. 2002;296:2225-9.

34. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. Primer3—new capabilities and interfaces. Nucleic Acids Res. 2012;40:e115-e115.

35. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Stat Soc Ser B (Methodol). 1995;57:289-300.

36. Liang Q, Ma D, Gao RF, Yu KD. Effect of Ki-67 expression levels and histological grade on breast cancer early relapse in patients with different immunohistochemical-based subtypes. Sci Rep. 2020;10:7648.

37. Li T-F, Zeng H-J, Shan Z, Ye R-Y, Cheang T-Y, Zhang Y-J, Lu S-H, Zhang Q, Shao N, Lin Y. Overexpression of kinesin superfamily members as prognostic biomarkers of breast cancer. Cancer Cell Int. 2020;20:123.

38. Singel SM, Cornelius C, Zaganjor E, Batten K, Sarode VR, Buckley DL, Peng Y, John GB, Li HC, Sadeghi N, et al. KIF14 promotes AKT phosphorylation and contributes to chemoresistance in triple-negative breast cancer. Neoplasia. 2014;16:247-256.e2.

39. Shasttiri A, Rostamian Delavar M, Baghi M, Dehghani Ashkezari M, Ghaedi K. SNP rs10800708 within the KIF14 miRNA binding site is linked with breast cancer. Br J Biomed Sci. 2019;76:46-8.

40. Hinnebusch AG, Ivanov IP, Sonenberg N. Translational control by 5'-untranslated regions of eukaryotic mRNAs. Science. 2016;352:1413-6.

41. Phan, L, Jin Y, Zhang H. Qiang W, Shekhtman E, Shao D, Revoe D, Villamarin R, Ivanchenko E, Kimura M, et al. ALFA: Allele Frequency Aggregator. National Center for Biotechnology Information, U.S. National Library of Medicine, 10 Mar. 2020. **http:// ww.ncbi.nlm.nih.gov/snp/docs/gsr/alfa/**.