

RESEARCH ARTICLE

Open Access

Breast tumors from *CHEK2 1100delC*-mutation carriers: genomic landscape and clinical implications

Taru A Muranen^{1,2}, Dario Greco¹, Rainer Fagerholm¹, Outi Kilpivaara¹, Kati Kämpjärvi¹, Kristiina Aittomäki³, Carl Blomqvist⁴, Päivi Heikkilä⁵, Åke Borg^{6,7,8} and Heli Nevanlinna^{1,2*}

Abstract

Introduction: *Checkpoint kinase 2 (CHEK2)* is a moderate penetrance breast cancer risk gene, whose truncating mutation *1100delC* increases the risk about twofold. We investigated gene copy-number aberrations and gene-expression profiles that are typical for breast tumors of *CHEK2 1100delC*-mutation carriers.

Methods: In total, 126 breast tumor tissue specimens including 32 samples from patients carrying *CHEK2 1100delC* were studied in array-comparative genomic hybridization (aCGH) and gene-expression (GEX) experiments. After dimensionality reduction with *CGHregions* R package, *CHEK2 1100delC*-associated regions in the aCGH data were detected by the Wilcoxon rank-sum test. The linear model was fitted to GEX data with R package *limma*. Genes whose expression levels were associated with *CHEK2 1100delC* mutation were detected by the bayesian method.

Results: We discovered four lost and three gained *CHEK2 1100delC*-related loci. These include losses of 1p13.3-31.3, 8p21.1-2, 8p23.1-2, and 17p12-13.1 as well as gains of 12q13.11-3, 16p13.3, and 19p13.3. Twenty-eight genes located on these regions showed differential expression between *CHEK2 1100delC* and other tumors, nominating them as candidates for *CHEK2 1100delC*-associated tumor-progression drivers. These included *CLCA1* on 1p22 as well as *CALCOCO1*, *SBEM*, and *LRP1* on 12q13. Altogether, 188 genes were differentially expressed between *CHEK2 1100delC* and other tumors. Of these, 144 had elevated and 44, reduced expression levels.

Our results suggest the WNT pathway as a driver of tumorigenesis in breast tumors of *CHEK2 1100delC*-mutation carriers and a role for the olfactory receptor protein family in cancer progression. Differences in the expression of the 188 *CHEK2 1100delC*-associated genes divided breast tumor samples from three independent datasets into two groups that differed in their relapse-free survival time.

Conclusions: We have shown that copy-number aberrations of certain genomic regions are associated with *CHEK2 1100delC* mutation. On these regions, we identified potential drivers of *CHEK2 1100delC*-associated tumorigenesis, whose role in cancer progression is worth investigating. Furthermore, poorer survival related to the *CHEK2 1100delC* gene-expression signature highlights pathways that are likely to have a role in the development of metastatic disease in carriers of the *CHEK2 1100delC* mutation.

Introduction

Large-scale gene-expression (GEX) profiling by DNA microarrays has become a routine method in breast cancer research. It has been widely used in tumor-subtype classification, and gene signatures predicting

clinical outcome have been defined in a number of studies [1-8].

The tumor genome is characterized by multiple translocations as well as gains and losses of chromosomal regions. Overall chromosomal instability is a result of arbitrary processes that take place during tumor progression, like breakage-fusion-bridge cycles [9], spontaneous breaks [10], and aberrant segregation of chromosomes in mitosis [11]. The cells that by chance

* Correspondence: heli.nevanlinna@hus.fi

¹Department of Obstetrics and Gynecology, Helsinki University Central Hospital, Haartmaninkatu 8, Helsinki, FI-00029, Finland
Full list of author information is available at the end of the article

gain advantageous changes survive and form a new more-malignant subpopulation of cancer cells.

Array-comparative genomic hybridization (aCGH) enables fine mapping of the locations of chromosomal breakpoints and detection of copy number on each location [12,13]. Combined aCGH and GEX profiling represents a valuable tool. aCGH can detect the events that have been selectively advantageous during the tumor progression as consistently gained or lost regions across a set of samples. GEX can further highlight the driver genes within these regions, revealing the ones whose expression levels have changed [14].

Understanding cellular processes that drive tumorigenesis is essential to select the optimal treatment for cancer [15]. Breast cancer can be divided into several subtypes that are characterized by differences in histopathologic features [16] as well as in genomic copy number [17,18] and gene-expression profiles [1,5,19,20]. These differences ultimately derive from different biological pathways that drive the tumor progression. For example, tumors from *BRCA1*- and *BRCA2*-mutation carriers exhibit distinct copy-number and gene-expression profiles and cluster into separate breast cancer subtypes [2,18,21], suggesting that events taking place during tumorigenesis are different and are influenced by the underlying inherited vulnerability of the key players, known as breast cancer tumor-suppressor genes. This raises a possibility also that moderate-penetrance mutations, such as *CHEK2 1100delC*, could be associated with certain genomic changes and gene-expression profiles.

CHEK2 (checkpoint kinase 2) is an intermediate-level breast cancer risk gene, whose truncating mutation *1100delC* doubles the risk of unselected women [22], but gives rise to much higher risk for women who have a family history of breast cancer [23]. Frameshift mutation *1100delC* causes premature translation stop at codon 381 in the middle of the kinase domain of the protein [24]. Truncated protein is highly unstable [25,26] as well as the mutated mRNA, which is rapidly degraded through nonsense-mediated mRNA decay [27].

The well-known role of *CHEK2* is to regulate cellular responses to DNA double-strand breaks (DSBs). *ATM* (ataxia telangiectasia mutated) phosphorylates *CHEK2* in response to DNA damage. This leads to *CHEK2* homodimerization, resulting in an active kinase whose targets include cell-cycle regulators *CDC25A*, *CDC25C*, *PLK* and *E2F1*, *BRCA1* involved in homologous recombination, as well as the master coordinator of apoptosis, *TP53* [28]. Recently *CHEK2* was reported to be involved also in regulation of the proper assembly of the mitotic spindle [29]. Even partial loss of *CHEK2* activity was sufficient to cause chromosomal instability.

Loss of 22q, where the *CHEK2* locus resides, is a common event in breast cancer [30,31]. Tumors from patients carrying *CHEK2* mutation *1100delC* show reduced *CHEK2* protein activity that can partly be explained by the loss of the wild-type allele in tumor cells [25,32]. However, loss of heterozygosity (LOH) does not comprehensively explain *CHEK2 1100delC*-related tumorigenesis, and likely also other molecular mechanisms exist, through which *CHEK2* deficiency contributes to malignant development [26,33]. *CHEK2 1100delC* mutation is associated with estrogen receptor (ER) positive and higher grade tumors, as well as with bilateral disease [34,35]. Breast cancer patients with *CHEK2 1100delC* have been found to have a poorer disease-free and overall survival [34,36], and this may partly be due to a disadvantageous response to treatment that causes DNA damage [37,38].

Here we report the first study investigating *CHEK2 1100delC*-related tumorigenesis by comparing gene-expression profiles and genomic changes in breast tumors of *CHEK2 1100delC*-mutation carriers to those of patients negative for any known germline mutations. As *CHEK2* deficiency is not restricted to tumors of *1100delC*-mutation carriers [25], similar copy number aberrations and gene-expression changes could also be present in the tumors of noncarriers, but enriched in the mutation carriers. The genomic aberrations as well as activated or silenced genes and pathways characteristic for *CHEK2 1100delC* mutation-carrier tumors are good candidates for *CHEK2 1100delC*-related tumor-progression drivers to be investigated further. Finally, we investigated the clinical significance of the discovered *CHEK2 1100delC* gene-expression signature.

Materials and methods

Patients and tumor tissue samples

In total, 126 tumor tissue samples from 121 breast cancer patients negative for *BRCA1*, *BRCA2*, or *TP53* germline mutations were obtained from Helsinki University Central Hospital, Department of Pathology. Thirty of the patients carried germline mutation *CHEK2 1100delC*. Tumor characteristics are described in detail in additional files (Additional file 1, Additional file 2).

The study was carried out with the permission from the Helsinki University Central Hospital Ethics Committee (Dnro207/E9/07) and with written informed consents from the patients. Both processed and raw data of aCGH and GEX experiments are available in Gene Expression Omnibus database [GEO: GSE24707] [39].

Gene-expression microarrays

Labeled tumor cDNA samples and reference samples (Universal Human RNA; Stratagene, La Jolla, CA, U.S.A.) were hybridized to custom-made cDNA microarrays

[GEO: GPL5345] produced at SCIBLU Genomics Centre, Lund University, Sweden, as described earlier [40]. Two samples were hybridized twice, and one sample, 3 times for quality control. Nucleic acid extraction, data acquisition, and preprocessing are described in detail in additional files (Additional file 2, Additional file 3).

Comparative genomic hybridization microarrays

The tiling-resolution genomic BAC (bacterial artificial chromosome) arrays [GEO: GPL4723] were described in detail previously [40], along with the slide treatment and hybridization protocol. Further details on nucleic acid extraction, hybridizations, data acquisition, preprocessing, and quality control are described in detail in additional files (Additional file 2, Additional file 3, Additional file 4).

aCGH data analysis

Data analysis was carried out in R software environment for statistical computing version 2.10. Soft calls (that is, loss, normal, and gain with probabilities) were calculated for each clone of the segmented data with the package *CGHcall* [41]. Sequences of clones with constant calls across all samples or across the *CHEK2 1100delC*-mutation carrier samples were defined by using R package *CGHregions* [42], allowing maximal information loss of 2.5%. Regions defined across all samples were tested for significant differences in calls between *CHEK2 1100delC*-mutation carrier and other tumors with the Wilcoxon rank-sum test, and those with nominal *P* values less than 0.01 were considered to be of interest. To assure that the results were not confounded by the estrogen receptor (ER) status, we performed similar analysis comparing ER-positive and ER-negative tumors to discover ER-associated regions. Regions defined only on basis of the *CHEK2 1100delC*-mutation carrier tumor samples by *CGHRegions* analysis were used primarily to explore the copy number of the genomic region covering the *CHEK2* locus.

Gene-expression data analysis

Gene-expression data analysis was carried out in R version 2.10 package *limma* [43-45]. For comparison of *CHEK2 1100delC* carrier and other tumors using multivariate analysis, a linear model with six covariates was fitted in the data. Specifically, as ER status is one of the most important tumor characteristics, which has a great impact also on the tumor gene-expression profile, we included it as a covariate in the multivariate linear model estimating the associations of the covariates on gene-expression levels in our dataset. The algorithm estimated which proportion of gene-expression changes was associated with *CHEK2*-mutation carrier status (independent of ER status), and simultaneously, which

proportion of gene expression was associated with ER status (independent of *CHEK2* mutation status) and similarly for all covariates. In addition to the tumor ER status and *CHEK2 1100delC*, the six covariates included a family history of breast cancer, tumor histopathologic types ductal or lobular, as well as *NQO1 rs1800566* genotype, for its rare allele has been associated with poor breast cancer survival [46], and the gene is likely to have a role in tumor progression. Frequencies of other histopathologic types were too low to have any detectable effect. For five samples, data for all covariates was not available, and these samples were omitted from the multivariate analysis. Statistical significance of differentially expressed genes was assessed with the empiric bayesian method. Genes with nominal *P* values less than 0.05 were considered to be differentially expressed. This resulted in a set of 862 genes that were used for functional enrichment analysis. The 188 genes used in the survival signature were selected by using a stricter *P*-value threshold, 0.01.

Functional annotation of differentially expressed genes

Functional annotation and enrichment analyses were carried out with the DAVID microarray functional annotation tool [47] and with the AmiGO [48] version 1.7 from the Gene Ontology [49,50] database during the period of July to October 2010, as well as from published literature. We also compared the *CHEK2 1100delC*-related list of 188 genes with previously published breast cancer gene-expression signatures designed for either subtype classification or survival prediction to detect possible similarities [1-8].

Survival analysis of public datasets

The survival effect of 188 *CHEK2 1100delC*-related genes was studied in a larger Helsinki gene-expression dataset of 183 unselected and familial breast tumor samples [GEO: GSE24450] as well as in three public datasets with the approach previously suggested [51]. The Helsinki dataset consisted of 151 breast cancer patients collected in three unselected cohorts previously described [35,46,52] and 32 additional familial breast cancer cases. The three public datasets consisted of two Swedish cohorts, one from Uppsala [GEO: GSE3494; GEO: GSE4922] [53,54] and another from Stockholm [GEO: GSE1456] [55]. Differences in sample selection and survival-analysis end points are summarized in additional files (Additional file 5). Of the 188 *CHEK2 1100delC*-related genes, 131 (69.7%) were present on the array used for the Helsinki dataset, and 108 (57.4%) on the array used for all Swedish datasets. The gene-expression matrices for the available genes were used as input of k-means clustering implemented in R v.2.11, with $k = 2$ and using 100,000 iterations to split the samples of each

dataset into two groups based on expression of the signature genes. Stabilization of the k-means results was secured by setting a random-number-generator schema for the initial k-means iteration. Differences in survival between the two groups of each dataset were evaluated by the log-rank test, and the Kaplan-Meier survival curves were generated for visualization.

Results

Areas with differing copy number

In the following sections, we use “copy number” to refer to the following three states: loss, normal, or gain. Seven chromosomal locations were found to differ in copy number between the *CHEK2 1100delC*-mutation carrier and other tumors, as defined by *P* values less than 0.01 when tested with the Wilcoxon rank-sum test. *CHEK2 1100delC*-associated regions included wide loss of 1p13.3-31.3, losses of 8p21.1-2, 8p23.1-2, and 17p12-13.1, as well as gains of 12q13.11-3, 16p13.3, and 19p13.3 (Figure 1, Table 1). The association with the *CHEK2* mutation was not confounded by the ER status (data not shown).

The genomic region of *CHEK2* locus in mutation-carrier tumors

The genomic region, which covered the *CHEK2* locus and where all *CHEK2 1100delC* samples had consistent calls, as defined by CGH regions analysis, spanned three successive BAC clones, almost 2 kb, and five genes. This region was lost in six, gained in four, and of normal copy number in 12 of the *CHEK2 1100delC* samples.

Gene-expression analysis

To estimate the independent effect of *CHEK2 1100delC*, we fitted a multivariate linear model on the gene-

expression data and included the tumor ER status and histopathologic type as well as patient’s family history of breast cancer and rs1800566 genotype as covariates. Comparison of gene-expression data from *CHEK2 1100delC*-mutation carrier and other tumors resulted in 188 (Additional file 6) and 862 differentially expressed genes with nominal *P*-value thresholds 0.01 and 0.05, respectively. Fold changes ranged from half to almost fourfold (Additional file 7). In the gene set resulting from the more stringent *P*-value cut-off, 144 genes had increased, and 44, reduced expression, and in the total of 862 genes, the respective numbers were 522 and 340.

Enriched functional groups among differentially expressed genes

The most significantly enriched functional group among the set of 862 differentially expressed genes was olfaction: altogether, 35 olfactory receptors (OR), 34 of them having elevated expression in *CHEK2 1100delC* tumors, were found (Additional file 7, Additional file 8). The 35 OR genes were located on 10 different chromosomes. Of these, 16 resided on chromosome 11, clustering in two blocks: 10 genes on 11p15.4 and five genes on a region flanking the centromere (11p11.2-11q12.1). Furthermore, we identified two chromosomal regions where more than two differentially expressed OR were located together with other differentially expressed genes, 6p21.33-6p22.1 and 12q13.11-12q13.3.

Other interesting enriched biologic entities among the set of genes with elevated expression (Additional file 8) were the interrelated WNT signaling pathway, cell adhesion, and calcium binding (Additional file 9). Also transcription regulation-related genes and members of zinc-finger, Ras-GEF, and EGF families were enriched.

In the set of genes with lower expression in *CHEK2 1100delC* tumors, the enriched functions (Additional file 10) included DNA modification, transcription, mRNA processing, ribosome, and translation. Additionally, we retrieved several genes related to mitochondria and cellular respiration, as well as to the cytoskeleton. Some genes involved in the recognition of DNA damage were downregulated as well.

Differentially expressed genes on regions with differing copy numbers

Significantly differentially expressed genes were found on four of the seven *CHEK2 1100delC*-associated regions: those on chromosomes 1, 8, 12, and 16. The overall level of gene expression in the *CHEK2 1100delC*-associated regions was not consistent with the corresponding copy number, which was as expected, because only partial overlap was noted between the sample sets in aCGH and GEX experiments. However, we identified a handful of genes as candidates driving *CHEK2*

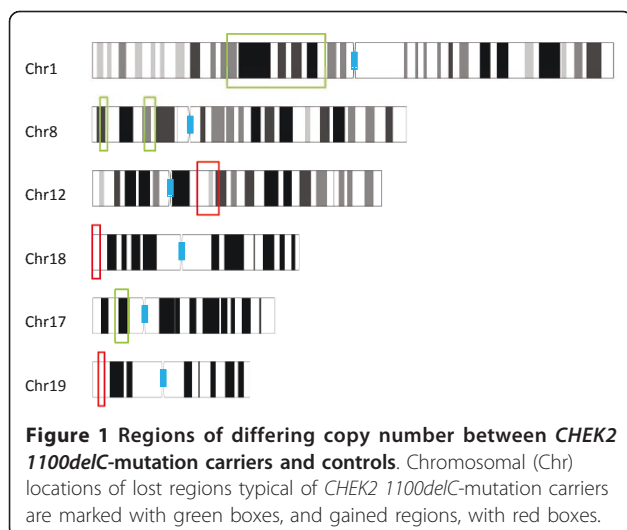


Table 1 Chromosomal regions with differing copy number between *CHEK2 1100delC*-mutation carriers and others

Chromosome	Start	End	Clones	Sum of clones	Length	Cytoband	P value
1	63762320	75867880	109	410	46865220	1p13.3-31.3	0.0031
1	76025260	87854764	110				0.0004
1	87964559	98030482	81				0.0001
1	98201778	110627540	110				0.0004
8	3329536	5083904	26	46	4279506	8p21.1-2	0.0062
8	5292190	6859255	15				0.0027
8	7041464	7609042	5				0.0066
8	24208215	29092984	35	41	5366218	8p23.1-2	0.0035
8	29293167	29574433	6				0.0064
12	46370995	47105108	10	103	9563151	12q13.11-3	0.0018
12	47191262	49654440	21				0.0022
12	49858115	51459707	21				0.0065
12	51492453	52261807	9				0.0025
12	52382320	54215964	23				0.0009
12	54367089	55934146	19				0.0044
16	11139	3142783	23	23	3131644	16p13.3	0.0091
17	9578676	15297799	54	54	5719123	17p12-13.1	0.0025
19	2122188	4798764	17	17	2676576	19p13.3	0.0045

Areas on chromosomes 1, 8, and 12 consist of multiple non-overlapping regions with differing copy numbers, whereas areas on chromosomes 16, 17, and 19 consist of single regions. Each row represents one region with constant calls across all samples. Sum of clones and length columns give the number of clones on all regions of each area and the total length of the area, respectively.

1100delC-related tumorigenesis. These included β -catenin-mediated transcription-related *CALCOCO1* [56], breast cancer poor-prognosis marker *SBEM* [57,58], and *SMARCC2*, a member of SWI/SNF chromatin-remodeling complex interacting with *BRCA1* [59]. All were located on 12q13 and showed elevated expression levels in *CHEK2 1100delC* tumors.

CLCA1 on 1p22.3 was among the top-ranking genes of differential expression, but its expression level was inconsistent with the *CHEK2 1100delC* typical copy number, for the region was frequently lost (Table 2).

Survival effect of 188 differentially expressed genes

We further tested whether the 188 *CHEK2 1100delC*-related genes (Additional file 7) might be associated with breast cancer prognosis. Based on their expression levels in the Helsinki and Uppsala gene-expression data sets, it was possible to divide the tumors into groups with different survival rates of the patients. Also in the third one, the smaller Stockholm dataset, there was evidence for survival difference, but it did not reach statistical significance. (Figure 2). Altogether, the *CHEK2 1100delC* gene-expression signature was associated with more rapid relapse of the disease in all three sample sets. In the largest data set (Uppsala), the *CHEK2 1100delC* signature also significantly also predicted 10-year breast cancer-specific survival.

Little similarity was seen between the gene list of 188 *CHEK2 1100delC*-associated genes and eight gene lists from previously published gene-expression signatures

predicting survival or breast cancer subtype (Table 3). Not more than about 1% of genes from any previously published gene list were shared by the *CHEK2 1100delC*-related list.

Discussion

We investigated the genomic copy number and gene-expression profiles of breast tumors of *CHEK2 1100delC*-mutation carriers. We aimed to identify patterns enriched in the *CHEK2 1100delC* group compared with the control group, to understand tumorigenesis, to which the *CHEK2* germline deficiency predisposes. We discovered two larger and five more narrow genomic regions whose copy-number aberrations were typical of *CHEK2 1100delC* tumors. We also identified 862 genes whose expression levels differed between the two groups of tumors, suggesting differential activity of especially WNT and olfactory pathways in these groups. Furthermore, we observed that the *CHEK2 1100delC* gene-expression signature is related to increased risk of breast cancer relapse.

Among the genes with lower expression levels in *CHEK2 1100delC* tumors, we observed several genes associated with centrosomes and the cytoskeleton or with DNA-damage signaling and apoptosis, as expected, owing to the role of *CHEK2* in response to DNA damage and in centrosome assembly. The apoptosis-related genes included pro-apoptotic *BAX* and *BAD* and DNA damage sensor *PARP1*.

BCL-2 family proteins, including *BAD* and *BAX*, are involved in the regulation of the normal breast tissue

Table 2 Differentially expressed genes on regions with differential copy number

Gene symbol	GEX: fold change	GEX: P value	Region	aCGH: gain/loss
<i>CLCA1</i>	1.8178	0.000007	chr1	Loss
<i>DPH5</i>	0.8254	0.0049	chr1	Loss
<i>C1orf62</i>	1.3277	0.0057	chr1	Loss
<i>ATXN7L2</i>	1.2804	0.007	chr1	Loss
<i>DR1</i>	0.8474	0.0108	chr1	Loss
<i>TMED5</i>	0.8728	0.0145	chr1	Loss
<i>ALG14</i>	0.8640	0.0389	chr1	Loss
<i>LRRC8D</i>	0.8854	0.0433	chr1	Loss
<i>DEFA5</i>	1.1076	0.0113	chr8	Loss
<i>DUSP4</i>	1.2724	0.0426	chr8	Loss
<i>OR6C3</i>	1.2558	0.0016	chr12	Gain
<i>CALCOCO1</i>	1.2579	0.0016	chr12	Gain
<i>SBEM</i>	3.0936	0.0217	chr12	Gain
<i>OR6C2</i>	1.1820	0.0485	chr12	Gain
<i>LRP1</i>	0.8171	0.0011	chr12	Gain
<i>OR10AD1</i>	1.2648	0.0061	chr12	Gain
<i>CSAD</i>	1.2758	0.0067	chr12	Gain
<i>C12orf10</i>	0.8867	0.0122	chr12	Gain
<i>MYO1A</i>	0.7468	0.0131	chr12	Gain
<i>LALBA</i>	1.2369	0.0187	chr12	Gain
<i>ITGA7</i>	1.2531	0.0201	chr12	Gain
<i>OR9K2</i>	1.4794	0.0231	chr12	Gain
<i>KRT5</i>	1.7237	0.0382	chr12	Gain
<i>FAM112B</i>	0.7840	0.0426	chr12	Gain
<i>SMARCC2</i>	1.1788	0.047	chr12	Gain
<i>KIAA1924</i>	0.8549	0.0156	chr16	Gain
<i>NTHL1</i>	0.8410	0.0233	chr16	Gain
<i>KIAA1171</i>	1.1890	0.0348	chr16	Gain

Five of eight differentially expressed genes on 1p13.3-31.3 show reduced expression levels. On 12q13.11-3, 11 of 15 genes show elevated expression. aCGH, array comparative genomic hybridization; GEX, gene expression.

development and involution. Their expression is often disturbed in malignant tissue, and reduced expression of BAD and BAX is associated with poor prognosis in breast cancer. All BCL-2 proteins regulate the mitochondrial apoptotic pathway. BAX induces apoptosis by increasing the permeability of the mitochondrial outer membrane to release cytochrome *c*. BAD is an upstream regulator of several pro- and antiapoptotic BCL-2 family proteins, BAX among them [60]. *BAX* is a transcriptional target of TP53 [61] and therefore also a downstream target of CHEK2. Its downregulation could be seen as a consequence of reduced CHEK2 activity, but at the same time, as a survival mechanism of the tumor cells.

aCGH analysis highlighted seven genomic locations, whose copy-number aberrations were enriched in tumors of *CHEK2 1100delC*-mutation carriers. These included losses of the proximal part of 1p, two regions

on 8p and a narrow region in the middle of 17p as well as gains of the proximal region on 12q and short distal regions on 16p and 19p. Loss of 8p is a common aberration in breast tumors, and several loci on 8p have been suggested to harbor tumor-suppressor genes [62]. Our data support these reports. The entire chromosome arm 8p was frequently lost in both *CHEK2 1100delC* and other tumors, but losses of certain sections of the arm were more frequent in *CHEK2 1100delC* tumors. Similarly, a region distal to 1p31 has been suggested to harbor tumor-suppressor genes, because it is commonly lost in breast and other solid tumors, and the loss is an early event in breast carcinogenesis [63]. Interestingly, this region covers only a small proportion of the *CHEK2 1100delC* associated more proximal to 1p13.3-31.3. Expression of a handful of genes located on 1p13.3-31.3 differed between *CHEK2 1100delC*-carrier tumors, the most significant of them being the calcium-activated chloride channel, *CLCA1*. *CLCA1* and its three paralogs *CLCA2*, *-3*, and *-4*, all reside on 1p22.3. *CLCA2* that shares 63% sequence similarity with *CLCA1* [64] has been suggested to be a breast cancer tumor-suppressor gene [65]. The *CLCA1* locus was lost in one third of *CHEK2 1100delC* tumors but only in 3.5% of other tumors. Conversely, *CLCA1* gene expression was significantly higher in *CHEK2 1100delC* tumors. From our data, we cannot state whether the loss of the *CLCA1* locus and elevated expression were simultaneous events. They could also be alternative events both affecting *CLCA1* normal function.

12q13 is a locus of a *WNT* gene cluster, hosting genes *WNT10B* and *WNT1*, as well as *WNT* pathway modifiers *CALCOCO1* and *LRP1* [66]. It is also a locus for a cluster of olfactory receptor genes. The results from GEX and aCGH analyses converge to highlight the importance of these two pathways. Expression of altogether 15 genes related to the *WNT* pathway was elevated. These included such central players as *WNT2* growth factor and *c-MYC* cellular oncogene. Expression differences of olfactory receptor genes were even more pronounced: up to 34 OR genes had higher expression in *CHEK2 1100delC*-mutation carriers than in the control samples.

In adult human breast tissue, *WNT2* expression is restricted to stromal cells, whereas in cancers, elevated *WNT2* expression has been detected also in epithelial cells, suggestive of an autocrine signaling loop [67]. Furthermore, elevated *WNT2* expression is a common feature in breast carcinomas and breast cancer cell lines [68,69]. It is associated with the epithelial mesenchymal transition, metalloproteinase induction through EGFR transactivation, invasive phenotype, and metastasis [67,69,70]. The oncogene *MYC* is one of the transcriptional targets of the *WNT* pathway, and it has been

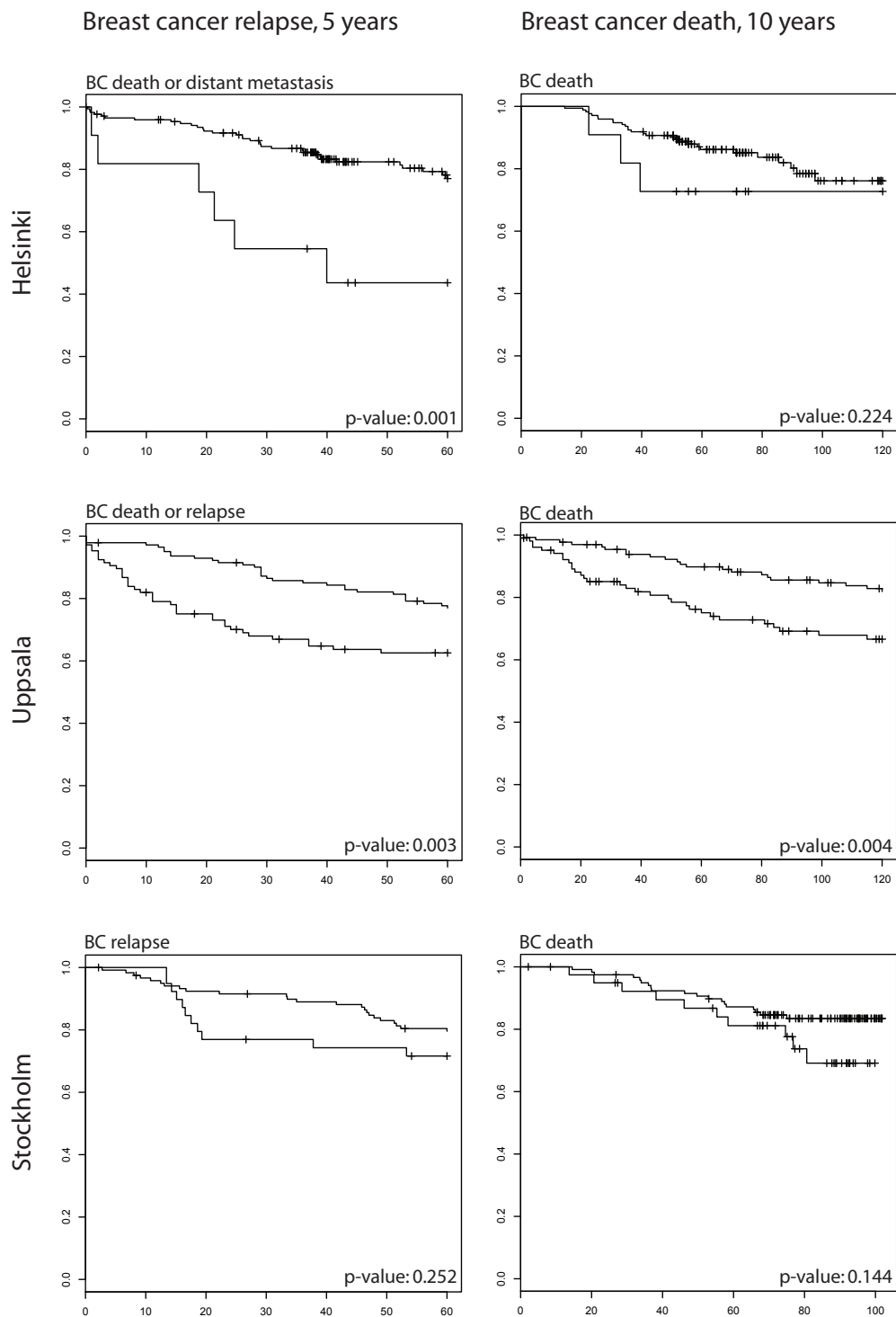


Figure 2 Breast cancer (BC)-related survival differences in three sample sets divided on the basis of the *CHEK2 1100delC* signature. Expression levels of the *CHEK2 1100delC*-related genes divide breast tumors into groups of better and poorer prognosis. Significant differences in overall or disease-free survival were detected in Helsinki and Uppsala cohorts, whereas in the Stockholm cohort, the differences were more subtle. The difference between the curves was measured with the log-rank test.

Table 3 Overlap between the 188 *CHEK2 1100delC*-associated genes and eight previously published gene-expression signatures

Gene-expression signature	Genes	Overlap	Proportion
SSP2003 [2]	500	5	0.010
SSP2006 [1]	306	2	0.007
PAM50 [5]	50	1	0.020
70 gene poor-prognosis signature [3]	70	0	0.000
76 gene metastasis signature [4]	76	0	0.000
Gene-expression grade index [6]	128	1	0.008
Recurrence score [7]	21	0	0.000
Recurrence predictor in tamoxifen-treated patients [8]	181	2	0.011
GO: proliferation [49,50]	1677	13	0.008

The first column contains names and references for the gene-expression signatures as well as the reference for proliferation-associated genes in the Gene Ontology database (GO). The second to fourth columns indicate the number of genes included in the signature, the number overlapping with each signature, and the 188 genes and proportion of genes of each signature among the 188 genes, respectively.

shown to enhance its own expression by promoting WNT signaling in tumors [71].

Elevated expression levels of *WNT2*, *MYC*, its transcriptional co-activator *MAX*, as well as several components of the WNT pathway, suggest that the pathway has a central role in *CHEK2 1100delC*-related tumorigenesis. The genes with elevated expression were involved in canonic and calcium-signaling branches of the WNT pathway. Interestingly, the expression of *LEF1*, the main transcription factor of the pathway, as well as *LRPI*, a cell-surface co-receptor of WNT ligand, was lower in *CHEK2 1100delC* tumors when compared with others. This, together with elevated *E cadherin* (*CDH1*) levels, implies that some balancing forces might act in control of the WNT pathway in the *CHEK2 1100delC* tumors.

Olfactory receptors (ORs) form the biggest mammalian gene family, with more than 800 genes present in the human genome and only about 45% of them being functional. ORs are located in clusters that are dispersed across almost all chromosomes. Great interindividual variation exists in the copy number of some OR genes, as well as in the proportion of functional alleles, as a result of low selection pressure [72,73]. ORs bind specifically a wide range of small chemical compounds, including steroid hormones and their analogs, and function in odor recognition [74,75]. Until now, only one OR gene, *prostate-specific G-protein-coupled receptor* (*PSGR/OR51E2*), has been connected to cancer. PSGR is activated by steroid hormones, and the activated receptor has been found to suppress prostate tumor cell proliferation through SAPK/JNK and p38 [75].

In addition to the neurons of the forebrain olfactory bulb [76], ORs are widely expressed across different tissues, but their functions in these locations remain obscure. ORs have been proposed to be involved in cell-cell recognition and in the organization of cells in developing tissues during embryogenesis [77], an analogous

function to the one that they have in the organization of the olfactory bulb [78]. But, since OR expression in different cell types does not show a clear pattern that would argue for specific functionality, it has been suggested that OR expression would partially result from transcriptional activity of neighboring genes [79]. The existence of neighboring driver genes is one possibility to explain the elevated expression OR genes in *CHEK2 1100delC* breast tumors. Our data do not strongly support it, although it cannot be totally excluded. On the basis of our observations, we suggest that the role of the olfactory receptor protein family in cancer cell proliferation and tissue invasion should be investigated further, as these are the two roles in which ORs have been previously implicated.

Converging evidence from the gene-expression and aCGH datasets studied here had the potential to reveal biologic functionalities and pathways that were relevant in *CHEK2 1100delC*-related tumorigenesis. This study concentrated on genomic-level changes that were associated with the *1100delC* germline mutation, whereas previous studies have shed light on the *CHEK2* protein expression in breast tumors. The protein expression is reduced or absent in the mutation-carrier tumors, but mechanisms leading to this may vary, and LOH has been reported for only part of *1100delC*-carrier tumors [25,26,32,33]. Here we found that the *CHEK2* locus was lost in fewer than one third of our mutation-carrier samples and gained in some, with the majority not showing evidence for copy-number variation. We have also determined *CHEK2* transcript levels in lymphoblastoid cell lines heterozygous for *CHEK2 1100delC* and have shown that constitutive *CHEK2* gene expression is significantly reduced in cells from heterozygous carriers of *1100delC* (fold change, 0.73; two-sided Student *t*-test $P = 3.4 \times 10^{-7}$; Muranen et al, unpublished data). These results support the hypothesis of haploinsufficiency being a possible factor driving *CHEK2 1100delC*-related tumorigenesis [80].

We further found an association between the *CHEK2 1100delC* signature of 188 genes and breast cancer relapse. Gene-expression differences of these genes were able to categorize patients into two groups that differed in their risk of breast cancer relapse. Association was observed in all three independent datasets, even though the analysis end points differed slightly between the data from different studies. In the largest dataset, the *CHEK2 1100delC* signature significantly predicted 10-year breast cancer-specific survival. This result is consistent with the previous reports suggesting that the *1100delC* germline genotype associates with worse survival for the carriers [36]. Further studies of larger materials will reveal whether this might be due to a worse therapy response.

Little overlap was found between the *CHEK2 1100delC* gene signature and previously published signatures. For the gene-expression analysis here, a multivariate analysis was used to identify *CHEK2 1100delC*-associated effects that were independent of possible confounding effects, especially the ER status, which is the most important mediator of breast cancer biology and subtypes. The number of genes shared between the published signatures overall is very low, even though they have been shown to perform equally well in predicting clinical outcome. It has been suggested that the prediction capability of these signatures is based on the presence of proliferation-associated genes [81]. Conversely, among the 188 *CHEK2 1100delC*-related genes, only 13 were associated with proliferation. Altogether, this suggests that the *CHEK2 1100delC*-associated signature of 188 genes is distinct from other signatures specifically built to predict breast cancer survival or classify tumors into subtypes and reflects the downstream effects of the germline *CHEK2 1100delC* mutation on tumor progression.

Conclusions

Altogether, differences between *CHEK2 1100delC* and others tumors were not strong, but we were able to point out chromosomal aberrations and gene-expression changes that were more pronounced in the *CHEK2 1100delC* group. Furthermore, the poor survival associated with the *CHEK2 1100delC* gene-expression signature suggests differential progression pathways in breast tumors with defective *CHEK2*. Our results converge to highlight the importance of the WNT pathway as a driver of tumorigenesis in breast tumors of *CHEK2 1100delC*-mutation carriers. Functional studies are needed to reveal the roles of different components of the pathway in different stages of cancer progression. Another interesting target for future investigation is the role of olfactory receptor family in cancer progression.

Additional material

Additional file 1: Tumor characteristics of study samples. Numbers of samples from estrogen-negative and -positive tumors, patients with or without a family history of breast cancer, patients with different rs1800566 genotypes, as well as tumors of different histologic and molecular subtypes in groups of *CHEK2 (checkpoint kinase 2)*-mutation carrier and other tumors for both gene-expression (GEX) and array-comparative genomic hybridization (aCGH) datasets. Molecular subtypes are defined on the basis of Immunohistochemistry results: Luminal A: ER⁺/PR⁺, HER2⁻; Luminal B: ER⁺/PR⁺, HER2⁺; HER2 positive: ER⁺/PR⁻, HER2⁺; Basal: ER⁻/PR⁻, HER2⁻, EGFR⁺; Other triple negative: ER⁻/PR⁻, HER2⁻, EGFR⁻. All data were not available for every sample. EGFR, epidermal growth factor receptor; ER, estrogen receptor; Her2, human epidermal growth factor receptor 2; PR, progesterone receptor.

Additional file 2: Supplementary methods. [82-94].

Additional file 3: Unsupervised clustering of array-comparative genomic hybridization (aCGH) and gene-expression (GEX) data. Left, aCGH samples do not cluster according to any of the covariates. Instead, the clustering is based on informative genomic regions and tumors' overall copy-number profiles. Right, unsupervised hierarchical clustering of gene-expression data suggests that estrogen receptor (ER) status and family history of breast cancer have an impact on the tumor's gene-expression profile. The effect of the other variables is likely to be smaller. Replicates that were used as a quality control cluster together as expected. Positions of *CHEK2 (checkpoint kinase 2)*, *1100delC*-mutation carrier and control (contr) tumors are indicated by the uppermost color block. *NQO1⁻* and *NQO1⁺* stand for samples with germline rs1800566 genotypes: homozygotes of the more common allele (CC) and heterozygotes or homozygotes of the rarer allele (CT or TT), respectively. Frequency of copy-number changes along the genome is depicted with a color scale from blue to red. Blue, few; red, frequent copy-number aberrations.

Additional file 4: Pair-wise correlation plot of aCGH (array-comparative genomic hybridization) quality control (QC) hybridizations. Heatmap of pair-wise correlation between the hybridization from fresh frozen tissue sample (fresh) and the dye-swap hybridizations from formalin-fixed paraffin-embedded (paraf) tissue.

Additional file 5: Sample-selection criteria and survival-analysis end points in Uppsala, Stockholm, and Rotterdam cohorts.

Additional file 6: Hierarchical clustering of samples according to 188 differentially expressed genes. All but two *CHEK2 (checkpoint kinase 2) 1100delC*-mutation carriers cluster together in two branches. Estrogen receptor (ER) status of each sample, either positive or negative, is indicated by the second color block.

Additional file 7: The 862 differentially expressed genes. Gene identifiers (ID), chromosomal locations, fold change (FC), *P* value, and Benjamini-Hochberg (BH) multiple testing corrected *P* value.

Additional file 8: Functional enrichment of differentially expressed genes with elevated expression levels. The most frequent annotation categories are involved in olfactory transduction, WNT pathway or zinc finger, Ras-GEF, or EGF protein families. FDR, false discovery rate; Pop, population.

Additional file 9: Differentially expressed genes associated with WNT pathway, cell adhesion, or calcium. Columns "*CHEK2 M*" and "other *M*" give average and range of expression values for each gene in *CHEK2 1100delC*-mutation carrier and other tumor samples correspondingly.

Additional file 10: Functional enrichment of differentially expressed genes with reduced expression levels. The most frequent annotation categories are involved in the protein-expression pathway, mitochondria, or cytoskeleton. (Pop, population; FDR, false discovery rate).

Abbreviations

aCGH: array comparative genomic hybridization; ATM: ataxia telangiectasia mutated; BAC: bacterial artificial chromosome; CDH1: E cadherin; CHEK2: checkpoint kinase 2; CTNND1: catenin delta; DSB: DNA double strand break; ER: estrogen receptor; FFPE: formalin fixed paraffin embedded; GEO: gene expression omnibus; GEX: gene expression; LOH: loss of heterozygosity; MB: megabase; OR: olfactory receptor; PSGR: prostate-specific G protein-coupled receptor.

Acknowledgements

The research leading to these results received funding from the Helsinki University Central Hospital Research Fund, the Sigrid Juselius Foundation, the Finnish Cancer Society, the Academy of Finland (132473), and the European Community's Seventh Framework Programme under grant agreement 223175 (HEALTH-F2-2009-223175).

We thank Dr. Kirsimari Aaltonen, as well as research nurses Hanna Jäntti and Irja Erkkilä, for their invaluable help in patient data collection and management. We thank the personnel of SCIBLU Genomics, Lund, for their willing and professional assistance with array handling, and Dr. Johan Vallon-Christersson, Dr. Markus Ringnér, and Dr. Göran Jönsson for their contribution to data management and guidance in using BASE environment as well as for their critical and constructive comments on the manuscript. The Finnish Cancer Registry is gratefully acknowledged for the cancer diagnostic and follow-up data.

Author details

¹Department of Obstetrics and Gynecology, Helsinki University Central Hospital, Haartmaninkatu 8, Helsinki, FI-00029, Finland. ²Institute for Molecular Medicine Finland, FIMM, Tukholmankatu 8, Helsinki, FI-00014, Finland. ³Department of Clinical Genetics, Helsinki University Central Hospital, Haartmaninkatu 2B, Helsinki, FI-00029, Finland. ⁴Department of Oncology, Helsinki University Central Hospital, Haartmaninkatu 4, Helsinki, FI-00014, Finland. ⁵Department of Pathology, Helsinki University Central Hospital, Haartmaninkatu 3, Helsinki, FI-00014, Finland. ⁶Department of Oncology, Clinical Sciences, Lund University and Skåne University Hospital, Barngatan 2B, SE 22185 Lund, Sweden. ⁷CREATE Health Strategic Center for Translational Cancer Research, Lund University, BMC C13, SE 22184, Lund, Sweden. ⁸Lund Strategic Research Center for Stem Cell Biology and Cell Therapy, Lund University, BMC B10, SE 22184, Lund, Sweden.

Authors' contributions

HN conceived the study and wrote the manuscript. ÅB conceived the study. RF conceived the study and performed all the steps from nucleic acid extraction to microarray data acquisition. OK and KK performed all the steps from nucleic acid extraction to microarray data acquisition. TM performed the GEX and aCGH data analyses and wrote the manuscript. DG performed the survival analyses, performed the GEX and aCGH data analyses, and wrote the manuscript. PH, CB, and KA contributed samples and clinical information. All authors contributed to and approved the final manuscript for publication.

Competing interests

The authors declare that they have no competing interests.

Received: 21 October 2010 Revised: 4 April 2011

Accepted: 20 September 2011 Published: 20 September 2011

References

- Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF, Livasy C, Carey LA, Reynolds E, Dressler L, Nobel A, Parker J, Ewend MG, Sawyer LR, Wu J, Liu Y, Nanda R, Tretiakova M, Ruiz Orrico A, Dreher D, Palazzo JP, Perreard L, Nelson E, Mone M, Hansen H, Mullins M, Quackenbush JF, Ellis MJ, Olopade OI, Bernard PS, Perou CM: **The molecular portraits of breast tumors are conserved across microarray platforms.** *BMC Genomics* 2006, **7**:96.
- Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lonning PE, Brown PO, Borresen-Dale AL, Botstein D: **Repeated observation of breast tumor subtypes in independent gene expression data sets.** *Proc Natl Acad Sci USA* 2003, **100**:8418-8423.
- van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH: **Gene expression profiling predicts clinical outcome of breast cancer.** *Nature* 2002, **415**:530-536.
- Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F, Talantov D, Timmermans M, Meijer-van Gelder ME, Yu J, Jatkoe T, Berns EM, Atkins D, Foekens JA: **Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer.** *Lancet* 2005, **365**:671-679.
- Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z, Quackenbush JF, Stijleman IJ, Palazzo J, Marron JS, Nobel AB, Mardis E, Nielsen TO, Ellis MJ, Perou CM, Bernard PS: **Supervised risk predictor of breast cancer based on intrinsic subtypes.** *J Clin Oncol* 2009, **27**:1160-1167.
- Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, Smeds J, Nordgren H, Farmer P, Praz V, Haibe-Kains B, Desmedt C, Larsimont D, Cardoso F, Peterse H, Nuyten D, Buyse M, Van de Vijver MJ, Bergh J, Piccart M, Delorenzi M: **Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis.** *J Natl Cancer Inst* 2006, **98**:262-272.
- Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Park T, Hiller W, Fisher ER, Wickerham DL, Bryant J, Wolmark N: **A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer.** *N Engl J Med* 2004, **351**:2817-2826.
- Loi S, Haibe-Kains B, Desmedt C, Wirapati P, Lallemand F, Tutt AM, Gillet C, Ellis P, Ryder K, Reid JF, Daidone MG, Pierotti MA, Berns EM, Jansen MP, Foekens JA, Delorenzi M, Bontempi G, Piccart MJ, Sotiriou C: **Predicting prognosis using molecular profiling in estrogen receptor-positive breast cancer treated with tamoxifen.** *BMC Genomics* 2008, **9**:239.
- Maser RS, DePinho RA: **Connecting chromosomes, crisis, and cancer.** *Science* 2002, **297**:565-569.
- Murnane JP: **Telomere loss as a mechanism for chromosome instability in human cancer.** *Cancer Res* 2010, **70**:4255-4259.
- Ganem NJ, Godinho SA, Pellman D: **A mechanism linking extra centrosomes to chromosomal instability.** *Nature* 2009, **460**:278-282.
- Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F, Pinkel D: **Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors.** *Science* 1992, **258**:818-821.
- Pinkel D, Seagraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y, Dairkee SH, Ljung BM, Gray JW, Albertson DG: **High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays.** *Nat Genet* 1998, **20**:207-211.
- Hyman E, Kauraniemi P, Hautaniemi S, Wolf M, Mousset S, Rozenblum E, Ringner M, Sauter G, Monni O, Elkahloun A, Kallioniemi OP, Kallioniemi A: **Impact of DNA amplification on gene expression patterns in breast cancer.** *Cancer Res* 2002, **62**:6240-6245.
- Goldhirsch A, Ingle JN, Gelber RD, Coates AS, Thurlimann B, Senn HJ, Panel members: **Thresholds for therapies: highlights of the St Gallen International Expert Consensus on the primary therapy of early breast cancer 2009.** *Ann Oncol* 2009, **20**:1319-1329.
- Blows FM, Driver KE, Schmidt MK, Broeks A, van Leeuwen FE, Wesseling J, Cheang MC, Gelmon K, Nielsen TO, Blomqvist C, Heikkilä P, Heikkinen T, Nevanlinna H, Akslen LA, Begin LR, Foulkes WD, Couch FJ, Wang X, Cafourek V, Olson JE, Baglietto L, Giles GG, Severi G, McLean CA, Southey MC, Rakha E, Green AR, Ellis IO, Sherman ME, Lissowska J, Anderson WF, Cox A, Cross SS, Reed MW, Provenzano E, Dawson SJ, Dunning AM, Humphreys M, Easton DF, Garcia-Closas M, Caldas C, Pharoah PD, Huntsman D: **Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short and long term survival: a collaborative analysis of data for 10,159 cases from 12 studies.** *PLoS Med* 2010, **7**:e1000279.
- Waddell N, Arnold J, Cocciardi S, da Silva L, Marsh A, Riley J, Johnstone CN, Orloff M, Assie G, Eng C, Reid L, Keith P, Yan M, Fox S, Devilee P, Godwin AK, Hogervorst FB, Couch F, kConFab Investigators, Grimmond S, Flanagan JM, Khanna K, Simpson PT, Lakhani SR, Chenevix-Trench G: **Subtypes of familial breast tumours revealed by expression and copy number profiling.** *Breast Cancer Res Treat* 2010, **123**:661-677.
- Jonsson G, Staaf J, Vallon-Christersson J, Ringner M, Holm K, Hegerdt C, Gunnarsson H, Fagerholm R, Strand C, Agnarsson BA, Kilpivaara O, Luts L, Heikkilä P, Aittomäki K, Blomqvist C, Loman N, Malmstrom P, Olsson H, Johannsson OT, Arason A, Nevanlinna H, Barkardottir RB, Borg A: **Genomic subtypes of breast cancer identified by array-comparative genomic hybridization display distinct molecular and clinical characteristics.** *Breast Cancer Res* 2010, **12**:R42.

19. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Eystein Lønning P, Borresen-Dale AL: **Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.** *Proc Natl Acad Sci USA* 2001, **98**:10869-10874.
20. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, van der Velde T, Bartelink H, Rodenhuis S, Rutgers ET, Friend SH, Bernard R: **A gene-expression signature as a predictor of survival in breast cancer.** *N Engl J Med* 2002, **347**:1999-2009.
21. van Beers EH, van Welsem T, Wessels LF, Li Y, Oldenburg RA, Devilee P, Cornelisse CJ, Verhoef S, Hogervorst FB, van't Veer LJ, Nederlof PM: **Comparative genomic hybridization profiles in human BRCA1 and BRCA2 breast tumors highlight differential sets of genomic aberrations.** *Cancer Res* 2005, **65**:822-827.
22. CHEK2 Breast Cancer Case-Control Consortium: **CHEK2*1100delC and susceptibility to breast cancer: a collaborative analysis involving 10,860 breast cancer cases and 9,065 controls from 10 studies.** *Am J Hum Genet* 2004, **74**:1175-1182.
23. Fletcher O, Johnson N, Dos Santos Silva I, Kilpivaara O, Aittomaki K, Blomqvist C, Nevanlinna H, Wasielewski M, Meijers-Heijerboer H, Broeks A, Schmidt MK, Van't Veer LJ, Bremer M, Dork T, Chekmariova EV, Sokolenko AP, Imyanitov EN, Hamann U, Rashid MU, Brauch H, Justenhoven C, Ashworth A, Peto J: **Family history, genetic testing, and clinical risk prediction: pooled analysis of CHEK2 1100delC in 1,828 bilateral breast cancers and 7,030 controls.** *Cancer Epidemiol Biomarkers Prev* 2009, **18**:230-234.
24. Bell DW, Varley JM, Szydo TE, Kang DH, Wahrer DC, Shannon KE, Lubratovich M, Verselis SJ, Isselbacher KJ, Fraumeni JF, Birch JM, Li FP, Garber JE, Haber DA: **Heterozygous germ line hCHK2 mutations in Li-Fraumeni syndrome.** *Science* 1999, **286**:2528-2531.
25. Vahteristo P, Bartkova J, Eerola H, Syrjakoski K, Ojala S, Kilpivaara O, Tamminen A, Kononen J, Aittomaki K, Heikkilä P, Holli K, Blomqvist C, Bartek J, Kallioniemi OP, Nevanlinna H: **A CHEK2 genetic variant contributing to a substantial fraction of familial breast cancer.** *Am J Hum Genet* 2002, **71**:432-438.
26. Staalens V, Falck J, Geisler S, Bartkova J, Borresen-Dale AL, Lukas J, Lillehaug JR, Bartek J, Lønning PE: **Alternative splicing and mutation status of CHEK2 in stage III breast cancer.** *Oncogene* 2004, **23**:8535-8544.
27. Anczukow O, Ware MD, Buisson M, Zetoune AB, Stoppa-Lyonnet D, Sinilnikova OM, Mazoyer S: **Does the nonsense-mediated mRNA decay mechanism prevent the synthesis of truncated BRCA1, CHK2, and p53 proteins?** *Hum Mutat* 2008, **29**:65-73.
28. Nevanlinna H, Bartek J: **The CHEK2 gene and inherited breast cancer susceptibility.** *Oncogene* 2006, **25**:5912-5919.
29. Stolz A, Ertych N, Kienitz A, Vogel C, Schneider V, Fritz B, Jacob R, Dittmar G, Weichert W, Petersen I, Bastians H: **The CHK2-BRCA1 tumour suppressor pathway ensures chromosomal stability in human somatic cells.** *Nat Cell Biol* 2010, **12**:492-499.
30. Chen LC, Kurisu W, Ljung BM, Goldman ES, Moore D, Smith HS: **Heterogeneity for allelic loss in human breast cancer.** *J Natl Cancer Inst* 1992, **84**:506-510.
31. Williams LH, Choong D, Johnson SA, Campbell IG: **Genetic and epigenetic analysis of CHEK2 in sporadic breast, colon, and ovarian cancers.** *Clin Cancer Res* 2006, **12**:6967-6972.
32. Oldenburg RA, Kroeze-Jansema K, Kraan J, Morreau H, Klijn JG, Hoogerbrugge N, Ligtenberg MJ, van Asperen CJ, Vasen HF, Meijers C, Meijers-Heijboer H, de Bock TH, Cornelisse CJ, Devilee P: **The CHEK2*1100delC variant acts as a breast cancer risk modifier in non-BRCA1/BRCA2 multiple-case families.** *Cancer Res* 2003, **63**:8153-8157.
33. Sodha N, Bullock S, Taylor R, Mitchell G, Guertl-Lackner B, Williams RD, Bevan S, Bishop K, McGuire S, Houlston RS, Eeles RA: **CHEK2 variants in susceptibility to breast cancer and evidence of retention of the wild type allele in tumours.** *Br J Cancer* 2002, **87**:1445-1448.
34. de Bock GH, Schutte M, Krol-Warmerdam EM, Seynaeve C, Blom J, Brekelmans CT, Meijers-Heijboer H, van Asperen CJ, Cornelisse CJ, Devilee P, Tollenaar RA, Klijn JG: **Tumour characteristics and prognosis of breast cancer patients carrying the germline CHEK2*1100delC variant.** *J Med Genet* 2004, **41**:731-735.
35. Kilpivaara O, Bartkova J, Eerola H, Syrjakoski K, Vahteristo P, Lukas J, Blomqvist C, Holli K, Heikkilä P, Sauter G, Kallioniemi OP, Bartek J, Nevanlinna H: **Correlation of CHEK2 protein expression and c.1100delC mutation status with tumor characteristics among unselected breast cancer patients.** *Int J Cancer* 2005, **113**:575-580.
36. Schmidt MK, Tollenaar RA, de Kemp SR, Broeks A, Cornelisse CJ, Smit VT, Peterse JL, van Leeuwen FE, Van't Veer LJ: **Breast cancer survival and tumor characteristics in premenopausal women carrying the CHEK2*1100delC germline mutation.** *J Clin Oncol* 2007, **25**:64-69.
37. Broeks A, Braaf LM, Huseinovic A, Nooijen A, Urbanus J, Hogervorst FB, Schmidt MK, Klijn JG, Russell NS, Van Leeuwen FE, Van't Veer LJ: **Identification of women with an increased risk of developing radiation-induced breast cancer: a case only study.** *Breast Cancer Res* 2007, **9**:R26.
38. Chrisanthar R, Knappskog S, Lokkevik E, Anker G, Ostenstad B, Lundgren S, Berge EO, Risberg T, Mjaaland I, Maehle L, Engebretsen LF, Lillehaug JR, Lønning PE: **CHEK2 mutations affecting kinase activity together with mutations in TP53 indicate a functional pathway associated with resistance to epirubicin in primary breast cancer.** *PLoS One* 2008, **3**:e3062.
39. *Gene Expression Omnibus* [http://www.ncbi.nlm.nih.gov/geo/].
40. Jonsson G, Staaf J, Olsson E, Heidenblad M, Vallon-Christersson J, Osoegawa K, de Jong P, Oredsson S, Ringner M, Hoglund M, Borg A: **High-resolution genomic profiles of breast cancer cell lines assessed by tiling BAC array comparative genomic hybridization.** *Genes Chromosomes Cancer* 2007, **46**:543-558.
41. van de Wiel MA, Kim KI, Vosse SJ, van Wieringen WN, Wilting SM, Ylstra B: **CGHcall: calling aberrations for array CGH tumor profiles.** *Bioinformatics* 2007, **23**:892-894.
42. van de Wiel MA, van Wieringen WN: **CGHregions: dimension reduction for array CGH data with minimal information loss.** *Cancer Inform* 2007, **3**:55-63.
43. Smyth GK: **Linear models and empirical Bayes methods for assessing differential expression in microarray experiments.** *Stat Appl Genet Mol Biol* 2004, **3**:Article3.
44. Smyth GK, Speed T: **Normalization of cDNA microarray data.** *Methods* 2003, **31**:265-273.
45. Ritchie ME, Silver J, Oshlack A, Holmes M, Diyagama D, Holloway A, Smyth GK: **A comparison of background correction methods for two-colour microarrays.** *Bioinformatics* 2007, **23**:2700-2707.
46. Fagerholm R, Hofstetter B, Tommiska J, Aaltonen K, Vrtel R, Syrjakoski K, Kallioniemi A, Kilpivaara O, Mannermaa A, Kosma VM, Uusitupa M, Eskelinen M, Kataja V, Aittomaki K, von Smitten K, Heikkilä P, Lukas J, Holli K, Bartkova J, Blomqvist C, Bartek J, Nevanlinna H: **NAD(P)H:quinone oxidoreductase 1 NQO1*2 genotype (P187S) is a strong prognostic and predictive factor in breast cancer.** *Nat Genet* 2008, **40**:844-853.
47. Huang da W, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.** *Nat Protoc* 2009, **4**:44-57.
48. Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S, AmiGO Hub, Web Presence Working Group: **AmiGO: online access to ontology and annotation data.** *Bioinformatics* 2009, **25**:288-289.
49. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: **Gene ontology: tool for the unification of biology: The Gene Ontology Consortium.** *Nat Genet* 2000, **25**:25-29.
50. **The Gene Ontology.** [http://www.geneontology.org/].
51. Lukes L, Crawford NP, Walker R, Hunter KW: **The origins of breast cancer prognostic gene expression profiles.** *Cancer Res* 2009, **69**:310-318.
52. Syrjakoski K, Vahteristo P, Eerola H, Tamminen A, Kivinummi K, Sarantausta L, Holli K, Blomqvist C, Kallioniemi OP, Kainu T, Nevanlinna H: **Population-based study of BRCA1 and BRCA2 mutations in 1035 unselected Finnish breast cancer patients.** *J Natl Cancer Inst* 2000, **92**:1529-1531.
53. Miller LD, Smeds J, George J, Vega VB, Vergara L, Ploner A, Pawitan Y, Hall P, Klaar S, Liu ET, Bergh J: **An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival.** *Proc Natl Acad Sci USA* 2005, **102**:13550-13555.
54. Ivshina AV, George J, Senko O, Mow B, Putti TC, Smeds J, Lindahl T, Pawitan Y, Hall P, Nordgren H, Wong JE, Liu ET, Bergh J, Kuznetsov VA, Miller LD: **Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer.** *Cancer Res* 2006, **66**:10292-10301.
55. Pawitan Y, Bjohle J, Amler L, Borg AL, Eghazi S, Hall P, Han X, Holmberg L, Huang F, Klaar S, Liu ET, Miller L, Nordgren H, Ploner A, Sandelin K, Shaw PM, Smeds J, Skoog L, Wedren S, Bergh J: **Gene expression profiling**

- sparing early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. *Breast Cancer Res* 2005, **7**:R953-64.
56. Yang CK, Kim JH, Li H, Stallcup MR: Differential use of functional domains by coiled-coil coactivator in its synergistic coactivator function with beta-catenin or GRIP1. *J Biol Chem* 2006, **281**:3389-3397.
57. Miksicek RJ, Myal Y, Watson PH, Walker C, Murphy LC, Leygue E: Identification of a novel breast- and salivary gland-specific, mucin-like gene strongly expressed in normal and tumor human mammary epithelium. *Cancer Res* 2002, **62**:2736-2740.
58. Skliris GP, Hube F, Gheorghiu I, Mutawe MM, Penner C, Watson PH, Murphy LC, Leygue E, Myal Y: Expression of small breast epithelial mucin (SBEM) protein in tissue microarrays (TMAs) of primary invasive breast cancers. *Histopathology* 2008, **52**:355-369.
59. Bochar DA, Wang L, Beniya H, Kinev A, Xue Y, Lane WS, Wang W, Kashanchi F, Shiekhattar R: **BRCA1 is associated with a human SWI/SNF-related complex: linking chromatin remodeling to breast cancer.** *Cell* 2000, **102**:257-265.
60. Schorr K, Li M, Krajewski S, Reed JC, Furth PA: **Bcl-2 gene family and related proteins in mammary gland involution and breast cancer.** *J Mammary Gland Biol Neoplasia* 1999, **4**:153-164.
61. Vogelstein B, Lane D, Levine AJ: **Surfing the p53 network.** *Nature* 2000, **408**:307-310.
62. Cooke SL, Pole JC, Chin SF, Ellis IO, Caldas C, Edwards PA: **High-resolution array CGH clarifies events occurring on 8p in carcinogenesis.** *BMC Cancer* 2008, **8**:288.
63. Ragnarsson G, Eiriksdottir G, Johannsdottir JT, Jonasson JG, Egilsson V, Ingvarsson S: **Loss of heterozygosity at chromosome 1p in different solid human tumours: association with survival.** *Br J Cancer* 1999, **79**:1468-1474.
64. Gruber AD, Schreier KD, Ji HL, Fuller CM, Pauli BU: **Molecular cloning and transmembrane structure of hCLCA2 from human lung, trachea, and mammary gland.** *Am J Physiol* 1999, **276**:C1261-70.
65. Li X, Cowell JK, Sossey-Alaoui K: **CLCA2 tumour suppressor gene in 1p31 is epigenetically regulated in breast cancer.** *Oncogene* 2004, **23**:1474-1480.
66. Katoh M: **WNT and FGF gene clusters (review).** *Int J Oncol* 2002, **21**:1269-1273.
67. Dale TC, Weber-Hall SJ, Smith K, Huguet EL, Jayatilake H, Gusterson BA, Shuttleworth G, O'Hare M, Harris AL: **Compartment switching of WNT-2 expression in human breast tumors.** *Cancer Res* 1996, **56**:4320-4323.
68. Watanabe O, Imamura H, Shimizu T, Kinoshita J, Okabe T, Hirano A, Yoshimatsu K, Konno S, Aiba M, Ogawa K: **Expression of twist and wnt in human breast cancer.** *Anticancer Res* 2004, **24**:3851-3856.
69. Schlange T, Matsuda Y, Lienhard S, Huber A, Hynes NE: **Autocrine WNT signaling contributes to breast cancer cell proliferation via the canonical WNT pathway and EGFR transactivation.** *Breast Cancer Res* 2007, **9**:R63.
70. Huber MA, Kraut N, Beug H: **Molecular requirements for epithelial-mesenchymal transition during tumor progression.** *Curr Opin Cell Biol* 2005, **17**:548-558.
71. Cowling VH, Cole MD: **Turning the tables: Myc activates Wnt in breast cancer.** *Cell Cycle* 2007, **6**:2625-2627.
72. Young JM, Endicott RM, Parghi SS, Walker M, Kidd JM, Trask BJ: **Extensive copy-number variation of the human olfactory receptor gene family.** *Am J Hum Genet* 2008, **83**:228-242.
73. Hasin Y, Olender T, Khen M, Gonzaga-Jauregui C, Kim PM, Urban AE, Snyder M, Gerstein MB, Lancet D, Korbel JO: **High-resolution copy-number variation map reflects human olfactory receptor diversity and evolution.** *PLoS Genet* 2008, **4**:e1000249.
74. Keller A, Zhuang H, Chi Q, Vossahl LB, Matsunami H: **Genetic variation in a human odorant receptor alters odour perception.** *Nature* 2007, **449**:468-472.
75. Neuhaus EM, Zhang W, Gelis L, Deng Y, Noldus J, Hatt H: **Activation of an olfactory receptor inhibits proliferation of prostate cancer cells.** *J Biol Chem* 2009, **284**:16218-16225.
76. Firestein S: **How the olfactory system makes sense of scents.** *Nature* 2001, **413**:211-218.
77. Dreyer WJ: **The area code hypothesis revisited: olfactory receptors and other related transmembrane receptors may function as the last digits in a cell surface code for assembling embryos.** *Proc Natl Acad Sci USA* 1998, **95**:9072-9077.
78. Feinstein P, Mombaerts P: **A contextual model for axonal sorting into glomeruli in the mouse olfactory system.** *Cell* 2004, **117**:817-831.
79. Feldmesser E, Olender T, Khen M, Yanai I, Ophir R, Lancet D: **Widespread ectopic expression of olfactory receptor genes.** *BMC Genomics* 2006, **7**:121.
80. Sodha N, Mantoni TS, Tavtigian SV, Eeles R, Garrett MD: **Rare germ line CHEK2 variants identified in breast cancer families encode proteins that show impaired activation.** *Cancer Res* 2006, **66**:8966-8970.
81. Desmedt C, Haibe-Kains B, Wirapati P, Buyse M, Larsimont D, Bontempi G, Delorenzi M, Piccart M, Sotiriou C: **Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes.** *Clin Cancer Res* 2008, **14**:5158-5165.
82. Smyth GK, Yang YH, Speed T: **Statistical issues in cDNA microarray data analysis.** *Methods Mol Biol* 2003, **224**:111-136.
83. Smyth GK: **Limma: linear models for microarray data.** In *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Edited by: Gentleman R, Carey V, Dudoit S, Izarrary R, Huber W. New York: Springer; 2005:397-420.
84. R Development Core Team: **R: A Language and Environment for Statistical Computing.** R Foundation for Statistical computing, Vienna, Austria; 2010.
85. Silver JD, Ritchie ME, Smyth GK: **Microarray background correction: maximum likelihood estimation for the normal-exponential convolution.** *Biostatistics* 2009, **10**:352-363.
86. Staaf J, Jonsson G, Ringner M, Vallon-Christersson J: **Normalization of array-CGH data: influence of copy number imbalances.** *BMC Genomics* 2007, **8**:382.
87. Saal LH, Troein C, Vallon-Christersson J, Gruber S, Borg A, Peterson C: **BioArray Software Environment (BASE): a platform for comprehensive management and analysis of microarray data.** *Genome Biol* 2002, **3**:SOFTWARE0003.
88. Vallon-Christersson J, Nordborg N, Svensson M, Hakkinen J: **BASE: 2nd generation software for microarray data management and analysis.** *BMC Bioinformatics* 2009, **10**:330.
89. Olshen AB, Venkatraman ES, Lucito R, Wigler M: **Circular binary segmentation for the analysis of array-based DNA copy number data.** *Biostatistics* 2004, **5**:557-572.
90. Venkatraman ES, Olshen AB: **A faster circular binary segmentation algorithm for the analysis of array CGH data.** *Bioinformatics* 2007, **23**:657-663.
91. Rhead B, Karolchik D, Kuhn RM, Hinrichs AS, Zweig AS, Fujita PA, Diekhans M, Smith KE, Rosenbloom KR, Raney BJ, Pohl A, Pheasant M, Meyer LR, Learned K, Hsu F, Hillman-Jackson J, Harte RA, Giardine B, Dreszer TR, Clawson H, Barber GP, Haussler D, Kent WJ: **The UCSC Genome Browser database: update 2010.** *Nucleic Acids Res* 2010, **38**:D613-9.
92. Van Wieringen WN, Van De Wiel MA, Ylstra B: **Weighted clustering of called array CGH data.** *Biostatistics* 2008, **9**:484-500.
93. Rennstam K, Ahlstedt-Soini M, Baldetorp B, Bendahl PO, Borg A, Karhu R, Tanner M, Tirkkonen M, Isola J: **Patterns of chromosomal imbalances defines subgroups of breast cancer with distinct clinical features and prognosis: a study of 305 tumors by comparative genomic hybridization.** *Cancer Res* 2003, **63**:8861-8868.
94. Chin K, DeVries S, Fridlyand J, Spellman PT, Roydasgupta R, Kuo WL, Lapuk A, Neve RM, Qian Z, Ryder T, Chen F, Feiler H, Tokuyasu T, Kingsley C, Dairkee S, Meng Z, Chew K, Pinkel D, Jain A, Ljung BM, Esserman L, Albertson DG, Waldman FM, Gray JW: **Genomic and transcriptional aberrations linked to breast cancer pathophysiologies.** *Cancer Cell* 2006, **10**:529-541.

doi:10.1186/bcr3015

Cite this article as: Muranen et al.: Breast tumors from CHEK2 1100delC-mutation carriers: genomic landscape and clinical implications. *Breast Cancer Research* 2011 **13**:R90.