

研究论文

乳腺癌特异性拷贝数变异致DNA修复失衡

高 猛¹ 程竹君² 张 成¹ 熊招平² 廖尉廷² 邓立彬^{1,2} 卢曲琴^{1*}(南昌大学医学部, 南昌 330006; ²南昌大学转化医学研究院, 南昌 330006)

摘要 该研究通过整合4个功能注释数据库(KEGG、GO、WIKI和PC)共发现56条DNA修复相关通路, 确定了725个功能基因; 并在筛查CCLE肿瘤细胞系和TCGA乳腺癌基因组数据的基础上, 发现了89个乳腺癌特异性的拷贝数变异基因, 其中包含了4个DNA修复相关的基因[*BRCA1*(breast cancer 1, early onset)、*ERBB2*(erb-b2 receptor tyrosine kinase 2)、*G6PC*(glucose-6-phosphatase, catalytic subunit)和*PSME3*(proteasome (prosome, macropain) activator subunit 3 (PA28 gamma; Ki))]. 进一步生物信息分析表明, 乳腺癌基因组中17q21.31区域的拷贝数变异可通过改变*BRCA1*和*PSME3*基因的表达, 从而影响DNA修复功能(如双链的断裂修复和DNA损伤相关检测点).

关键词 乳腺癌; 拷贝数变异; DNA修复

Imbalance of DNA Repair Caused by Breast Cancer Specific Copy Number Variation

Gao Meng¹, Cheng Zhujun², Zhang Cheng¹, Xiong Zhaoping², Liao Weiting², Deng Libin^{1,2}, Lu Quqin^{1*}¹School of Medicine, Nanchang University, Nanchang 330006, China;²Institute of Translational Medicine, Nanchang University, Nanchang 330006, China)

Abstract We integrated 56 pathways related to DNA repair from databases (KEGG, GO, WIKI and PC), and identified 725 DNA repair genes. Based on the genomic data from CCLE and TCGA, we detected four candidates related to DNA repair (*BRCA1*, *ERBB2*, *G6PC* and *PSME3*) from 89 genes mapped to the copy number variation (CNV) restricted in breast cancer. Further interpretation revealed that the CNV of 17q21.31 could affect the DNA repair (such as “double strand breaks repair” and “DNA damage related detection point”) by changing the expression of *BRCA1* and *PSME3* in breast cancer.

Keywords breast cancer; copy number variation; DNA repair

乳腺癌是女性常见的恶性肿瘤之一, 最新流行病学调查表明, 在中国乳腺癌的发病率已跃居女性恶性肿瘤的第一位。乳腺癌病因复杂, 是遗传因素和环境因素协同作用的结果。前期基于家族性乳腺癌的研究发现, 相当比例的患者存在DNA修复相关基因[*BRCA1*(breast cancer 1, early onset)、*BRCA2*、*TP53*(tumor protein p53)等]的失功能性(loss

of function, LOF)的突变, 提示DNA修复失衡在其发生中具有不容忽视的作用^[1]。DNA修复相关基因主要参与修复内外环境因素所致的DNA损伤。这类基因的突变可直接造成DNA修复能力的缺陷或低下, 致使体内DNA损伤不断积累, 最终增加癌变风险。近年来, 大量针对非家族性乳腺癌组学筛查发现, 虽然多数肿瘤样品中存在已知DNA修复相关

收稿日期: 2014-12-26 接受日期: 2015-02-26

江西省自然科学基金(批准号: 20114BAB205035、20133BCB23007)、江西省教育厅科研项目(批准号: GJJ12110、GJJ13091)和南昌大学大学生创新训练项目(批准号: 2012185、201410403049)资助的课题

*通讯作者。Tel: 0791-86361579, E-mail: quqinlu@ncu.edu.cn

Received: December 26, 2014 Accepted: February 26, 2015

This work was supported by the Natural Science Foundation of Jiangxi Province (Grant No.20114BAB205035, 20133BCB23007), the Research Project of Jiangxi Province Education Department (Grant No.GJJ12110, GJJ13091) and Student Innovation Training Project of Nanchang University (Grant No.2012185, 201410403049)

*Corresponding author. Tel: +86-791-86361579, E-mail: quqinlu@ncu.edu.cn

网络出版时间: 2015-05-04 16:55 URL: <http://www.cnki.net/kcms/detail/31.2035.Q.20150504.1655.001.html>

基因(如*BRCA1*)的表达量改变,但并未发现已知的LOF突变^[2]。这些结果提示,筛查新的DNA修复相关基因和(或)已知基因新的功能性改变是理解非家族性乳腺癌发病机制过程中亟需解决的问题。

拷贝数变异(copy number variation, CNV)是人类基因组中重要的一类遗传变异,可直接通过剂量效应改变特定基因表达量,引起功能紊乱。Stranger等^[3]发现,拷贝数变异与人类基因组中17.7%的基因表达活性相关,且大多数呈正相关,即拷贝数增加使得覆盖区域或临近区域的基因表达活性增强。Shlien等^[4]的研究也证实拷贝数变异可通过改变癌基因或抑癌基因的基因剂量,间接影响个体对肿瘤的易感性。在乳腺癌方面,Ciriello等^[5]在2014年发表的基因组学研究提示,拷贝数变异是非家族性乳腺癌基因组的重要标志,但拷贝数变异对DNA修复过程的影响还尚待深入研究。因此,针对DNA修复相关基因的拷贝数变异筛查和后续功能分析,将有助于阐明非家族性乳腺癌的发生机制,促进乳腺癌的预防、诊断及治疗新策略的提出。

本研究基于CCLE和TCGA两个大型肿瘤基因组学资源,从拷贝数变异的角度探讨DNA损伤修复在乳腺癌中的作用机制。研究发现,17q21.31区域存在乳腺癌特异性拷贝数变异;在乳腺癌临床样本中,该变异所覆盖的*BRCA1*和*PSME3*基因可通过表达量改变影响DNA损伤修复过程。

1 材料与方法

1.1 肿瘤基因组数据

1.1.1 CCLE(Broad-Novartis Cancer Line Encyclopedia)肿瘤细胞系基因组数据 CCLE数据库(<http://www.broadinstitute.org/ccle>)^[6]是一个记录了人类肿瘤细胞系详细信息的大型数据库,其中记录分析了超过1 000个细胞系相关DNA拷贝数、mRNA表达、突变等方面相关的信息。我们从中提取出细胞系的DNA拷贝数信息,包括乳腺癌细胞系59个,非乳腺癌细胞系1 037个。

1.1.2 临床乳腺癌基因组数据 乳腺癌临床样本的基因组数据从cBioPortal下载(<http://www.cbioportal.org/>)^[7]。选取的数据集为Breast Invasive Carcinoma(TCGA, Provisional)。截至2014年7月,该数据集包含了1 069例乳腺癌临床样本。

1.2 全基因组拷贝数变异预测

由于本研究中利用到的乳腺癌临床样本拷贝数

变异片段的计算方法是GISTIC2.0。所以为了保证我们结果的一致性,在利用CCLE细胞系数据计算细胞系拷贝数变异片段的时候我们采用的也是该方法。

根据样本基因组杂交信号及相应版本(hg19, 2009年2月)基因组注释数据(<https://tcga-data.nci.nih.gov/docs/integration/adfs/vendor/>),采用在线版GISTIC2.0分析确定乳腺癌基因组中拷贝数变异的区域及其所覆盖的已知基因。参数设置均为默认值。

1.3 基因表达差异分析

采用MEV(MeV_4_6_2版本)中的t Tests方案分析特定类别的乳腺癌之间的差异表达基因,参数设置均为默认值,差异基因筛查P值小于0.01。

1.4 基因富集分析

本研究采用GSEA(GseaPreranked)分析框架,评价不同功能通路所包含基因在不同类别乳腺癌之间差异表达的整体富集情况,参数设置均为默认筛查,GSEA-P<0.01为显著阈值,表示该功能在相应的差异表达背景下也显著。

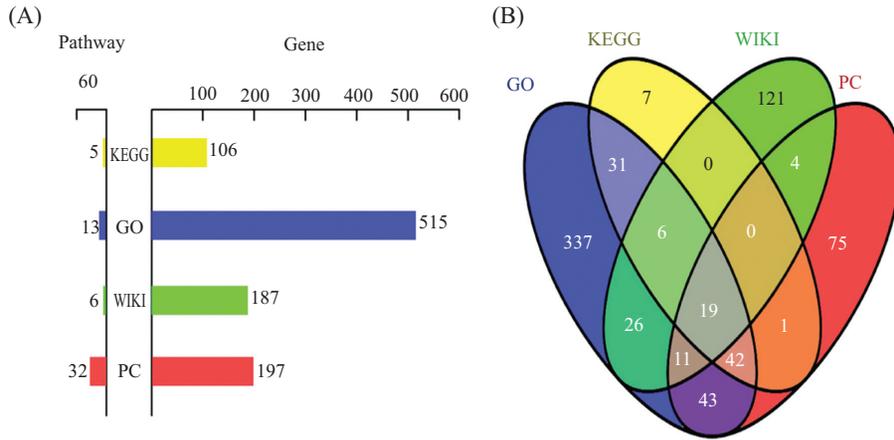
2 结果

2.1 DNA修复相关基因扩展

为了探讨DNA修复失衡在乳腺癌发生中的作用,我们首先从GO^[7]、KEGG^[8]、WIKI^[9]和PC^[10]数据库中确定了56个DNA修复相关通路,建立了包含725个蛋白质的DNA修复相关基因集合(图1A)。其中,KEGG数据库包含了DNA修复机制相关的5个通路,碱基切除修复(base excision repair, BER)、核苷酸切除修复(nucleotide excision repair, NER)、错配修复(mismatch repair, MMR)、同源重组修复(homologous recombination repair, HRR)和非同源末端连接修复(non-homologous end joining, NHEJ),共106个基因。如图1B所示,KEGG所提供的基因与另三个数据库的DNA修复相关基因有较大比例的共享;而其他三个数据库所收录的DNA修复基因在功能各有侧重。与其他数据库相比,PC数据库包含了75个特有基因(38.1%),除了与DNA修复机制相关的基因外,还收录了与DNA修复过程相关的细胞周期检测点通路。而GO和WIKI数据库则分别包含了337个(65.4%)和121个(64.7%)特有基因,功能上偏重于DNA损伤的应对过程。

2.2 乳腺癌特异的拷贝数变异基因筛查

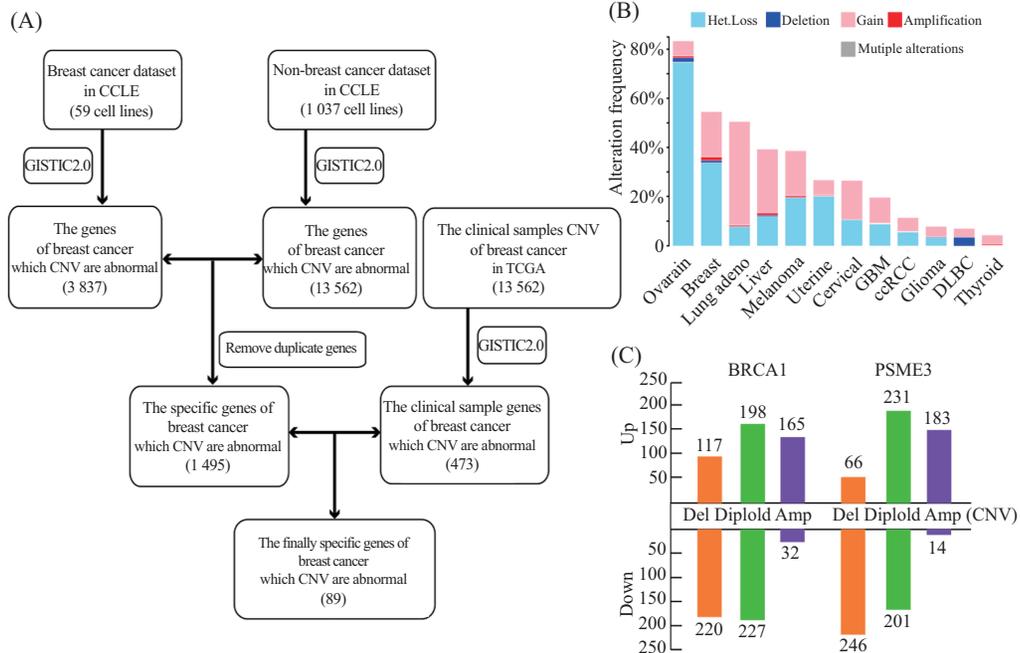
为了发现乳腺癌特异的拷贝数变异基因,本研



A: 本研究整合了4个数据库, 56个通路, 共725个DNA修复基因; B: 4个数据库所收录的DNA修复相关基因的共享情况。

A: in this study, 56 pathways related to DNA repair from 4 databases were integrated, and 725 DNA repair genes were identified; B: the common genes from 4 databases for DNA repair.

图1 DNA修复相关通路整合
Fig.1 Integration of DNA repair related pathways



A: 乳腺癌特异拷贝数变异基因的筛查流程; B: 拷贝数变异区域17q21.31在不同肿瘤临床样本中的拷贝数情况; C: 17q21.31区域中, 拷贝数变异与BRCA1、PSME3基因表达的关系。GBM: 多形性成胶质细胞瘤; DLBC: 淋巴瘤弥漫性大B细胞淋巴瘤。

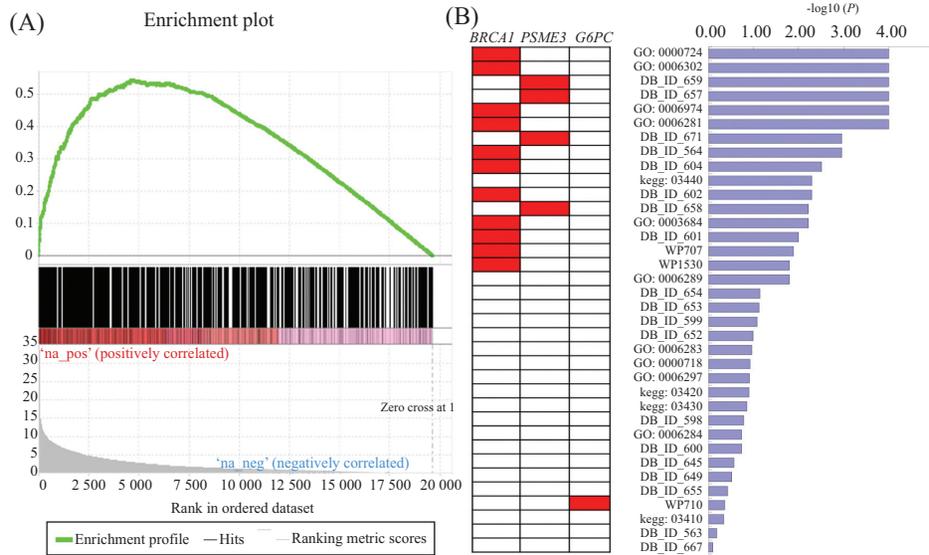
A: the screening process of breast cancer specific genes which CNV are abnormal; B: the CNV of 17q21.31 in different cancers; C: the relationship between the CNV and mRNA expression for BRCA1 and PSME3 from 17q21.31. GBM: glioblastoma multiforme; DLBC: lymphoid neoplasm diffuse large B-cell lymphoma.

图2 乳腺癌特异的拷贝数变异基因筛查
Fig.2 The screening of specific CNV genes in breast cancer

究从CCLE数据库中分别下载了59个乳腺癌和1 037个非乳腺癌细胞系的基因组数据, 并利用GISTIC2.0算法筛选两大类肿瘤细胞系中存在拷贝数变异基因。如图2A所示, 在乳腺癌细胞系中共发现3 837个存在拷贝数变异的基因; 其中1 495个基因为乳腺癌独有, 在非乳腺癌细胞系中未发现明显拷贝数变化。

同时, 根据TCGA所提供的1 044个乳腺癌基因组数据, 本研究筛选到473个在临床样本中存在拷贝数变异的基因, 其中富集了89个肿瘤细胞系所提示的乳腺癌特异性拷贝数变异的基因。

将89个候选基因与已整合的725个DNA修复相关基因进行比较, 本研究最终确定了4个与DNA



A: DNA修复相关基因在H、L组之间呈现差异表达; B: 37个DNA修复相关基因集合的基因富集分析(GSEA), 其中与BRCA1和PSME3基因有关的基因集合呈明显的差异表达模式。

A: the DNA repair genes tend to present the differential expression between groups H/L. The DNA repair genes tend to present the differential expression between groups H/L; B: the GSEA analysis results of 37 effective collections of DNA repair genes. The gene set which was associated with BRCA1 and PSME3 present more obvious difference in expression patterns.

图3 拷贝数变异对DNA修复过程的影响

Fig.3 Effects of CNV on DNA repair process

修复过程相关的拷贝数变异基因(BRCA1、ERBB2、G6PC、PSME3), 其中3个(BRCA1、G6PC、PSME3)聚集分布于基因组17q21.31区域。如图2B所示, 虽然在TCGA数据库所提供的多种临床肿瘤样本中均能发现该区域的拷贝数变异, 但乳腺癌的变异比例达到了54.7%, 远高于其他几种肿瘤, 在所分析的13类肿瘤中排名第二。这一结果印证了肿瘤细胞系拷贝数分析所提示的乳腺癌特异性。

为了讨论拷贝数变异对DNA修复相关基因的功能影响, 本研究进一步整合了相应肿瘤样本的基因表达数据。如图2C所示, 17q21.31区域中拷贝数变异显著改变了BRCA1与PSME3基因在相应肿瘤样本中的表达量。在拷贝数倍增的样本中, 83.8%的肿瘤组织BRCA1基因高表达(即高于所有临床样本表达量的中位数), 这一数值显著高于拷贝数正常或拷贝数降低的临床样本比例(46.6%和34.7%)。而17q21.31区的拷贝数变异亦对PSME3基因的表达具有相似的影响, 当拷贝数下降时, PSME3的表达水平也呈明显下降水平。这些结果提示, 17q21.31区的拷贝数变异可通过BRCA1与PSME3基因表达量的变化影响临床肿瘤的DNA修复功能。

2.3 拷贝数变异致DNA修复失衡的评价

为了评价拷贝数变异是否通过其所覆盖基

因的表达量变化影响DNA修复过程, 本研究根据17q21.31区域拷贝数变异及BRCA1、PSME3基因表达量, 分别选取了229例拷贝数增加且两个基因均高表达的样本组成H组, 175例拷贝数减少且两个基因均低表达的样本组成L组。通过t Tests, 确定了6792个组间差异表达基因($P < 0.01$), 其中富集了318个(318/725, 43.9%)DNA修复相关基因。这一结果提示, 该区域的拷贝数变异通过靶基因(BRCA1、PSME3)的表达量变化影响了组织中DNA修复过程。

为了进一步揭示拷贝数变异对DNA修复过程的影响, 本研究根据GSEA方案的原理, 以所有基因在H、L组之间的差异表达P值分布为背景, 针对DNA修复相关的基因集合进行富集分析。如图3A所示, 比较所获得的ES(enrichment score)为0.5432, 模拟检验(Permutation test)提示, 这一数值大于模拟数据的ES数值(GSEA- $P < 0.001$)。结果显示, DNA修复相关的基因倾向于在H、L组之间呈现差异表达(即富集于 $|\log_{10}(P)|$ 大的区域)。本研究进一步利用GSEA方案评价了所有56个DNA修复通路所涉及基因在H、L组间的表达差异的富集情况。如图3B所示, 在37个有效基因集合中(所含基因数大于15), 17个DNA修复相关通路受到了17q21.31区域拷贝数变异的影响(GSEA- $P < 0.05$); 其中, 11个通路与BRCA1的

功能有关,主要参与了DNA双链断裂(DNA double-strand break, DSB)修复过程,而*PSME3*相关的4个DNA损伤相关检测点通路亦明显受到了该拷贝数变异的影响。

3 讨论

DNA修复系统能够有效修复高致命的遗传变异,对维持细胞基因组的完整性和稳定性具有至关重要的作用。在细胞中,这类基因的单碱基突变、拷贝数变异或其他遗传改变均可改变DNA修复平衡,触发细胞癌变的潜能,机体发生癌变的风险随之增加^[11]。本研究利用生物信息学手段,首先扩展了DNA修复相关基因集合,进一步从拷贝数变异的角度,筛选了在乳腺癌中具有特异性结构变化DNA修复相关的基因,并对其后续的功能影响进行了预测。在研究中我们发现,*BRCA1*和*PSME3*属于乳腺癌特异性拷贝数变异基因,并在乳腺癌临床样本中对DNA损伤修复具有重要影响。

*BRCA1*作为DNA修复的重要调控因子,参与细胞的许多重要的生物学过程^[12],如DNA损伤修复、细胞周期调控、基因调控及染色体的重塑等。其中,以参与修复DSB(double-strand breaks)最为重要。研究指出,有50%~80%的遗传性乳腺癌与*BRCA1*基因的单碱基突变有关^[2]。而在非遗传性乳腺癌患者中,虽然有30%的患者伴有*BRCA1*表达水平降低,但只有5%的患者监测到*BRCA1*基因的单碱基突变^[13],提示*BRCA1*基因可能发生了其他类型的功能性改变。而本次筛查证实,相当比例的非遗传性乳腺癌患者中存在*BRCA1*基因拷贝数的变异,且该拷贝数变异可通过改变*BRCA1*基因的表达影响DSB修复过程。

DSB是DNA损伤最严重的形式,可以在DNA复制过程中自发地产生,也可在外部因素作用下产生,如接触电离辐射及应用基因毒性制剂等^[12]。这种严重的DNA损伤可导致细胞死亡或导致肿瘤发生。而在肿瘤细胞中,*BRCA1*基因的表达异常可导致DSB修复失衡,DSB损伤得不到有效的修复,增加了基因组不稳定性(即:基因组中累积基因缺失、易位、杂合子缺失、染色体缺失或染色体融合)。这一推论与Ciriello等^[5]的结果一致,提示拷贝数变异是非家族性乳腺癌基因组的重要标志。

同时,本次筛查还发现,*BRCA1*所在的17q21.31拷贝数变异区域中还存在另两个DNA修复相关基

因(*G6PC*和*PSME3*)。生物信息学分析进一步显示,该拷贝数变异也改变了*PSME3*的表达,而*PSME3*相关的4个DNA损伤相关检测点通路亦明显受到了该拷贝数变异的影响。但由于尚缺乏*PSME3*与乳腺癌之间的研究,该基因在非家族性乳腺癌发生发展中的作用以及*PSME3*同*BRCA1*之间是否存在协同作用还需进一步的研究。

参考文献 (References)

- 1 Vargas AC, Reis-Filho JS, Lakhani SR. Phenotype-genotype correlation in familial breast cancer. *J Mammary Gland Biol Neoplasia* 2011; 16(1): 27-40.
- 2 James CR, Quinn JE, Mullan PB, Johnston PG, Harkin DP. *BRCA1*, a potential predictive biomarker in the treatment of breast cancer. *Oncologist* 2007; 12(2): 142-50.
- 3 Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, *et al.* Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 2007; 315(5813): 848-53.
- 4 Shlien A, Tabori U, Marshall CR, Pienkowska M, Feuk L, Novokmet A, *et al.* Excessive genomic DNA copy number variation in the Li-Fraumeni cancer predisposition syndrome. *Proc Natl Acad Sci USA* 2008; 105(32): 11264-9.
- 5 Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C. Emerging landscape of oncogenic signatures across human cancers. *Nat Genet* 2013; 45(10): 1127-33.
- 6 Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, *et al.* The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 2012; 483(7391): 603-7.
- 7 Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 2013; 6(269): p11.
- 8 Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: Back to metabolism in KEGG. *Nucleic Acids Res* 2014; 42(Database issue): D199-205.
- 9 Kelder T, Pico AR, Hanspers K, van Iersel MP, Evelo C, Conklin BR. Mining biological pathways using WikiPathways web services. *PLoS One* 2009; 4(7): e6447.
- 10 Cerami EG, Gross BE, Demir E, Rodchenkov I, Babur O, Anwar N, *et al.* Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res* 2011; 39(Database issue): D685-90.
- 11 Liu C, Srihari S, Cao KA, Chenevix-Trench G, Simpson PT, Ragan MA, *et al.* A fine-scale dissection of the DNA double-strand break repair machinery and its implications for breast cancer therapy. *Nucleic Acids Res* 2014; 42(10): 6106-27.
- 12 Moynahan ME, Jasin M. Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. *Nat Rev Mol Cell Biol* 2010; 11(3): 196-207.
- 13 Butcher DT, Rodenhiser DI. Epigenetic inactivation of *BRCA1* is associated with aberrant expression of CTCF and DNA methyltransferase (DNMT3B) in some sporadic breast tumours. *Eur J Cancer* 2007; 43(1): 210-9.