Bolton *et al. BMC Res Notes (2016) 9:287*
DOI 10.1186/s13104-016-2086-3

BMC Research Notes

SHORT REPORT

Open Access

A novel polymorphic repeat in the upstream regulatory region of the estrogen-induced gene *EIG121* is not associated with the risk of developing breast or endometrial cancer

Katherine A. Bolton[1,2], Elizabeth G. Holliday[3,4], John Attia[3,4], Nikola A. Bowden[1,2], Kelly A. Avery-Kiejda[1,2] and Rodney J. Scott[1,2,5,6*]

Abstract

Background: The estrogen-induced gene 121 (*EIG121*) has been associated with breast and endometrial cancers, but its mechanism of action remains unknown. In a genome-wide search for tandem repeats, we found that *EIG121* contains a short tandem repeat (STR) in its upstream regulatory region which has the potential to alter gene expression. The presence of this STR has not previously been analysed in relation to breast or endometrial cancer risk.

Results: In this study, the lengths of this STR were determined by PCR, fragment analysis and sequencing using DNA from 223 breast cancer patients, 204 endometrial cancer patients and 220 healthy controls to determine if they were associated with the risk of developing breast or endometrial cancer. We found this repeat to be highly variable with the number of copies of the AG motif ranging from 27 to 72 and having a bimodal distribution. No statistically significant association was identified between the length of this STR and the risk of developing breast or endometrial cancer or age at diagnosis.

Conclusions: The STR in the upstream regulatory region of *EIG121* is highly polymorphic, but is not associated with the risk of developing breast or endometrial cancer in the cohorts analysed here. While this polymorphic STR in the regulatory region of *EIG121* appears to have no impact on the risk of developing breast or endometrial cancer, its association with disease recurrence or overall survival remains to be determined.

Keywords: *EIG121*, *KIAA1324*, Short tandem repeats, STR, Microsatellites, Regulatory region, Breast cancer, Endometrial cancer, Cancer risk

Findings

The recent emphasis on high-throughput assays in the search to find genetic variants responsible for familial cancer risk has failed to account for a significant proportion of cases. Instead, the use of genome-wide association studies and next-generation sequencing has revealed variants that account for a small portion of heritability and has now resulted in the phrase "missing heritability" to explain that which remains unaccounted for [1, 2]. Another form of genetic variation that has been overlooked, largely due to their inability to be analysed on a large scale, is that of variable tandem repeats (TRs) which are common throughout the human genome and highly mutable [2]. TRs may be drivers of phenotypic variation as they are known to be the cause of several neurological disorders and are associated with complex diseases such as diabetes and cancer [3].

*Correspondence: rodney.scott@newcastle.edu.au
[6] Head of the Discipline of Medical Genetics, School of Biomedical Sciences and Pharmacy, Faculty of Health and Medicine, University of Newcastle, University Drive, Callaghan, NSW 2308, Australia
Full list of author information is available at the end of the article

In a recent study we identified short tandem repeats (STRs) in the upstream regulatory region of genes that are candidates for conferring cancer risk [4]. One such STR is a dinucleotide AG repeat upstream of the estrogen-induced gene, *EIG121* (also known as *KIAA1324*). In endometrial cancer cases, *EIG121* is highly induced by estrogen in the endometrium and differentially expressed in endometrial cancer types [5, 6]. Studies suggest that EIG121, a transmembrane protein, has an important cellular function, as it is highly conserved across species and confers survival upon cells that have been starved of nutrients or exposed to cytotoxic chemotherapeutics [7]. Our analysis from publicly-available datasets, using the Oncomine™ Platform (http://www.oncomine.com), shows *EIG121* to be over-expressed in breast cancer compared to other cancer types (Additional file 1: Table S1; [8]) and compared to normal breast tissue (Additional file 2: Table S2; [9, 10]).

Breast and endometrial cancers are estrogen-driven malignancies, and in both diseases, higher expression of estrogen-induced genes is associated with tumours that tend to be low-grade and less aggressive [5, 11] suggesting involvement of these genes in cancer risk and/or development. As *EIG121* has already been associated with estrogen levels and cancer, we analysed the variability of this newly identified STR in a series of breast and endometrial cancer cases and in a healthy control population to determine if there was any association between its length and the risk of developing these estrogen-driven cancers.

This study included 223 breast cancer cases, 204 endometrial cancer cases and 220 healthy controls from whom blood-derived genomic DNA had been collected for previous studies in Newcastle, New South Wales (NSW), Australia [12–14]. Study participant demographics are shown in Table 1. All participants provided written informed consent for the samples to be used for research.

The STR (a dinucleotide AG repeat) situated 518 bp upstream of the transcription start site for *EIG121* was genotyped by polymerase chain reaction (PCR) and fragment analysis using forward (5′-aggctaatccaggaga atctcttg-3′) and reverse (5′-aggctaatccaggagaatctcttg-3′) primers designed to amplify a 232 bp length fragment. PCR was performed with Platinum *Taq* DNA Polymerase High Fidelity (Invitrogen), an annealing temperature of 61 °C and 1.5 mM $MgSO_4$. Fragment analysis was conducted on the ABI3730 DNA Analyzer (Applied Biosystems (AB)) after denaturation in the presence of HiDi Formamide (AB) and GeneScan 600 LIZ Size Standard (AB). The resulting electropherograms were analysed using Peak Scanner v1.0 software (AB). Sanger sequencing [12] on at least 10 % of each sample cohort, using the same primer sequences as described above, confirmed STR lengths. A line of best fit was generated to correct lengths obtained from fragment analysis as described by Pasqualotto and co-workers [15].

Table 1 Demographic characteristics of the participants used in this study

Characteristic	Breast cancer ($n = 223$)	Endometrial cancer ($n = 204$)	Healthy controls ($n = 220$)
Sex	All female	All female	All female
Age (at ascertainment; in years)			
Range	N/A	40–92	67–86
Median		68	73
Mean (SD)		67.9 (9.5)	73.4 (4.6)
Age (at diagnosis; in years)			
Range	22–57	37–86	N/A
Median	41	63.5	
Mean (SD)	39.8 (7.3)	63.2 (9.0)	
BMI (in kg/m^2)[a]			
Range	N/A	16.9–66.6	17.4–47.1
Median		30.0	27.9
Mean (SD)		31.3 (7.8)	28.5 (5.3)
Underweight (BMI < 18.5)		$n = 1$	$n = 1$
Normal (18.5 ≤ BMI < 25)		$n = 37$	$n = 58$
Overweight (25 ≤ BMI < 30)		$n = 56$	$n = 91$
Obese (BMI ≥ 30)		$n = 94$	$n = 70$
Not specified		$n = 16$	$n = 0$

[a] *BMI* body mass index

Statistical analyses were performed using the Stata 11.1 software package (StataCorp LP, College Station, TX, USA) and involved non-parametric Mann–Whitney U tests, Cox proportional hazard regression, Pearson's Chi squared and Fisher's exact tests. The significance levels of all tests were set at p value < 0.05 (two-sided) and corrected for multiple comparisons using the Bonferroni method.

Based on genotyping results, the AG repeat in the upstream regulatory region of *EIG121* was highly variable, and showed a bimodal distribution of lengths with sizes ranging from 27 to 72 copies across all three cohorts (Fig. 1). The mean values for number of copies of the AG

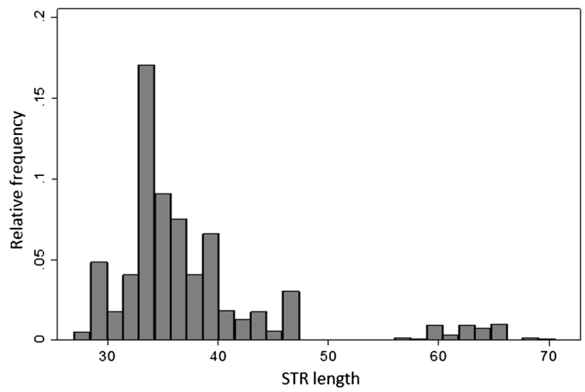

Fig. 1 *Histogram* showing the bimodal distribution of *EIG121* STR lengths across all three cohorts (breast cancer, endometrial cancer and healthy control samples)

motif were 37.14, 38.34 and 37.55 for the breast cancer, endometrial cancer and healthy control cohorts respectively and the median was 35 for all three cohorts.

A non-parametric Mann–Whitney U test was used to test for any association between STR lengths, using both allele lengths for each individual, and cancer types. This demonstrated a lack of association between STR length and breast and endometrial cancers when compared to healthy controls (Table 2). To test if there was any association between STR lengths and age at diagnosis of the cancers, Cox proportional hazard regression was performed. This showed no association between STR length and age at diagnosis for breast and endometrial cancers, including when BMI was taken into account for endometrial cancer (Table 2). BMI data was not available for the breast cancer cohort. When allelic (short (S) vs long (L)) and genotypic (SS, SL and LL) analyses were performed, using a threshold of 50 copies for calling short and long alleles, there were no statistically significant associations for either cancer type (Table 2). In the breast cancer cohort, resulting p values for the allelic and genotypic analyses ($p = 0.185$ and $p = 0.102$, respectively; Table 2) are inclined towards a weak association between STR length and breast cancer risk. Further analysis of larger sample sizes or breast cancer subtypes would be required to confirm any possible association.

In conclusion, the AG dinucleotide repeat in the upstream regulatory region of *EIG121* is a highly polymorphic STR, making it a variable genetic element with the potential to influence the expression of *EIG121* and subsequently impact disease risk and/or severity. No statistically significant association was identified between the length of this dinucleotide repeat and the age at diagnosis or risk of developing breast or endometrial cancer in the cohorts analysed. Hence, while this STR in the regulatory region of *EIG121* is highly polymorphic, it is unlikely to be associated with the risk of developing breast or endometrial cancer. We cannot exclude its involvement in recurrence or overall survival as this information was not available for this study.

Table 2 Hazard ratios (HR), 95 % confidence intervals (CI) and *p* values for breast and endometrial cancer analysis in relation to *EIG121* STR lengths

Category	Statistical test	Breast cancer		Endometrial cancer	
		HR (95 % CI)	*p* value	HR (95 % CI)	*p* value
Both allele lengths with cancer risk	Mann–Whitney U test	N/A	0.985	N/A	0.262
Both allele lengths with age at diagnosis	Cox proportional hazard regression	0.994 (0.981–1.007)	0.343	1.006 (0.995–1.017)	0.304
Both allele lengths with age at diagnosis (BMI considered)	Cox proportional hazard regression	N/A	N/A	1.003 (0.992–1.015)	0.572
Allelic analysis (S/L)	Pearson's Chi squared test	N/A	0.185	N/A	0.393
Genotypic analysis (SS/SL/LL)	Fisher's exact test	N/A	0.102	N/A	0.545

Additional files

Additional file 1: Table S1. Statistics from analysis of the expression of *EIG121* (*KIAA1324*) from Oncomine™ Platform comparisons in multi-cancer datasets (http://www.oncomine.com). The values shown refer to over-expression of *EIG121* in breast cancer cases relative to other cancer types. The total number of samples, *n*, in each dataset are shown.

Additional file 2: Table S2. Statistics from analysis of the expression of *EIG121* (*KIAA1324*) from Oncomine™ Platform comparisons of breast cancer versus normal breast samples (http://www.oncomine.com). The values shown refer to over-expression of *EIG121* in breast cancer cases compared to normal breast. The number (*n*) of breast cancer cases and normal breast samples are shown for each dataset.

Authors' contributions

KAB provided the concept, designed the study, performed the experiments and analysis, and wrote the manuscript. EGH and JA assisted with statistical analysis and manuscript development. KAAK and NAB assisted with the study design, interpretation of results and manuscript development. RJS assisted with the concept and study design, interpretation of results and manuscript development. All authors read and approved the final manuscript.

Author details

[1] Centre for Bioinformatics, Biomarker Discovery and Information-Based Medicine, Hunter Medical Research Institute, Newcastle, NSW, Australia. [2] Priority Research Centre for Cancer, School of Biomedical Sciences and Pharmacy, Faculty of Health and Medicine, University of Newcastle, Newcastle, NSW, Australia. [3] Centre for Clinical Epidemiology and Biostatistics, School of Medicine and Public Health, Faculty of Health and Medicine, University of Newcastle, Newcastle, NSW, Australia. [4] Clinical Research Design, IT and Statistical Support Unit, Hunter Medical Research Institute, Newcastle, NSW, Australia. [5] Molecular Medicine, Pathology North, John Hunter Hospital, Newcastle, NSW, Australia. [6] Head of the Discipline of Medical Genetics, School of Biomedical Sciences and Pharmacy, Faculty of Health and Medicine, University of Newcastle, University Drive, Callaghan, NSW 2308, Australia.

Acknowledgements

This work was supported by an Australian Postgraduate Award (K.A. Bolton) and a National Health and Medical Research Centre Training (post-doctoral) Fellowship and Cancer Institute NSW Career Development Fellowship (N.A. Bowden). We wish to thank the breast cancer patients (Hunter Area Pathology Service, HAPS), endometrial cancer patients (John Hunter Hospital, JHH) and healthy controls (Hunter Community Study, HCS) who kindly consented to the use of their DNA for research. We wish to thank Dr. Michelle Wong-Brown for helping to source the breast cancer samples from HAPS; Dr. Katie Ashton, Dr. Anthony Proietto and Dr. Geoffrey Otton for sourcing and collecting the endometrial cancer samples from JHH; and Mr. Mark McEvoy and Mr. Stephen Hancock for sourcing samples and clinical data for the HCS cohort.

Availability of supporting data

Oncomine™ Platform datasets are available at http://www.oncomine.com. Oncomine™ gene expression analysis results are available as Additional file 1: Table S1 and Additional file 2: Table S2.

Competing interests

The authors declare that they have no competing interests.

Ethics statement

Ethics approval was obtained from the Human Research Ethics Committee, University of Newcastle (H-2009-0265; H-050-065) and the Hunter New England Human Research Ethics Committee, Hunter New England Health Service (09/05/20/5.01; 05/03/09/3.14), Newcastle, NSW, Australia.

Received: 7 February 2016 Accepted: 11 May 2016
Published online: 26 May 2016

References

1. Brookes KJ. The VNTR in complex disorders: the forgotten polymorphisms? A functional way forward? Genomics. 2013;101(5):273–81.
2. Hannan AJ. Tandem repeat polymorphisms: modulators of disease susceptibility and candidates for 'missing heritability'. Trends Genet. 2010;26(2):59–65.
3. Press MO, Carlson KD, Queitsch C. The overdue promise of short tandem repeat variation for heritability. Trends Genet. 2014;30(11):504–12.
4. Bolton KA, Ross JP, Grice DM, Bowden NA, Holliday EG, Avery-Kiejda KA, Scott RJ. STaRRRT: a table of short tandem repeats in regulatory regions of the human genome. BMC Genom. 2013;14(1):795.
5. Westin SN, Broaddus RR, Deng L, McCampbell A, Lu KH, Lacour RA, Milam MR, Urbauer DL, Mueller P, Pickar JH, et al. Molecular clustering of endometrial carcinoma based on estrogen-induced gene expression. Cancer Biol Ther. 2009;8(22):2126–35.
6. Deng L, Broaddus RR, McCampbell A, Shipley GL, Loose DS, Stancel GM, Pickar JH, Davies PJ. Identification of a novel estrogen-regulated gene, EIG121, induced by hormone replacement therapy and differentially expressed in type I and type II endometrial cancer. Clin Cancer Res. 2005;11(23):8258–64.
7. Deng L, Feng J, Broaddus RR. The novel estrogen-induced gene EIG121 regulates autophagy and promotes cell survival under stress. Cell Death Dis. 2010;1(1):e32.
8. Yu K, Ganesan K, Tan LK, Laban M, Wu J, Zhao XD, Li H, Leung CH, Zhu Y, Wei CL, et al. A precisely regulated gene expression cassette potently modulates metastasis and survival in multiple solid cancers. PLoS Genet. 2008;4(7):e1000129.
9. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 2012;486(7403):346–52.
10. Finak G, Bertos N, Pepin F, Sadekova S, Souleimanova M, Zhao H, Chen H, Omeroglu G, Meterissian S, Omeroglu A, et al. Stromal gene expression predicts clinical outcome in breast cancer. Nat Med. 2008;14(5):518–27.
11. Oh DS, Troester MA, Usary J, Hu Z, He X, Fan C, Wu J, Carey LA, Perou CM. Estrogen-regulated genes predict survival in hormone receptor-positive breast cancers. J Clin Oncol. 2006;24(11):1656–64.
12. Wong MW, Nordfors C, Mossman D, Pecenpetelovska G, Avery-Kiejda KA, Talseth-Palmer B, Bowden NA, Scott RJ. BRIP1, PALB2, and RAD51C mutation analysis reveals their relative importance as genetic susceptibility factors for breast cancer. Breast Cancer Res Treat. 2011;127(3):853–9.
13. Ashton KA, Proietto A, Otton G, Symonds I, McEvoy M, Attia J, Gilbert M, Hamann U, Scott RJ. The influence of the Cyclin D1 870 G>A polymorphism as an endometrial cancer risk factor. BMC Cancer. 2008;8:272.
14. McEvoy M, Smith W, D'Este C, Duke J, Peel R, Schofield P, Scott R, Byles J, Henry D, Ewald B, et al. Cohort profile: the hunter community study. Int J Epidemiol. 2010;39(6):1452–63.
15. Pasqualotto AC, Denning DW, Anderson MJ. A cautionary tale: lack of consistency in allele sizes between two laboratories for a published multilocus microsatellite typing system. J Clin Microbiol. 2007;45(2):522–8.