

# Targeted Resequencing of the Coding Sequence of 38 Genes Near Breast Cancer GWAS Loci in a Large Case–Control Study



Brennan Decker<sup>1,2</sup>, Jamie Allen<sup>1</sup>, Craig Luccarini<sup>3</sup>, Karen A. Pooley<sup>1</sup>, Mitul Shah<sup>3</sup>, Manjeet K. Bolla<sup>1</sup>, Qin Wang<sup>1</sup>, Shahana Ahmed<sup>3</sup>, Caroline Baynes<sup>1</sup>, Don M. Conroy<sup>1</sup>, Judith Brown<sup>1</sup>, Robert Luben<sup>1</sup>, Elaine A. Ostrander<sup>2</sup>, Paul D.P. Pharoah<sup>1,3</sup>, Alison M. Dunning<sup>3</sup>, and Douglas F. Easton<sup>1,3</sup>

## Abstract

**Background:** Genes regulated by breast cancer risk alleles identified through genome-wide association studies (GWAS) may harbor rare coding risk alleles.

**Methods:** We sequenced the coding regions for 38 genes within 500 kb of 38 lead GWAS SNPs in 13,538 breast cancer cases and 5,518 controls.

**Results:** Truncating variants in these genes were rare, and were not associated with breast cancer risk. Burden testing of rare missense variants highlighted 5 genes with some suggestion of an association with breast cancer, although none met the multiple testing thresholds: *MKL1*, *FTO*, *NEK10*, *MDM4*,

and *COX11*. Six common alleles in *COX11*, *MAP3K1* (two), and *NEK10* (three) were associated at the  $P < 0.0001$  significance level, but these likely reflect linkage disequilibrium with causal regulatory variants.

**Conclusions:** There was no evidence that rare coding variants in these genes confer substantial breast cancer risks. However, more modest effect sizes could not be ruled out.

**Impact:** We tested the hypothesis that rare variants in 38 genes near breast cancer GWAS loci may mediate risk. These variants do not appear to play a major role in breast cancer heritability.

## Introduction

Genome-wide association studies (GWAS) have identified approximately 180 risk loci (1), all of which are common variants that confer modest disease risks. Fine-mapping and functional analyses suggest that most causal variants modulate risk via regulatory effects, though a few lead SNPs, including in *DCLRE1B* and *EXO1*, are missense substitutions (1). In some diseases, GWAS association signals have been shown to be mediated, at least in part, by rare, high-risk coding variants in nearby genes (2). Moreover, even if GWAS signals are due to regulatory variants, rare coding variants in the target genes are biologically plausible candidates for modulating risk. In this study, we tested this hypothesis by sequencing the coding exons and intron–exon

boundaries of 38 genes that are potential targets for GWAS-identified causal variants.

## Materials and Methods

The subjects, DNA enrichment, sequencing, and variant calling employed in this study have been described elsewhere (3). Sequencing primers, coverage statistics, quality metrics, and variants are in Supplementary Tables S1–S4.

## Results

A total of 3,839 variants were identified, and most were rare, with 3,564 (92.8%) found in <0.1% of all sequenced subjects (noncoding variants) or ExAC European subjects (coding variants; Supplementary Fig. S1; Supplementary Tables S3 and S4). Only 131 truncating variants were identified, and all were uncommon in the population (Supplementary Tables S3 and S4). Burden testing showed that truncating variants were not associated with risk for any gene, even at the nominal significance threshold ( $P < 0.05$ ; Supplementary Table S5).

The aggregate of rare missense variants were not associated for any gene at  $P < 0.0001$ ; however, 5 genes were associated at  $P < 0.05$  (Fig. 1; Table 1). Stratification with SIFT, PolyPhen2, and CADD effect predictions showed that only *NEK10* variants with a CADD score >20 conferred a significantly higher risk than predicted benign variants (OR = 2.73 and 0.86, respectively;  $P_{diff} = 0.010$ ; Supplementary Table S6). This signal was partially driven by variants within the highly conserved *NEK10* protein kinase domain, which were more strongly associated than variants

<sup>1</sup>Department of Public and Primary Care, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, United Kingdom. <sup>2</sup>Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Institute, NIH, Bethesda, Maryland. <sup>3</sup>Department of Oncology, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, United Kingdom.

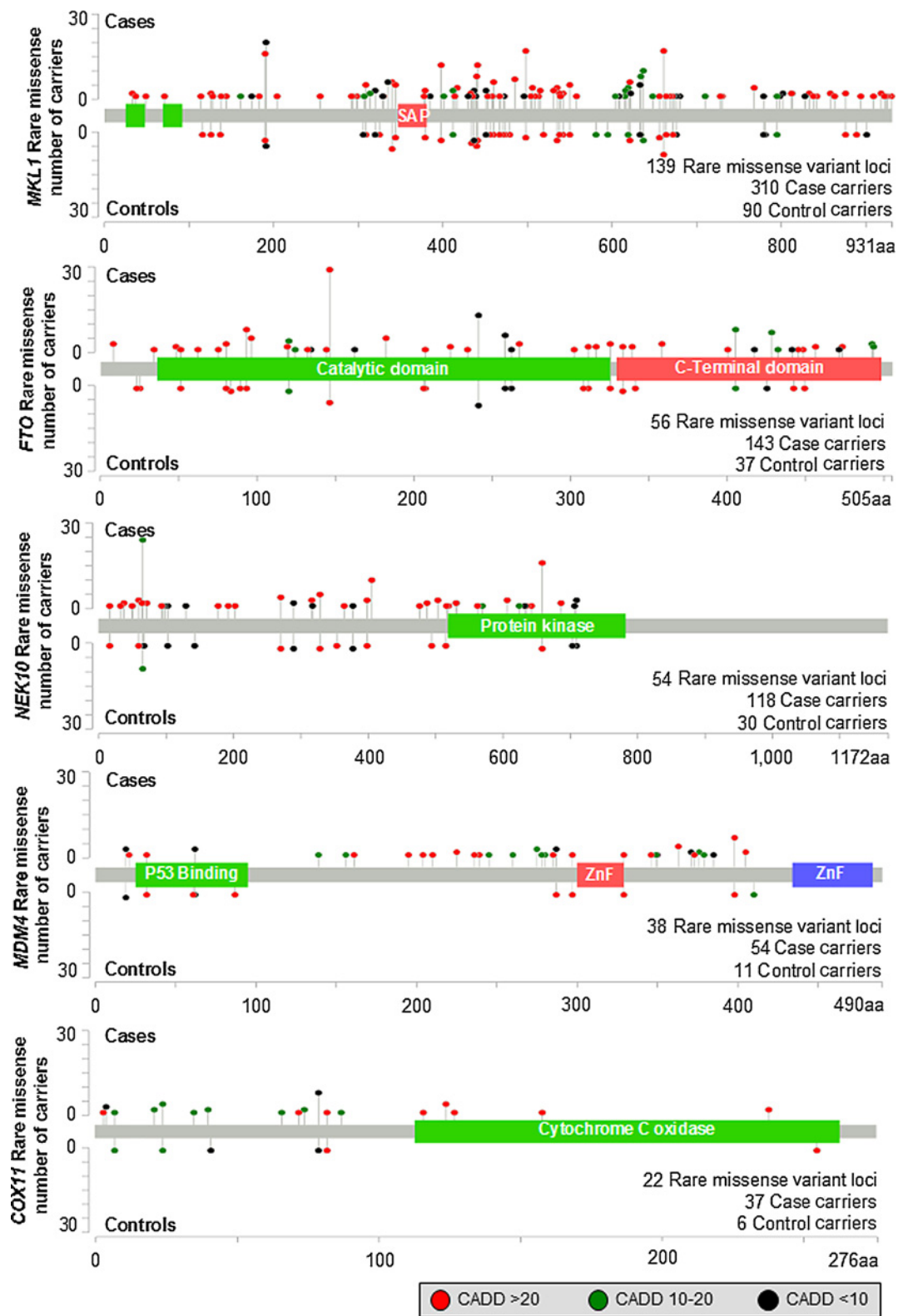
**Note:** Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (<http://cebp.aacrjournals.org/>).

Current address of B. Decker: Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts.

**Corresponding Author:** Douglas F. Easton, University of Cambridge, Strangeways Research Laboratory, Worts Causeway, Cambridge, CB1 8RN, United Kingdom. Phone: +44 1223 748629; Fax: +44 1223 748628; E-mail: dfe20@medschl.cam.ac.uk

doi: 10.1158/1055-9965.EPI-18-0298

©2019 American Association for Cancer Research.



**Figure 1.** Variant position and frequency for rare missense variants in MKL1 (A), FTO (B), NEK10 (C), MDM4 (D), and COX11 (E). Variants are color-coded by CADD prediction. ZnF, Zinc finger domain; SAP, putative DNA/RNA binding domain.

Downloaded from <http://aacrjournals.org/cebp/article-pdf/28/4/822/2286637/822.pdf> by guest on 04 November 2024

**Table 1.** Risk estimates for all rare missense variants in the 38 GWAS genes, as well as subsets of variants that are either localized within or outside of Pfam domain regions

Gene	Overall risk				In Pfam domain				Not In Pfam domain				$P_{diff}$
	Case carriers	Control carriers	OR (95%CI)	P	Case carriers	Control carriers	OR (95%CI)	P	Case carriers	Control carriers	OR (95%CI)	P	
<i>C6orf211</i>	60	21	1.17 (0.70-1.92)	0.63	54	17	1.30 (0.75-2.24)	0.42	6	4	0.61 (0.17-2.17)	0.49	0.28
<i>CASP8</i>	57	32	0.72 (0.47-1.12)	0.18	34	18	0.77 (0.43-1.36)	0.45	23	14	0.67 (0.34-1.30)	0.31	0.93
<i>CCDC170</i>	116	50	0.95 (0.68-1.32)	0.81	0	0	NA	1.00	116	50	0.95 (0.68-1.32)	0.81	NA
<i>CCND1</i>	25	14	0.73 (0.38-1.40)	0.44	20	13	0.63 (0.31-1.26)	0.26	5	1	2.04 (0.24-17.5)	0.68	0.55
<i>CDCA7</i>	38	13	1.19 (0.63-2.24)	0.70	3	1	1.22 (0.13-11.8)	1.00	35	12	1.19 (0.62-2.29)	0.72	1.00
<i>CDKN2A</i>	38	22	0.70 (0.42-1.19)	0.24	78	31	1.03 (0.68-1.56)	0.99	3	2	0.61 (0.10-3.66)	0.63	0.96
<i>CDKN2B</i>	14	4	1.42 (0.47-4.34)	0.61	0	0	NA	1.00	14	4	1.43 (0.47-4.34)	0.61	NA
<i>CDYL2</i>	144	46	1.23 (0.92-1.79)	0.17	69	26	1.08 (0.69-1.70)	0.82	75	20	1.53 (0.93-2.51)	0.11	0.40
<i>COX11</i>	37	6	2.52 (1.06-5.97)	0.045	9	1	3.67 (0.46-29.0)	0.30	28	5	2.29 (0.88-5.92)	0.086	1.00
<i>DCLRE1B</i>	112	44	1.04 (0.73-1.47)	0.91	36	16	0.92 (0.51-1.65)	0.89	76	28	1.11 (0.72-1.71)	0.73	0.75
<i>DNAJC1</i>	89	47	0.77 (0.54-1.10)	0.18	27	15	0.73 (0.39-1.38)	0.43	62	32	0.79 (0.51-1.21)	0.33	1.00
<i>ESR1</i>	78	34	0.93 (0.62-1.40)	0.82	54	25	0.88 (0.55-1.42)	0.69	24	9	1.09 (0.50-2.34)	0.98	0.82
<i>EXO1</i>	186	85	0.89 (0.69-1.15)	0.42	24	13	0.75 (0.38-1.48)	0.52	162	72	0.92 (0.69-1.21)	0.59	0.73
<i>FGFR2</i>	64	25	1.04 (0.66-1.66)	0.95	33	11	1.22 (0.62-2.42)	0.68	31	14	0.90 (0.48-1.70)	0.88	0.69
<i>FOXQ1</i>	20	9	0.91 (0.41-1.99)	0.97	15	5	1.22 (0.44-3.37)	0.89	5	4	0.51 (0.14-1.90)	0.29	0.40
<i>FTO</i>	143	37	1.58 (1.10-2.27)	0.016	139	35	1.63 (1.12-2.36)	0.012	4	2	0.82 (0.15-4.45)	1.00	0.78
<i>HNF4G</i>	46	17	1.10 (0.63-1.93)	0.84	27	7	1.57 (0.68-3.61)	0.37	19	10	0.77 (0.36-1.67)	0.65	0.34
<i>IGFBP5</i>	43	15	1.17 (0.65-2.11)	0.71	22	6	1.50 (0.61-3.69)	0.50	21	9	0.95 (0.44-2.08)	1.00	0.66
<i>MAP3K1</i>	154	50	1.26 (0.91-1.73)	0.18	14	7	0.81 (0.33-2.02)	0.84	140	43	1.33 (0.94-1.88)	0.12	0.47
<i>MDM4</i>	54	11	2.00 (1.05-3.84)	0.045	5	5	0.41 (0.12-1.41)	0.17	49	6	3.34 (1.43-7.79)	5.0E-03	9.0E-03
<i>MKL1</i>	310	90	1.41 (1.12-1.79)	4.8E-03	9	2	1.83 (0.40-8.49)	0.74	301	88	1.40 (1.10-1.78)	6.4E-03	1.00
<i>MRPS30</i>	79	39	0.82 (0.56-1.21)	0.38	79	39	0.82 (0.56-1.21)	0.38	0	1	NA	0.29	0.73
<i>MYC</i>	65	26	1.02 (0.65-1.61)	1.00	65	26	1.02 (0.65-1.61)	1.00	0	0	NA	1.00	NA
<i>NEK10</i>	118	30	1.61 (1.08-2.40)	0.025	32	2	6.53 (1.57-27.3)	1.9E-03	86	28	1.25 (0.82-1.92)	0.35	0.033
<i>NRBF2</i>	32	11	1.19 (0.60-2.35)	0.75	28	10	1.14 (0.55-2.35)	0.86	4	1	1.63 (0.18-14.6)	1.00	1.00
<i>NTN4</i>	104	42	1.01 (0.70-1.45)	1.00	68	26	1.07 (0.68-1.68)	0.87	36	16	0.92 (0.51-1.65)	0.89	0.84
<i>PDE4D</i>	81	40	0.82 (0.56-1.21)	0.37	18	4	1.84 (0.62-5.43)	0.35	63	36	0.71 (0.47-1.07)	0.13	0.13
<i>PTH1R</i>	78	26	1.22 (0.78-1.91)	0.43	40	15	1.09 (0.60-1.97)	0.90	38	11	1.41 (0.72-2.76)	0.40	0.73
<i>PTH1H</i>	13	2	2.65 (0.60-11.9)	0.26	6	0	5.30 (NA)	0.19	7	2	1.43 (0.30-6.87)	1.00	0.49
<i>RAD51B</i>	76	29	1.07 (0.70-1.64)	0.85	44	18	1.00 (0.58-1.73)	1.00	32	11	1.19 (0.60-2.35)	0.75	0.87
<i>RNF15</i>	51	18	1.16 (0.67-1.98)	0.69	8	3	1.09 (0.29-4.10)	1.00	43	15	1.17 (0.65-2.11)	0.71	1.00
<i>TBX3</i>	88	29	1.24 (0.81-1.89)	0.37	46	18	1.04 (0.60-1.80)	0.99	59	27	0.92 (0.58-1.47)	0.83	0.80
<i>TERT</i>	81	42	0.78 (0.54-1.14)	0.24	22	2	4.49 (1.06-19.1)	0.024	59	27	0.89 (0.56-1.41)	0.70	0.034
<i>TET2</i>	255	117	0.89 (0.71-1.11)	0.31	66	37	0.73 (0.48-1.09)	0.15	189	80	0.96 (0.74-1.25)	0.83	0.31
<i>TGFBP2</i>	45	18	1.02 (0.59-1.76)	1.00	28	14	0.81 (0.43-1.55)	0.65	17	4	1.73 (0.58-5.15)	0.47	0.37
<i>TOX3</i>	158	52	1.24 (0.91-1.70)	0.20	3	1	1.22 (0.13-11.8)	1.00	155	51	1.24 (0.90-1.72)	0.21	1.00
<i>ZNF365</i>	84	22	1.56 (0.97-2.50)	0.078	0	0	NA	1.00	84	22	1.56 (0.97-2.50)	0.078	NA

Abbreviation: CI, Confidence interval.

outside of the domain (Table 1;  $P_{diff} = 0.033$ ). In contrast, for *MDM4*, the association was stronger for variants outside Pfam-defined domains (Table 1).

Six common variants were associated with breast cancer risk at  $P < 0.0001$  (Supplementary Table S7), and all were in linkage disequilibrium (LD) with the reported lead GWAS SNP: a 3'-untranslated region variant in *COX11* (rs1802212), two synonymous variants in *MAP3K1* (p.Gln1028Gln, rs3822625; and p.Thr522Thr, rs2229882), and one missense and two synonymous variants in *NEK10* (p.Lys513Ser, rs10510592; p.Thr670Thr, rs11129280; and p.Thr687Thr, rs3213930). In each case, the strength of the associations was compatible with those seen by Michailidou and colleagues (1), but the associations were much weaker than for the corresponding lead SNP (rs2787486, rs62355902, and rs4973769).

Among variants associated at  $P < 0.05$ , *DCLRE1B* p.His49Tyr (rs11552449; OR = 1.10) was also the lead GWAS SNP and conferred a similar risk to that seen in the initial study (OR = 1.07,  $P = 1.8 \times 10^{-8}$ ; ref. 4).

## Discussion

Exon sequencing of genes in GWAS regions did not identify clear novel associations. There was no evidence for association with truncating variants in any genes, and although the variants were too rare to establish reliable estimates, these are unlikely to

be large contributors to breast cancer risk. There was limited evidence of association for rare missense variants in 5 genes, whereas 1.9 genes would have been expected to be associated by chance. Larger targeted studies will be required to establish whether any of these associations can be confirmed; if so, these may indicate novel associations distinct from the common variant associations identified through GWAS.

Six common variants were associated with breast cancer after correcting for multiple testing, but all were in LD with the lead GWAS SNP, and therefore probably do not represent novel risk loci. Moreover, other noncoding SNPs in these regions were more strongly associated, suggesting that these associations are "passenger" associations reflecting LD with causal regulatory variants (5).

## Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

## Authors' Contributions

Conception and design: B. Decker, D.F. Easton

Development of methodology: B. Decker, J. Allen, S. Ahmed, A.M. Dunning, D.F. Easton

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): B. Decker, C. Luccarini, M. Shah, S. Ahmed, R. Luben, P.D.P. Pharoah, A.M. Dunning, D.F. Easton

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): B. Decker, J. Allen, K.A. Pooley, D.F. Easton

**Writing, review, and/or revision of the manuscript:** B. Decker, J. Allen, M.K. Bolla, R. Luben, P.D.P. Pharoah, D.F. Easton

**Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases):** B. Decker, J. Allen, C. Luccarini, M. Shah, M.K. Bolla, Q. Wang, C. Baynes, D.M. Conroy, J. Brown, R. Luben, E.A. Ostrander, D.F. Easton

**Study supervision:** B. Decker, D.F. Easton

### Acknowledgments

B. Decker and E.A. Ostrander were supported by the Intramural Research Program of the National Human Genome Research Institute. SEARCH is funded by a programme grant from Cancer Research UK (C490/A10124) and supported

by the UK NIH Research Biomedical Research Centre at the University of Cambridge (Cambridge, UK). Targeted sequencing in SEARCH was supported by Cancer Research UK grants (C1287/A16563, to D.F. Easton; C8197/A16565, to A.M. Dunning).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received March 27, 2018; revised December 9, 2018; accepted December 14, 2018; published first January 14, 2019.

### References

1. Michailidou K, Lindström S, Dennis J, Beesley J, Hui S, Kar S, et al. Association analysis identifies 65 new breast cancer risk loci. *Nature* 2017;551:92–4.
2. van Leeuwen EM, Sabo A, Bis JC, Huffman JE, Manichaikul A, Smith AV, et al. Meta-analysis of 49 549 individuals imputed with the 1000 Genomes Project reveals an exonic damaging variant in *ANGPTL4* determining fasting TG levels. *J Med Genet* 2016;53:441–9.
3. Decker B, Allen J, Luccarini C, Pooley KA, Shah M, Bolla MK, et al. Rare, protein-truncating variants in *ATM*, *CHEK2* and *PALB2*, but not *XRCC2*, are associated with increased breast cancer risks. *J Med Genet* 2017;54:441–9.
4. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* 2013;45:353–61.
5. Milne RL, Burwinkel B, Michailidou K, Arias-Perez J-I, Zamora MP, Menéndez-Rodríguez P, et al. Common non-synonymous SNPs associated with breast cancer susceptibility: findings from the Breast Cancer Association Consortium. *Hum Mol Genet* 2014;23:6096–111.