Bromodomain protein 4 discriminates tissue-specific super-enhancers containing disease-specific susceptibility loci in prostate and breast cancer

Zuber, V., Bettella, F., Witoelar, A., Andreassen, O. A., Mills, I. G., Urbanucci, A., & PRACTICAL Consortium (2017). Bromodomain protein 4 discriminates tissue-specific super-enhancers containing disease-specific susceptibility loci in prostate and breast cancer. *BMC Genomics, 18*(1), 270. https://doi.org/10.1186/s12864-017-3620-y

Published in:
BMC Genomics

Document Version:
Publisher's PDF, also known as Version of record

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

Publisher rights
Copyright the authors 2017.
This is an open access article published under a Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution and reproduction in any medium, provided the author and source are cited.

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Download date: 09. Jun. 2020
Bromodomain protein 4 discriminates tissue-specific super-enhancers containing disease-specific susceptibility loci in prostate and breast cancer

Verena Zuber, Francesco Bettella, Aree Witoelar, Ole A. Andressen, Ian G. Mills, and Alfonso Urbanucci

Abstract

Background: Epigenetic information can be used to identify clinically relevant genomic variants single nucleotide polymorphisms (SNPs) of functional importance in cancer development. Super-enhancers are cell-specific DNA elements, acting to determine tissue or cell identity and driving tumor progression. Although previous approaches have been tried to explain risk associated with SNPs in regulatory DNA elements, so far epigenetic readers such as bromodomain containing protein 4 (BRD4) and super-enhancers have not been used to annotate SNPs. In prostate cancer (PC), androgen receptor (AR) binding sites to chromatin have been used to inform functional annotations of SNPs.

Results: Here we establish criteria for enhancer mapping which are applicable to other diseases and traits to achieve the optimal tissue-specific enrichment of PC risk SNPs. We used stratified Q-Q plots and Fisher test to assess the differential enrichment of SNPs mapping to specific categories of enhancers. We find that BRD4 is the key discriminant of tissue-specific enhancers, showing that it is more powerful than AR binding information to capture PC specific risk loci, and can be used with similar effect in breast cancer (BC) and applied to other diseases such as schizophrenia.

Conclusions: This is the first study to evaluate the enrichment of epigenetic readers in genome-wide associations studies for SNPs within enhancers, and provides a powerful tool for enriching and prioritizing PC and BC genetic risk loci. Our study represents a proof of principle applicable to other diseases and traits that can be used to redefine molecular mechanisms of human phenotypic variation.

Keywords: BRD4, Genome-wide association studies, SNPs, Functional annotation, Chromatin, Risk loci, Prostate cancer risk, breast cancer risk, schizophrenia, super-enhancer

Background

Genome-wide association studies (GWASs) have linked more than ten thousand of single nucleotide polymorphisms (SNPs) to human diseases and traits [1]. Given that a great part of associated variants are located in known tissue-specific enhancers, a recent study by Tehranchi and colleagues [2] found that these non-coding variants affect transcription factors (TFs) binding and gene expression. Although they found that CCCTC-binding factor (CTCF) is likely to play a pioneering role in translating natural genetic variation in chromosomal architecture [2], we still strive to understand tumor-specific epigenetic features that render possible progression toward such disease. For instance, previous approaches have been adopted to explore disease risk association with regulatory DNA elements [3–6].

In prostate cancer (PC) the androgen receptor (AR) binds predominantly to gene-distal sites and has been used by multiple groups to functionally annotate genetic
risk loci based on overlaps with risk single nucleotide polymorphisms (SNPs) as measured in genome-wide association studies (GWAS), which in some cases are also predicted to affect AR binding [7, 8].

Epigenetic marks such as acetylation on Histone 3 lysine 27 (H3K27ac) have been used as annotation of enhancers [9]. Moreover, regions of extended H3K27ac bound by combinations of mediator complex subunit 1 (MED1) and bromodomain containing protein 4 (BRD4) have been defined as super-enhancers important to determine cell identity [10–12]. BRD4 has proven to be involved in several diseases thanks to the small molecule inhibitor JQ1 [13]. In PC cells, BRD4 was recently shown to bind to the AR and affect its activity [14] while components of the mediator complex such as MED1 and MED12 were recently found to be implicated in advanced PC [15, 16].

SNPs associated with common diseases have been found to lie within enhancers driving transcriptional output and have been identified using different methods [9]. For PC, the most recent methods include genotyping matched to expression quantitative trait loci analysis and epigenetic marks such as H3K27Ac combined with chromatin accessibility [17, 18] or additional combination of binding information for key TFs such as AR and FOXA1 [19]. Here we combined information on H3K27ac profile with binding site data for BRD4 and MED12 to improve the functional annotation of PC risk SNPs based on a previously described enrichment analysis [18].

We show that this method is able to capture SNPs associated not only with PC but also in the context of Breast Cancer (BC) and Lung Cancer (LC) susceptibility. We find that BRD4 is the key discriminant of tissue-specific super-enhancers and binds disease specific PC and BC low p-value risk SNPs. Enrichment of disease-specific risk SNPs is higher when BRD4 binding profile information is incorporated with other epigenetic marks such as H3K27Ac and MED components, than for binding profiles of key TFs implicated in disease development and progression such as the AR or estrogen receptor (ER). Inhibitors for BRD4 are in clinical trials. However, little is known about the contribution of BRD4 to brain diseases. In order to evaluate if similar principles apply also for heritable mental disorders we extended our framework to epigenetic marks including BRD4 binding derived from Schwann cells and applied the enrichment analysis to GWAS studies of mental disorders from the Psychiatric Genetics Consortium (PGC) [20, 21].

Methods

Data source for enhancers' annotation
AR binding information in both LNCaP and VCaP cells was retrieved from Massie et al., (2011) [22]. Raw data were aligned with novoalign to human genome version hg19, and peaks were called with MACS using default parameters after filtering low quality reads (score below 20). Resulting peaks were then overlapped using Bedtools. MED1 binding information and H3K27Ac profile in LNCaP cells was retrieved from Wang et al., (2012) [23] and re-analyzed as described above. To define the degree of overlap with super-enhancers, we also downloaded super-enhancers coordinates from dbSUPER database [24], BRD4 binding information and H3K27Ac profile in VCaP cells was retrieved from Asangani et al., (2014) [14], ER and BRD4 binding information were retrieved from Nagarajan et al., (2014) [25]. H3K27Ac profile in MCF7 was retrieved from Theodorou et al., (2013) [26]. BRD4 and MED1 binding information, and H3K27Ac profile for small cell lung cancer (SCLC) cell line H2171 and Schwann cells were retrieved from cistrome [27]. All cell-specific datasets were equally analyzed to ensure comparability within a tissue type.

Enhancers were defined in LNCaP based on (1) extended H3K27Ac marked regions ranging from 3000 bp to 200 kb (Additional file 1); (2) an intersection of these H3K27Ac marked regions with MED12 binding sites (Additional file 2). In VCaP cells enhancers were defined (3) as an extended H3K27Ac marked regions ranging from 3000 bp to 200 kb (Additional file 3) (4) the intersection of H3K27Ac stretches longer than 2000 bp and BRD4 binding sites (in VCaP cultured in presence of androgens) (Additional file 4) or (5) as BRD4 sites alone (Additional file 5). (6) To achieve a consensus map of super-enhancers in PC (Additional file 6) we selected super-enhancers found in LNCaP cells that were found to have H3K27Ac and BRD4 binding also in VCaP cells.

Enhancers in MCF7 cells were identified following the criteria described in Hnisz et al., (2013) [12]. First, H3K27Ac peaks closer than 100 bp were merged, then only stretches longer than 2000 bp were selected (Additional file 7). Different compendia of enhancers were then created based on the presence of BRD4 (Additional file 8) and ER binding (Additional file 9) or the combination of these features (Additional files 10 and 11). The same type of algorithm was followed to identify enhancers in H2171 and Schwann cells 90-8TL (Additional files 12, 13, 14, 15, 16, 17 and 18). DNase I hypersensitive sites (DHS) profiles for LNCaP cells were retrieved from He et al., (2012) [28] and from ENCODE (Additional files 19 and 20). A more stringent profile of these two based on overlap (Additional file 21) was also included.

Data source for summary statistics of genome-wide association studies
We obtained summary statistics from large meta-analyses of the traits of interest. In particular, the
To assess the significance of the association enrichment, we quantified SNPs enrichment within sets of enhancers. Enrichment is defined by the presence of lower p-values than expected by chance. Quantile-quantile (Q-Q) plots are tools commonly used in genetics to visualize enrichment [18]. Typically, the observed p-value quantiles on the y-axes are plotted against the theoretical p-value quantiles under the assumption of no association (i.e., following the quantiles of the uniform distribution) on the x-axes. In case of no association, a Q-Q plot follows a straight 0–1 line starting from the origin. In the presence of association, the enrichment (of low p-values) is described by the deflection of the Q-Q plot from this theoretical line of no association. We used stratified Q-Q plots to assess differential enrichment of SNPs mapping to specific categories of enhancers. Stratified Q-Q plots have been used previously to demonstrate enrichment of general location annotation categories such as 5′UTR SNPs [18].

**Random pruning**

The statistical models underlying the SNP enrichment analyses carried out here generally assume independence of the data. Far from resembling independent samples, SNPs are linked by complex correlation patterns reflected in their linkage disequilibrium (LD) structure. In order to adhere more closely to the independence assumption, and to rule out bias due to confounding factors such as LD, and assess whether the intrinsic capacity of functional annotations to enrich specific SNP sets was due to such confounding factors, the SNPs were randomly pruned prior to the analyses by randomly selecting representatives from all 1Mbase LD blocks of SNPs with pairwise $r^2 \geq 0.2$. Iterating the random pruning procedure 100 times and subsequently averaging the corresponding test statistics compensated the arbitrariness in the choice of representative SNPs. These analyses were performed and shown in Additional file 22: Figures S1, S3, and S6.

**Results**

To assess whether tissue or cell-specific enhancers could mark tissue-specific risk SNPs associated with development of PC, we analyzed datasets from two studies that profiled MED12 binding and H3K27Ac map in LNCaP cells [23], and BRD4 and H3K27Ac in VCaP cells [14]. MED12, is a subunit of the same chromatin looping mediator complex as MED1 [34] therefore we used it for our PC study assuming that these two subunits would have similar binding profiles in the same cells.

Enhancers were defined in LNCaP based on (1) extended H3K27Ac marked regions (Additional file 1); (2) an intersection of these H3K27Ac marked regions with MED12 binding sites (Additional file 2). In VCaP cells enhancers were defined as (3) extended H3K27Ac marked regions (Additional file 3) (4) the intersection of H3K27Ac stretches and BRD4 binding sites (Additional file 4) or (5) as BRD4 sites alone (Additional file 5). (6) To achieve a consensus map of enhancers in PC (Additional file 6) we intersected the enhancers found in both LNCaP and VCaP cells characterized by all three epigenetic features and responded to the definition of super-enhancers [12] (Table 1 and Fig. 1).

**Enrichment of SNPs associated with prostate cancer in regions bound by MED and BRD4, marked by H3K27Ac in prostate cancer cells**

First, we overlaid genome coordinates of enhancers in PC cell lines, as defined previously, with genomic coordinates of all SNPs in the PC iCOGS dataset [29]. To visualize differential enrichment patterns of specific epigenetic markers with respect to their genetic association with PC risk we generated stratified Q-Q plots which is a method for visualizing the enrichment of statistical
association relative to that expected under the global null hypothesis [18]. Q-Q plots show that SNPs within regions with different genomic features (H3K27ac, BRD4, and MED12, or a combination of these) had different enrichment patterns compared to all SNPs (Fig. 2a). The SNPs contained in common PC enhancers, and therefore characterized by BRD4 and MED12 binding, and a long stretch of H3K27Ac had lower p-values than SNPs contained in enhancers identified in VCaP cells by mapping long stretches of H3K27Ac and BRD4 binding. SNPs associated with PC risk were more enriched within BRD4 binding sites alone than within H3K27Ac sites or H3K27Ac/MED12 overlapping sites in LNCaP. In addition, we focused on SNPs achieving genome wide significance and compared overrepresentation of these SNPs mapping to the above-described enhancers (Additional file 22: Table S1). 12% and 3% of the SNPs contained in PC enhancers achieved genome-wide significance in the iCOGS and in the PRACTICAL GWAS respectively. SNPs that achieved significance in iCOGS were listed in Additional file 22: Table S2. These results highlight that combining generic epigenetic marks such as H3K27Ac with generic epigenetic readers such as BRD4 and with MED binding increases the capacity of capturing SNPs associated with PC.

Importantly, to rule out possible confounding factors, we first randomly pruned the SNPs, selecting one representative SNP per LD block. The random pruning did not change the enrichments patterns caused by the functional annotations (Additional file 22: Figure S1). Secondly, in order to rule out that the enrichment merely results from the non-independence of the SNPs in the enhancer regions or other confounding features of these, we compared the observed enrichment to the one attained on a set of SNPs numerically matching those in the enhancer regions on minor allele frequencies and mutual LD r² (Additional file 22: Figure S2). The numerically matched SNP set was also used as control set to assess the enrichment significance (Additional file 22: Table S1, S3, S4) by means of Fisher’s hypergeometric test (see Methods).

Enrichment of prostate cancer associated SNPs within androgen receptor binding information.

We also compared the genomic coordinates of the SNPs to the coordinates for AR binding sites (ARBSs). Despite the use in the literature of ARBSs for functional annotation of GWAS significant PC SNPs, intersecting enhancer information with AR binding data did not lead to any further enrichment of SNPs associated with PC compared with enhancer information alone (Fig. 2b and Additional Table 1: SNPs mapping to epigenetic marks used to define enhancers in prostate and breast cancer

| Cell line/Tumor type | Epigenetic marks | Key transcription factora | Number of SNPs within the intervals covered by the array: |
|----------------------|-----------------|--------------------------|--------------------------------------------------------|
|                      |                 |                          | iCOGS | PRACTICAL | BCAC | All |
| LNCaP H3K27Ac        |                 | AR                       | 1605  | 3092      | 13482| 13503|
| LNCaP H3K27Ac + MED12|                 | AR                       | 685   | 1271      | 5428 | 5442 |
| VCaP H3K27Ac + BRD4  |                 | AR                       | 587   | 983       | 4218 | 4239 |
| VCaP BRD4            |                 | AR                       | 342   | 502       | 2148 | 2154 |
| VCaP H3K27Ac         |                 | AR                       | 859   | 1595      | 7984 | 8066 |
| PC H3K27Ac + MED12 + BRD4 |            | AR                       | 23    | 60        | 233  | 233  |
| PC H3K27Ac           |                 | AR                       | 3896  | 8440      | 38150| 38444|
| MCF7 H3K27Ac + BRD4  |                 | ER                       | 8783  | 21058     | 93969| 94858|
| MCF7 H3K27Ac         |                 | ER                       | 19270 | 48028     | 215997| 217382|
| MCF7 BRD4            |                 | ER                       | 280   | 495       | 2617 | 2641 |
| MCF7                 |                 | ER                       | 34    | 66        | 282  | 282  |
|                      |                 |                          | 158   | 371       | 1638 | 1639 |

aBinding information for key transcription factors such as androgen receptor (AR) or estrogen receptor (ER) where used alone or in combination with the epigenetic mark profiles in order to assess their capacity to refine enrichment of risk SNPs.

Zuber et al. BMC Genomics (2017) 18:270
In particular, although intersecting AR binding information induced a slight left-shift of the Q-Q plot for enhancers marked by H3K27Ac, MED12, and BRD4 binding, and for enhancers marked by H3K27Ac and BRD4, the enrichment was caused by the same SNPs responsible for the enrichment without AR binding information (see Additional file 22: Table S1 and S3). Furthermore, enhancer information outperformed ARBSs profile alone, or in combination with H3K27Ac profile, in enriching for genome-wide significant p-valued SNPs in PC (Additional file 22: Table S3), and overlapping AR with BRD4 binding sites did not alter the superior capability of BRD4 (as in Fig. 2a) to enrich for disease associated SNPs. Interestingly, although DHSs have been used to predict locations of common disease-associated variation [3], DHSs profiles enriched less than ARBSs alone (Additional file 22: Figure S3).

Validation of the enrichment method on an independent GWAS for prostate cancer.

Finally, we validated our results on the independent PC GWAS obtained from the PRACTICAL consortium measured on a smaller UK-based cohort [30] (Fig. 2c). Again, we observed the strongest SNP enrichment in PC super-enhancers marked by H3K27Ac, MED12, and BRD4 binding.

BRD4 binding sites derived from prostate cancer cells do not enrich for SNPs associated with breast cancer

To test the specificity of BRD4, MED12 and H3K27Ac profiles in PC cells in identifying tissue-specific SNPs, we performed a similar enrichment analysis for genetic association with BC risk measured on the genotype array content from the BCAC [31] (Fig. 2d). Enhancers defined on the basis of BRD4 binding profile in PC cells failed to enrich specifically for BC associated SNPs. Whilst H3K27ac and MED12 together achieved some enrichment of BC SNPs, the addition of BRD4 depleted this enrichment entirely. Importantly, once again, randomly pruning the SNPs did not alter the results of the enrichment analysis (Additional file 22: Figure S4). These results are in stark contrast to the analysis on PC datasets in which inclusion of BRD4 enhanced enrichment of low p-valued SNPs associated with PC, and suggests a hierarchical determination of tissue-specificity, based on the subsequent deposition of these epigenetic marks. Taken together, this indicates that BRD4

![Fig. 1](image-url)
substantially contributes to prostate-specific SNP enrichment within super-enhancers.

Of note, the genomic distribution of the BCAC SNP array mirrored the genomic distribution of the SNP arrays used for iCOGS with the majority of SNPs located within intronic (48% and 57%, respectively) and intergenic (48% and 34%, respectively) regions of the genome (Additional file 22: Figure S5) thus meaning that whilst the number of SNPs differed between the PC and BC studies, there was no genomic distribution bias for imputed SNPs. The SNPs included within the enhancers defined in this study reflected similar distributions, with the only exception of SNPs lists derived from LNCaP cells that were slightly biased toward intergenic regions. Around 69% to 77% of the SNPs were located within intergenic regions (data not shown).

Enrichment of SNPs associated with breast cancer in regions bound by BRD4, marked by H3K27Ac in breast cancer cells

Next, we sought to identify whether using BC-specific epigenetic profiles for the same markers derived from the BC cell line MCF7, we would be able to repeat the same performance as in the PC enrichment analysis. Therefore we retrieved genome-wide profiles of H3K27Ac, ER, and BRD4 binding in MCF7 [25], compiled a similar list of enhancers (Table 1 and Additional file 22: Figure S6), and performed an enrichment analysis of association with BC risk on the BCAC GWAS (Fig. 2e and Additional file 22: Table S4). Information on MED binding is not available for BC cell lines. However, BRD4 binding information alone caused the strongest enrichment of SNPs associated with BC (Additional file 22: Table S5). These data confirm that BRD4 alone is an important enhancer and super-enhancer discriminant, which binds disease-specific susceptibility loci in a tissue specific fashion. Randomly pruning the SNPs involved, proved not to alter the capacity of BRD4 of capturing disease-specific associated SNPs (Additional file 22: Figure S7). Interestingly, pruning the SNPs revealed that ER capability to capture disease associated SNPs in combination with other epigenetic features was enhanced, possibly suggesting a different contribution of ER and AR in breast and PC pathogenesis, respectively.
As counterproof, we tested whether BC epigenetic profiles caused any enrichment in iCOGS PC associations, but no such enrichment was detected (Additional file 22: Figure S8). These results are consistent with BRD4 binding being cell and tissue-specific [35]. Moreover, these results pinpoint the tissue-specificity of risk loci and hint that BRD4 activity may be influenced by genetic variations as it is for TFs [2].

Enrichment of risk SNPs associated with lung cancer and psychiatric traits using H3K27Ac profiles, BRD4, and MED binding sites derived from relevant cell lines

To understand whether the properties of BRD4 binding to clinically relevant genetic risk loci is confined to PC and BC only, or such selectivity can also be observed to other diseases and traits, we retrieved binding information for BRD4, MED1 and H3K27Ac profiles available for the lymphoblastoid cell line H2171 derived from a metastatic site in a LC patient [10] and from the malignant peripheral nerve sheath tumor Schwann cells 90-8TL [36] (Additional files 12, 13, 14, 15, 16, 17 and 18). To retrieve associations of these epigenetic features with other phenotypes we collected summary statistics for LC [32], Alzheimer’s disease [33], schizophrenia [21], and bipolar disorder [20].

BRD4 binding information alone caused the strongest enrichment of associations with LC, although combined information for BRD4 and MED1 binding, also combined with H3K27Ac profile failed to improve the enrichment of low p-value SNPs (Additional file 22: Figure S9a). We speculate that the LC cell line H2171 might not reflect characteristics of the tissue of origin, as well as the PC and BC cell lines. However, upon assessing the enrichment using epigenetic features related to PC cells for the same LC GWAS (Additional file 22: Figure S9b), as expected, we detected none, confirming that BRD4 binding information in H2171 retains some tissue-specificity and capacity to enrich for LC tissue-specific risk SNPs.

Fig. 3 Circular plots of GWAS significant SNPs overlapping with putative super-enhancers. The outmost circles depict chromosome-wise histograms showing p-values of SNP loci (LD r^2 < 0.2 within 1Mbase) representatives for SNPs in iCOGS (a), for SNPs in BCAC (b), and for SNPs associated with schizophrenia according to PGC (c). GWAS-significant SNPs are labeled and the nearby genes are also indicated. The inmost circle represents super-enhancers regions identified in prostate cancer cells (SE_PC_BRD4_MED12_H3K27Ac) (a), breast cancer cells (SE_MCF7_BRD4) (b), and in Schwann cells (SE_Schwann_BRD4) (c) that were most enriched of low p-value SNPs.
Next, we applied our enrichment method to perform an inverse analysis in which we sought to understand whether any association could be found between epigenetic features related to Schwann cells (the only brain cells for which H3K27Ac profile and BRD4 binding information were publicly available) and three diseases affecting the brain. No enrichment for SNPs associated with Alzheimer disease and bipolar disorder was detected (Additional file 22: Figure S10a&b). However, low $p$-valued SNPs associated with schizophrenia were highly enriched within BRD4 binding sites (Fig. 2f). Interestingly H3K27Ac profiles also enriched substantially for clinically relevant SNPs associated with schizophrenia. These data suggest that BRD4 activity in Schwann cells could potentially be involved in the etiology of schizophrenia [37], and grant further investigation on the molecular mechanism underlying these findings.

Discussion

With the discovery of significant numbers of cancer genetic risk loci through GWAS there is now a major focus on the functional annotation of these loci to prioritize them for further biological study. So far this annotation has been undertaken Post-GWAS and has often employed classifiers of open chromatin, for example DHS, as a primary annotation followed by genome-wide binding maps for tissue-specific transcription factors such as the AR for PC or the ER for BC, while combining this information with H3K27Ac and open chromatin in a tissue-specific manner [38]. In this study we ask whether it is possible to use binding sites data and chromatin marks upfront to enrich for genetic risk factors in a cancer type-specific manner. We show that an enhancer signature comprising a number of factors but dominated by BRD4 allows for the enrichment of PC-specific and BC-specific genetic risk loci (Fig. 3a and b). Interestingly, these chromatin features have been previously reported to be characteristic of super-enhancer-like profiles [10–12, 35]. We found a strong degree of tissue-specificity, that is when profiles are derived from cell-lines associated with specific cancer types such as the cancer of the breast and prostate, they become far more effective at enriching for cancer-type specific risk loci than other widely used cancer type-specific TFs such as the AR, ER or DHSs alone. We also applied this enrichment strategy to infer that BRD4 binding information may allow in future for the upfront nomination of genomic-regions for high-coverage sequencing in risk studies for schizophrenia (Fig. 3c). Functional determination of the impact of risk SNPs have been the priority of several consortia aiming to uncover the effects on epigenetics mediated by clinically relevant risk variants located in non-exonic regions [39]. Our study implies a conserved and important relationship between enhancers and cancer-associated risk loci, which is being pinpointed also by recent work linking the effect of genetic variation to TFs binding [2]. Our approach is the first one that implies an effect of such genetic variations on the activity of generic epigenetic

![Fig. 4 Tissue-specific super-enhancers usage and identification of clinically relevant genetic variations associated with diseases and traits. The method for prioritization of clinically relevant SNPs is based on the identification of risk SNPs with GWAS significance that are associated with BRD4 binding to chromatin, within tissue-specific super-enhancers rather transcription factors binding](image-url)
readers. This is also the first time that such epigenetic readers have been evaluated as enrichment factors for SNPs without prior filtering based on published p-values for risk association.

We highlight the possibility that SNPs lying within super-enhancers marked by BRD4 are more likely to be associated with an increased susceptibility to BC, PC, and schizophrenia. The expression of the genes regulated by enhancers identified in these diseases could be altered by the presence of specific SNPs lying therein (Additional file 22: Figure S11). This is a concept that has recently been postulated for cancer mutations occurring in a chromatin-specific context [40].

Conclusions

In conclusion we have discovered that BRD4-bound super-enhancers provide a powerful tool for enriching and prioritizing PC and BC genetic risk loci (Fig. 4), and have shown that key TFs such as AR or ER, despite being pivotal tissue-specific TFs, do not contribute to tissue-specific genetic risk enrichment more than epigenetic factors. We propose to refine disease specific risk loci enrichment with the identification of potential binding of BRD4 combined with key MED components and acetylation profiles. Our study will promote the use of BRD4 for SNP annotation as the genetic landscape for different diseases goes on expanding.

Additional files

Additional file 1: SE_LNCaP_H3K27Ac. (XLS 84 kb)
Additional file 2: SE_LNCaP_MED12_H3K27Ac. (XLS 45 kb)
Additional file 3: SE_VCaP_H3K27Ac. (BED 344 kb)
Additional file 4: SE_VCaP_BRD4_H3K27Ac. (XLS 42 kb)
Additional file 5: SE_VCaP_BRD4_4. (XLS 945 kb)
Additional file 6: SE_PC_BRD4_MED12_H3K27Ac. (XLS 28 kb)
Additional file 7: SE_MCF7_H3K27Ac. (XLS 539 kb)
Additional file 8: SE_MCF7_BRD4_4. (XLS 72 kb)
Additional file 9: SE_MCF7_H3K27Ac_4. (XLS 87 kb)
Additional file 10: SE_MCF7_BRD4_H3K27Ac_4. (XLS 54 kb)
Additional file 11: SE_MCF7_BRD4_H3K27Ac. (XLS 137 kb)
Additional file 12: SE_H2171_BRD4_MED1_H3K27Ac. (BED 53 kb)
Additional file 13: SE_H2171_BRD4. (BED 1857 kb)
Additional file 14: SE_H2171_H3K27Ac. (BED 88 kb)
Additional file 15: SE_H2171_MED1_H3K27Ac. (BED 53 kb)
Additional file 16: SE_Schwann_BRD4_H3K27Ac. (BED 153 kb)
Additional file 17: SE_Schwann_BRD4. (BED 916 kb)
Additional file 18: SE_Schwann_H3K27Ac. (BED 195 kb)
Additional file 19: DHS_consensus. (BED 3694 kb)
Additional file 20: DHS_encode. (BED 6397 kb)
Additional file 21: DHS_Hc. (BED 4673 kb)
Additional file 22: Including Supplementary Material such as Supplementary Figures S1–S11, Supplementary Tables S1–S5, and Supplementary References. (DOCX 2037 kb)
The University of Surrey, Guildford, Surrey, GU2 7XH, UK

COGS acknowledgement and funding: This study would not have been possible without the contributions of the following: Per Hall (COGS); Douglas F. Easton, Paul Pharoah, Kyriaki Michailidou, Manjeet K. Bolla, Qin Wang (BCAC), Andrew Berchuck (OCAC), Rosalind A. Eccles, Douglas F. Easton, Ali Amin Al Olama, Zsofia Kote-Jarai, Sara Benlloch (PRACTICAL), Georgia Chenexv-Trench, Antonis Antoniou, Lesley McGuffog, Ferguson Cusker and Ken Offit (CIMBA), Joe Dennis, Alison M. Dunning, Andrew Lee, and Ed Dicek, Craig Lucarrini and the staff of the Centre for Genetic Epidemiology Laboratory, Javier Benitez, Anna Gonzalez-Neira and the staff of the CNIO genotyping unit, Jacques Simard and Daniel C. Tessier, Francois Bacot, Daniel Vincent, Sylvie Laboissiere and Frederic Robidoux and the staff of the McGill University and Genome Quebec Innovation Centre, Stig E. Bojesen, Sune F. Nielsen, Borge G. Nordestgaard, and the staff of the Copenhagen DNA laboratory, and Julie M. Cuningham, Sharon A. Windebank, Christopher A. Hikker, Jeffrey Meyer and the staff of Mayo Clinic Genotyping Core Facility Funding for the iCOGS infrastructure came from: the European Community’s Seventh Framework Programme under grant agreement n° 223175 (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C1287/A10118, C1287/A10710, C1292/A111174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692, C8107/A16565), the National Institutes of Health (CA128978) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19 CA148065 and 1U19 CA148112 - the GAME-ON initiative), the Department of Defence (W81XWH-10-1-0341), the Canadian Institutes of Health Research (CIHR) for the CIRH Team in Familial Risks of Breast Cancer, Komen Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund.

Breast Cancer Association Consortium (BCAC) (http://bcac.ccgemedschl.cam.ac.uk) BCAC Lead Investigators: Douglas Easton, Ph.D., Paul Pharoah, Ph.D., Georgia Chenexv-Trench, Ph.D., Manjeet M. Hikker.

University of Cambridge, 2 Cambridge Cancer Center, 3 Queensland Institute of Medical Research.

Transdisciplinary Research in Cancer of the Lung (TRICL) Research Team: Hung RJ, Han Y, Brennan P, Bickeboller H, Rosenberger A, Houlston RS, Caporaso N, Landi MT, Heinrich J, Risch A, Wu X, Ye Y, Christians DC, Amos CI.

Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Ontario, Canada, 1 Department of Biomedical Data Science, Geisel School of Medicine, Dartmouth College, Hanover, NH, 2Genetic Epidemiology Group, International Agency for Research on Cancer (IARC), Lyon, France, 3Department of Genetic Epidemiology, University Medical Center, Georg-August-University Göttingen, Göttingen, Germany, 4Division of Genetics and Epidemiology, The Institute of Cancer Research, London, United Kingdom, 5Division of Cancer Epidemiology and Genetics, National Cancer, Institute, National Institutes of Health, Bethesda, MD, USA, 6Helmholtz Centre Munich, German Research Centre for Environmental Health, Institute of Epidemiology, I, Neuherberg, Germany, 7Department of Molecular Biology, University of Salzburg, Salzburg, Austria, 8Department of Epidemiology, UT MD Anderson Cancer Center, Houston, TX, 9Massachusetts General Hospital, Boston, Massachusetts, 10 Department of Environmental Health, Harvard School of Public Health, Boston, Massachusetts.

Funding
A.U. is supported by the South-East Norway Health Authorities (Helse Sor-Ost grant ID 20140406) at the Oslo University Hospital, and the Norwegian Centre for Molecular Medicine. I.G.M. is supported by funding from the Research Council of Norway (RCN), South East Norway Health Authority (SENHA) and the University of Oslo through the Centre for Molecular Medicine (Norway), which is part of the Nordic EMLB (European Molecular Biology Laboratory) partnership and also supported by Oslo University Hospitals. I.G.M. is also supported by the Norwegian Cancer Society and by EU FP7 funding. I.G.M. holds a visiting scientist position with Cancer Research UK through the Cambridge Research Institute and a Senior Visiting Research Fellowship with Cambridge University through the Department of Oncology. A.U. is funded by the SENHA at the Oslo University Hospital. V. Z. is supported by the Centre for Molecular Medicine (Norway) and together with A.W., F.B and O.A.A. supported by the Norwegian Centre of Research in Mental Disorders (NORMENT) with funding from the RCN, SENHA, Norwegian Health Association and KG Jebsen Foundation. This work was supported by the Kristian Gerhard Jebsen Foundation, Centre for Molecular Medicine Norway, Research Council of Norway (213837, 223273), South-East Norway Health Authorities (2013–123), National Institutes of Health (R01AG031224, R01EB007090 and RC2DA249745). I.G.M. and group members participate in the NIH Genetic Associations and Mechanisms in Oncology (GAME-ON), A Network of Consortia for Post-Genome Wide Association (Post-GWA) Research (prostate: 1U19CA148537-01).

This work was also supported by Cancer Research UK Grant CS047/A3354. We would also like to thank the following for funding support: the Institute of Cancer Research and the Everman Campaign, the Prostate Cancer Research Foundation, Prostate Research Campaign UK (now known as Prostate Cancer UK), the National Cancer Research Network UK and the National Cancer Research Institute (NCR) UK. The ProtecT study is ongoing and is funded by the Health Technology Assessment Programme (projects 96/20/06, 96/20/99). The ProtecT trial and its linked ProMPT and CAP (Comparison Arm for ProtecT) studies are supported by Department of Health, UK, Cancer Research UK grant number CS22/A8649, Medical Research Council (UK) grant number G0501066, ID 75466 and the NCR, UK. The epidemiological data for ProtecT were generated through funding from the Southwest National Health Service Research and Development.

Availability of data and materials
The data supporting the results of this research paper are included within this article and its additional supplementary files. Summary statistics from the GWAS studies used in this manuscript are available through application to the relevant consortia (PRACTICAL, BCAC, TRICL and iCOGS).

Authors’ contributions
I.G.M, AU and VZ conceived the study, performed data analysis and wrote the manuscript. I.G.M helped the conceptual design of the study, the preparation of the manuscript, and the interpretation of the results. FB and AW provided scripts for random pruning and edited the manuscript, OAA edited the manuscript. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Consent of publication
Not applicable

Ethics approval and consent to participate
No patient samples were collected and analysed during this study. All GWAS data were provided as summary statistics by the consortia acknowledged in this study having been collected in accordance with ethical regulations in the partner countries and as defined in original research publications by such consortia.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
1 Prostate Cancer Research Group, Centre for Molecular Medicine Norway (NCMM), Nordic EMLB Partnership, Faculty of Medicine, University of Oslo, Oslo, Norway, 2 NORMENT, KG Jebsen Centre for Psychosis Research, Institute of Clinical Medicine, University of Oslo, Oslo, Norway, 3 Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway, 4 European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK, 5 Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, UK, 6 Department of Molecular Oncology, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway, 7 PCUK Movember Centre of Excellence, CCRCB, Queen’s University, Belfast, UK.

Received: 3 March 2016 Accepted: 11 March 2017

Published online: 31 March 2017

References
1. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorf L et al. The NHGRI GWAS Catalog, a curated
