Published in final edited form as:

Mamm Genome. 2015 February ; 26(0): 57–79. doi:10.1007/s00335-014-9551-x.

***In-silico* QTL mapping of postpubertal mammary ductal development in the mouse uncovers potential human breast cancer risk loci**

Darryl L. Hadsell^{1,2,*}, Louise A. Hadsell¹, Walter Olea¹, Monique Rijnkels³, Chad J. Creighton⁴, Ian Smyth^{5,6}, Kieran M. Short⁶, Liza L. Cox⁷, and Timothy C. Cox⁷

¹USDA/ARS Children's Nutrition Research Center, Department of Pediatrics, Houston TX. 77030. U.S.A

²Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston TX. 77030. U.S.A

³Department of Veterinary Integrative Biosciences, College of Veterinary Medicine & Biomedical Sciences, Texas A&M University, College Station TX 77843

⁴Dan L. Duncan Cancer Center, Division of Biostatistics, Department of Medicine, Baylor College of Medicine, Houston, TX. 77030. U.S.A

⁵Department of Anatomy and Developmental Biology, Monash University, Clayton, Victoria 3800, Australia

⁶Department of Biochemistry and Molecular Biology, Monash University, Clayton, Victoria 3800, Australia

⁷Department of Pediatrics, University of Washington School of Medicine, & Center for Developmental Biology and Regenerative Medicine, Seattle Children's Research Institute, Seattle, WA 98101. U.S.A

Abstract

Genetic background plays a dominant role in mammary gland development and breast cancer (BrCa). Despite this, the role of genetic diversity in mammary gland development is only partially understood. This study used strain-dependent variation in an inbred mouse mapping panel, to identify quantitative trait loci (QTL) underlying structural variation in mammary ductal development, and determined if these QTL correlated with genomic intervals conferring breast cancer susceptibility in humans. For about half of the traits, the observed variation among the complete set of strains in this study was greater ($P < 0.05$) than that observed with previously studied strains or with strains that are in current common use for mammary gland biology. Correlations were also detected with previously reported variation in mammary tumor latency and metastasis. *In silico* genome-wide association (GWAS) identified 20 mammary development QTL (*Mdq*). Of these, 5 were syntenic with previously reported human BrCa loci. The most highly significant ($P = 1 \times 10^{-11}$) association of the study was on MMU6 and contained the genes *Plxna4*,

*Direct all correspondence to Darryl Hadsell, Department of Pediatrics, USDA/ARS Children's Nutrition Research Center, 1100 Bates St. Suite 10072, Mail Stop: BCM-320, Houston TX. 77030-2600. Phone, (713) 798-7043. FAX, (713) 798-7057. dhadsell@bcm.edu.

Plxna4os1, and *Chchd3*. On MMU5, a QTL was detected ($p=8\times 10^{-7}$) that was syntenic to a human BrCa locus on h12q24.5 containing the genes *Tbx3* and *Tbx5*. Intersection of high-association SNP ($r^2 > 0.8$) with genomic and epigenomic features, and intersection of candidate genes with gene expression and survival data from human BrCa highlighted several for further study. These results support the conclusion that genetic variation in mammary ductal development is greater than previously appreciated. They also suggest that mammary tumor latency and metastatic index may be influenced by variations in the same factors that control normal mammary ductal development and that further studies of genetically diverse mice can improve our understanding of the connection between breast development and breast cancer in humans by identifying novel susceptibility genes.

INTRODUCTION

Branching processes are central to homeostasis in all multicellular organisms. The purpose of branching is to maximize the ratio surface area to volume for efficient exchange of molecules across a cellular boundary. The mammary gland is one of several well studied ductal systems that use branching mechanisms during development (Ochoa-Espinosa and Affolter 2012). In the mouse, mammary ductal development starts on embryonic day 11.5 with the formation of mammary lines that condense into five pairs of mammary placodes (Propper et al. 2013). These placodes then go on to form rudimentary ductal trees which are each embedded in a stroma termed the mammary fat pad (Propper et al. 2013; Sakakura et al. 2013). Although little development of the ductal tree occurs during the immediate postnatal period, the onset of puberty causes allometric expansion and arborization (Sternlicht 2006). This expansion produces an invasion of the mammary fat pad by primary ducts. The principle driver of this invasion is cellular proliferation within specialized structures known as terminal endbuds (TEB). These TEB not only mediate the invasion of the fat pad but also execute bifurcation, one of the two major branching modes within the gland (Silberstein 2001). Lateral or side branching, which is the second branching mode, occurs through a combination of both cell migration and proliferation at intervals along pre-established ducts (Lain et al. 2013; Richert et al. 2000).

Although many of the players involved in mammary ductal morphogenesis have been identified, comparison with other branching systems suggests that not only are there a large number of regulatory genes left to be discovered, but that these genes need better annotation with regard to their specific functions in mammary ductal development. For example, in *Drosophila* at least 200 different gene products are known to be involved in tracheal branching and can be placed into various functional categories (Ghabrial et al. 2011; Gjorevski and Nelson 2011; McNally and Martin 2011). Likewise, in the kidney, expression of hundreds of differentiation-dependent genes has been identified and mapped to anatomically and functionally distinct regions providing a better functional context (Brunskill et al. 2011). Recent analysis of gene expression in the developing mammary ductal system has led to the identification and compartmentalization of hundreds of differentially expressed genes (Kouros-Mehr and Werb 2006). In addition, a genome-wide “targetome” (cistrome and transcriptome) analysis of PR-dependent gene expression has identified genes that potentially underlie early events in mammary ductal side-branching

(Lain et al. 2013). The main limitation here is that these studies do little to place these gene candidates into the functional context of a normal developing mammary ductal system. To do this requires some way to relate changes in ductal patterning with gain or loss of gene function. A common way to do this in the mouse has been to produce genetically engineered models in inbred strains one gene at a time. This approach has been clearly effective, but an alternative approach that has not been fully appreciated yet is to take advantage of natural genetic variation with the mouse genome as a means to identify genes with functional significance to mammary branching.

The role of natural genetic variation in mammary ductal development and susceptibility to cancer has been documented for many years (Medina 2010). In this regard, early studies on select inbred mouse strains demonstrated a link between the extent of ductal development and later development of mammary tumors (Apolant 1906; Gardner and Strong 1935; Gibson 1930; Haaland 1911). These studies also found that certain strains exhibited altered mammary gland number and placement and displayed estrus-cycle-dependent variation in ductal morphology (Gardner and Strong 1935). This work ultimately led to the conclusion that mammary ductal development was controlled by three factors; 1) the hormonal stimulation to which the mammae were subject, 2) the presence of an agent transmitted through the milk which was later identified as mouse mammary tumor virus (MMTV), and 3) the genetic constitution of the animal (Huseby and Bittner 1946). More recent work with additional strains identified a role for the mammary stoma in determining strain-dependent variation in ductal development (Naylor and Ormandy 2002; Yant 1998), and documented strain-dependent variation in response to diet and exogenous estrogen and progesterone (Aupperlee et al. 2009; Lain et al. 2013; Olson et al. 2010). Despite these, the variation of normal mammary ductal development remains undocumented across a large number of commonly available inbred mouse strains. In addition there have only been 5 genetic loci directly linked to mammary development (Bamshad et al. 1997; Chang et al. 2009; Davenport et al. 2003; Finlay and Marks 1978; Howard et al. 2005; Howard and Gusterson 2000a, b; Megarbane et al. 2008; van Bokhoven et al. 2001; van Genderen et al. 1994; van Steensel et al. 1999), and only one of these was identified in the mouse. Lastly, although there has been a considerable work on the genetics of mammary tumor susceptibility in mice (Hunter 2004, 2012) little has been done to map quantitative trait loci and their genes to variations in normal mammary ductal development.

Recent re-sequencing efforts of inbred mouse strains have allowed for the development of high-density SNP databases for use in genome-wide association studies (GWAS) (Frazer et al. 2007; Kirby et al. 2010). These resources, coupled with the curated collections of mouse strains available through the Jackson Laboratory and other facilities have allow researchers to identify and map gene loci for both Mendelian-inherited and quantitative traits through *in-silico* GWAS (Grupe et al. 2001; Pletcher et al. 2004). This approach has now been successfully used to identify QTL and their underlying genes in a number of situations (Burgess-Herbert et al. 2009; Davis et al. 2013; Ghazalpour et al. 2012; Grupe et al. 2001; Miller et al. 2010; Pletcher et al. 2004; Tang et al. 2009). The primary goals of this study were to describe the degree of phenotypic variation in mammary ductal development across 43 strains within the mouse diversity panel (MDP), to identify QTL associated with

variations in mammary ductal development, and to determine whether any correlate with known BrCA loci in humans.

MATERIAL AND METHODS

Ethics statement

The experiments described in this paper were conducted in accordance with procedures outlined in the *NIH Guide to Care and Use of Experimental Animals*, and they were approved by the Baylor College of Medicine Animal Care and Use Committee. The anesthesia used in these studies was isoflurane. Animals were euthanized by carbon monoxide asphyxiation.

Animals and study design

This study used females (N=4/strain except where otherwise noted) from 43 inbred mouse strains considered a part of the mouse diversity panel (MDP). All animals were purchased from the Jackson Laboratory at 4 weeks of age. The strains studied included C57BL/6J, 129S1/SvImJ, A/J, AKR/J, BALB/cByJ, BPL/1J, BTBR_T+_{tf}/J, BUB/BnJ, C3H/HeJ, C57BLKS/J, C57BR/cdJ, C57L/J, C58/J, CAST/EiJ, CBA/J, CE/J, CZECHII/EiJ, DBA/2J, DDY/JclSidSeyFrkJ, FVB/NJ, I/LnJ, KK/HIJ, LG/J, LP/J, MA/MyJ, MOLF/EiJ, MRL/MpJ, MSM/Ms, NOD/ShiLtJ, NON/ShiLtJ, NZL/LtJ, NZO/HILtJ, NZW/LacJ, PL/J, PWD/PhJ, PWK/PhJ, Qsi5, RIIS/J, SEA/GnJ, SJL/J, SM/J, SWR/J, WSB/EiJ of age. On arrival animals were maintained at an ambient temperature of 21°C, and a light:dark cycle of 14:10 and fed the 2020× pelleted diet (Harlan Teklad, Indianapolis, IN). Biopsies of the #4 mammary gland were collected at 6 and 12 weeks of age. Prior to each biopsy, the study animals were administered a subcutaneous injection of pregnant mare serum gonadotropin (Calbiochem, San Diego, CA) followed 45 to 47 hours later with a subcutaneous injection of human chorionic gonadotropin (Ferring Pharmaceuticals, Parsippany, NJ.). Biopsies were collected at 24 hours following HCG-injection. These treatments are known to induce ovulation and synchronize all of the animals to the estrus phase of the cycle (Hong et al. 2010; Owen et al. 2013).

Processing and imaging of hematoxylin-stained mammary wholemounts

Inguinal (#4) mammary glands were collected and flattened between two microscope slides during fixation. Glands were fixed at least 12 hours in Tellyezniczky's fixative and then stained with iron hematoxylin following a previously described procedure (Rasmussen 2000). Stained wholemounts were imaged in color at 7200 DPI using the Pathscan Slide Scanner (Meyer Instruments, Houston, TX). Color digital images were processed and segmented using Image Pro 7.0 (Media Cybernetics, Rockville, MD). The original image (Figure S1 A) was first subjected to the application of a kernel filter with a 7-pixel feature width to create a background image (Figure S1 B). A background-subtracted image of the ductal tree (Figure S1C) was then created by subtracting the original image from the background image. A manually cleaned, background-subtracted, image was then produced by manually selecting regions of interest that contained non-ductal elements and filling these with the black background present in each image. Each image was then converted to grayscale and then thresholded in order to count and measure the resulting ductal tree with

the count-and-size function in Image-Pro (Figure S1 D). The step also imposed a 500 pixel size filter to remove small non-ductal objects that were missed in manual processing steps. The resulting count mask (Figure S1 E) was then used to generate a skeletonized representation of the ductal tree for further measurements (Figure S1 F). Length and area measurements were converted from pixels to mm or mm² by using a reference calibration image.

Description of ductal development traits

Ductal traits were measured in females at either 6 wk or 12 wk of age. To make the measurement mammary wholemounts were stained and imaged (Figure S2 A). The resulting images were processed to produce binary images of the ductal tree (Figure S2 B). Ductal area (Figure S2 B) was measured in square millimeters by determining the number of pixels that occupied the ductal tree and making the appropriate conversion based on image resolution. Ductal perimeter was measured in millimeters as the length of the line outlining the ductal tree including holes (Figure S2 C). The ductal tree was then digitally eroded to produce a skeleton of single pixel width and the branch points were detected and counted (Figure S2 D). Branch density was determined as the ratio of total branch points to the sum of the lengths of all ductal segments within the skeletonized tree. Hereafter the traits measured as 6 wks of age will be referred to as area_6, perimeter_6, branches_6, length_6 and density_6. Likewise the traits measured at 12 wks of age will be referred to as area_12, perimeter_12, branches_12, length_12 and density_12. Lastly because both the 6 wk and 12 wk data were collected from wholemounts taken from the same animal individual rates of change were calculated for all 5 traits as the difference between the 12wk measurement and the 6wk measurement. These will be referred to as the D_traits: area_D, perimeter_D, length_D, branches_D, and density_D.

Wholemount immunostaining

To provide samples for optical projection tomography (OPT), mammary gland wholemounts were fixed for 6 h at 4°C in 4% paraformaldehyde in phosphate-buffered saline (PBS) at pH 7.4 containing 137 mM NaCl, 2.7mM KCl, 4.3 mM Na₂HPO₄-7H₂O, and 1.4 mM KH₂PO₄. The samples were then washed twice for 1 hr each at room temperature in PBS and once in Tris-buffered saline at pH 7.5 (TBS) containing 100 mM Tris base, and 0.9% NaCl. The wholemounts were then permeabilized in TBS containing 0.1% triton-X 100 (TBST). Following permeabilization the samples incubated for 5 min at room temperature in TBST containing 40 µg/ml proteinase K (Promega Corp., Madison, WI.). The samples were then treated with 0.5 mM phenyl methyl sulfonyl fluoride in TBST and then washed an additional 3 times for 5 min each in TBST. The samples were then blocked at 4°C in TBST with 10% normal goat serum (Thermo Scientific, Waltham, MA.). Following the blocking step samples were incubated overnight at 4°C in a 1:500 dilution of anti-e-cadherin antibody (Invitrogen Life Technologies, Grand Island, NY.). The samples were then washed 5 times for 1 hr each at room temperature and then incubated overnight at 4°C with a 1:1000 dilution of Texas-red-conjugated goat-anti-rat (Life Technologies). The samples were then washed 5× for 1 hr each at room temperature in TBST and post-fixed overnight at 4°C in 4% PFA. After post-fixing the samples were shipped in PBS to Seattle Children's Research

Institute for embedding and optical projection tomography (OPT) scanning at the Small ANimal Tomographic Analysis (SANTA) Facility.

OPT scanning and image processing

Stained mammary wholemounts were embedded in 1.1% low-melting point agarose and dehydrated in 100% methanol for 3 days with the methanol changed each day. Embedded samples were then cleared for three to four days, depending on the size of the specimen, with 1:2 benzyl alcohol:benzyl benzoate (BABB), with the BABB changed each day. Imaging was then performed at standard resolution (512×512) using a Biotronics model 3001M Optical Projection Tomography (OPT) scanner (Skyscan, Belgium). The ultraviolet light imaging mode, using a Texas-red filter, was used to capture fluorescence specifically from the antibody staining. Independent scans varied in their pixel resolution (~31–39 µm) and exposure time depending on the specimen size and intensity of staining, respectively. Scans were routinely carried out using a rotation step of 0.9 degrees, ie. data acquisition through 400 rotational positions. Raw image data (16-bit TIFFs) were reconstructed to multiplanar slice data on a workstation cluster using NRecon (V 1.6.9) software (Skyscan, Belgium) and saved as an 8-bit BMP stack. Histogram ranges were adjusted as needed during reconstruction to improve signal intensity. A smoothing value of 1 was used for all reconstructions. Reconstructed data were then imported into Analyze 10.0 (Mayo Clinic) for digital removal of the mammary node. Segmented data were then rendered for viewing as a 3-dimensional volume using Drishti v2.3 Volume Exploration Software (<http://sf.anu.edu.au/Vislab/drishti>). Images and movies were captured directly from within Drishti. Skeletonization of the ductal trees and generation of phylograms was conducted with Tree Surveyor and Tree Graph, respectively (Short et al. 2013; Stover and Muller 2010).

Data Processing and Analysis

For GWAS, all trait data quality control and analysis were conducted using packages within the R environment (Team 2010). Comparisons of trait variation, along with the generation of histograms and scatterplots, for the data was accomplished using the base R functions. Descriptive statistics were obtained for each of the strains using the doBy package (Hojsgaard 2013). Broad sense heritability estimates were calculated as previously described (Miller et al. 2010) using an excel spreadsheet kindly provided by Dr. Mathew Pletcher (Pfizer Inc. New York, NY). The clustered heatmap of strain-means comparing traits was created using the heatmap.2 function in gplot (Warnes et al.), in conjunction with Mkmisc (Kohl 2013) and RcolorBrewer (Neuwirth 2007).

Genome-Wide Association

GWA was conducted using the EMMA package in R (Kang et al. 2008) in conjunction with a composite SNP dataset that was obtained from the Mouse Phenome Database (<http://phenome.jax.org/>). The data set consisted of 600,831 genotype calls based on the CGD-MDA2, Broad2, and WTCHG1 SNP sets. Coordinates for this dataset are based on GRCm38/mm10. Any SNPs with minor allele frequencies less than 0.05, no-call rate greater than 0.2, and SNP for the Y chromosome were excluded from the analysis. This resulted in a dataset of 437,730 SNP we refer to as the 400k set. All phenotype data were transformed into normal scores using the qqnorm function in base R. Genome-wide significance

thresholds were set by running EMMA on 1000 permutations of the data from each trait (Churchill and Doerge 1994). Manhattan plots were constructed using publicly available code for a function written for ggplot2 (Stephen Turner, <http://gettinggeneticsdone.blogspot.com/>). Regional association plots were generated from the 400K set using an R-function (<https://www.broadinstitute.org/diabetes/scandinavs/figures.html>) obtained from the Broad Institute website. Estimates of local recombination rate were derived from recently published analysis (Brunschwig et al. 2012). Long-range LD was also estimated for significant loci by calculating r^2 .

Identification of potential causal SNPs and genes

Analysis of linkage disequilibrium (LD) in the regions surrounding associations was conducted by calculating r^2 for SNP within 10 MBP on either side of the lead SNP. This was initially done with 400K set, but was then repeated using additional SNP genotype calls obtained by combining data from several additional datasets in the Mouse Phenome Database. This set includes genotype calls from Perlegen2, CGD-MDA1, Broad2, Chicago1, UNS-MUGA1, JAX-SNP1, CGD-IMP2, and CNB1 and will be referred to as the HD set. Only high-confidence calls were used from CGD-IMP2 dataset. Using the HD set, an LD-block, and the genes contained within it, was defined as the distance between the most distant flanking high-LD SNP on either side of the lead SNP. High-LD SNPs were defined as those that had a no call rate of ≤ 0.3 and an r^2 of ≥ 0.8 with the lead SNP for each QTL region. The accompanying functional annotations for these high-LD SNP were also downloaded from the Mouse Phenome Database. To identify the genomic features which overlapped with the *mdq* loci, the lead and high-LD SNP for each QTL, were intersected with ChIP-seq data for histone methylation (GSE25105) (Rijnkels et al. 2013), with STAT5 (GSE48685) (Yamaji et al. 2013) and progesterone receptor (GSE42887) (Lain et al. 2013) binding sites, and with consensus transcription factor binding sites defined by Hypergeometric Optimization of Motif Enrichment (HOMER, <http://homer.salk.edu/homer/>). The position information for these SNP in mouse genome assembly mm9 was retrieved and intersection was performed using the UCSC genome browser table-browser (<http://genome.ucsc.edu/>) or Galaxy (<http://galaxyproject.org/>). The potential effect of 3' UTR SNP on miRNA target sites were analyzed using PolymiRTS Database (<http://compbio.uthsc.edu/miRSNP/>).

RESULTS

Analysis of mammary ductal development traits at 6 and 12 weeks of age

To evaluate the impact of genetic background on mammary ductal development we compared mammary wholemounts from 43 different inbred mouse strains. The collection of the left and right inguinal mammary glands from each animal at 6 and 12 wks of age allowed for an analysis at different developmental stages and facilitated the calculation of developmental rates for each trait based on the difference between the two ages. Because stage of the estrus cycle is known to cause variations in mammary development, this was corrected for by synchronizing the animals with gonadotropin injections prior to collection of the biopsies. Consequently all animals should have been in metestrus at the time of biopsy collection. To accomplish the wholemount analysis, digital images were processed to

produce a binary image that was then measured with the count and size function of ImagePro Plus (Figure S1). A total of 5 different quantitative measurements (Figure S2) were made on the ductal trees that were segmented from these binary images. The analysis of these at both ages along with the calculation of the difference for each of the traits produced a total of 15 trait measurements for each animal.

Correlation Structure among ductal development traits with body weight and sexual maturation

To visualize the correlation structure among the different traits, a lattice plot was constructed showing histograms for each trait on the diagonal, scatterplots for each trait to the lower left diagonal of diagonal, and Pearson correlations to the upper right of diagonal (Figure 1) This visualization demonstrated that within each of the time points all of the traits were moderately to highly correlated. The highest Pearson's r values (0.98) were observed between length_6, and area_6, perimeter_6, and branches_6 (Figure 1 A). Density_6 was also correlated to these traits, but the r values were more moderate. A similar correlation structure was observed among the 12-wk traits (Figure 1B) and the D-traits (Figure 1C), but these were lower than that present at 6 weeks. The D-traits were also correlated to the 12-wk traits, but not the 6-wk traits. In addition for area, perimeter, length, and branches, there were moderately high Pearson's correlations with body weight at both 6 and 12 weeks of age. These were highest within an age group and decreased when compared across ages. This observation supports the suggestion that larger strains had higher ductal development.

Because the rate of ductal development is slow during the early neonatal period and is increased with the onset of puberty, differences in ductal development would have been expected in response to variations in the timing of puberty (Flux 1954). Although we did not control for this in the current study, we were able to determine the degree to which onset of puberty contributed to the variations in mammary ductal development. To determine the relationship between ductal development and onset of puberty, previously published (Yuan et al. 2012) observations on vaginal patency were obtained from the Mouse Phenome Data base (MPD). For 29 of the 43 strains in our survey, measurements of vaginal patency could be obtained from the MPD. Data were unavailable for BPL/1J, C58/J, CER/J, CZECHII/EiJ, DDY/JclSidSeyFrk/J, I/LnJ, LG/J, Ma/MyJ, MSM/Ms, NZL/LtJ, NZO/HiLtJ, PWK/PhJ, Qsi5, and SEA/GnJ. Comparative analysis using strain means revealed that strains with a lighter body weight had delayed ($r = -0.55$, $P < 0.0005$) vaginal patency (Figure 1). In addition, all of the ductal development traits except density_6, density_12, and density_D tended to be higher in strains that underwent vaginal opening earlier. Importantly, all of the D traits had little or no correlation to vaginal patency. This result indicates that although the extent of ductal development in any one biopsy is influenced both by the rate of growth and sexual maturation, the growth rate of the mammary ductal tree, as indicated by the difference between two sequential biopsies from the same individual, is independent of the timing pubertal onset.

Comparison of ductal development at 6 and 12 weeks

To visualize the amount of variation present in mammary ductal development and to determine the effect of age, the distributions for each of the five traits at 6 and 12 weeks of

age were compared (Figure 2). For all of the traits except density, the distributions were skewed. In addition, the average values for all of the 12-wk traits were higher ($p<0.05$) than those for the 6-wk traits. This observation was consistent with the known fact that the mammary ductal system continues to expand between these two ages in the mouse. Importantly, strain accounted for a highly significant ($p<0.0001$) proportion of the observed variation in all of the traits (Table S1). However, not all strains demonstrated increases in all of the ductal traits with age. Estimates of broad-sense heritabilities were generally highest for the 6-week traits and lower for both the 12-week and D-traits. These ranged from a low of 0.32 for density_6 to a high of 0.65 for length_6 (Table S1). This result demonstrates that ductal development traits in the mouse are moderately to highly heritable and emphasizes the importance of genetic background as a determinant of mammary ductal development.

Hierarchical clustering of the Strains by developmental changes

To determine if the strains could be grouped based on temporal patterns of mammary development, a heat map was produced and strains were clustered on the basis of their average normal scores for each trait (Figure 3A). This revealed that strains could be grouped into at least 4 clusters. Cluster 1 strains exhibited the highest overall development at 6 weeks with only modestly high D for the traits. Cluster 2 strains had average to below average development at 6 weeks, but some of the highest development observed in the entire population at 12 weeks. This cluster also contained strains with the highest rates of development between 6 and 12 weeks. Cluster 3 had above average development at 6 weeks, average to below average development at 12 weeks, and average to below average developmental rate. Cluster 3 also included the population average, along with the commonly studied strains FVB/NJ and C57BL/6J. Cluster 4 strains had below average development at 6 weeks, below average development at 12 weeks, and average to below average developmental rates. Alignment of the heatmap with a genetic distance matrix (Figure 3B) demonstrated that with the exception of 4 of the 6 wild-derived strains that were analyzed, there was only limited clustering of the strains by kinship. This was further illustrated by the fact that although there were five C57-related strains in the analysis, they did not cluster together in the heatmap but were distributed across clusters 3 and 4. A more extreme example of this is the comparison between BTBR^{+/tf}/J and 129S1/SvImJ. These two strains, though closely related, were quite different ($P<0.05$) in their degree of mammary ductal development. This result illustrates that variation in the timing of ductal developmental onset, as well as in the rate of ductal development is influenced by genetic differences among even closely related strains.

Comparisons to previously studied strains

Although inbred strain comparisons of mammary gland development have been made by others (Gardner and Strong 1935; Gibson 1930; Huseby and Bittner 1946; Naylor and Ormandy 2002; Olson et al. 2010; Yant 1998), the 17 different strains described in all of these studies combined is still small in comparison to the roughly 224 inbred strains described in the mouse phenome database (<http://phenome.jax.org/>). Of these 17 “classical” strains, 7 were available and included in the current work. In addition, a subset of 5 strains that are in common current use for mammary biology were also included in the current work. Examination of the heatmap (Figure 4) demonstrated that there were 9 additional strains

within the complete set described here which display more extreme differences in the ductal traits than the classical or commonly studied strains. In addition comparison of the variances for the 15 traits among the classical and common strains with that observed for the entire set of strains (Table S2) revealed that for 14 of the 15 traits, these variances were larger in the entire strain set. Comparing to classical strains, these increases were significant ($p < 0.05$) for 4 of the 5 traits at 6 weeks, and 2 of the 5 traits at 12 weeks of age. For the difference traits, the variances were not significantly different. These results support conclusion that in analyzing additional strains, the current work has been able to capture a larger amount of phenotypic variation in mammary ductal development than has previously been described.

Identification of new phenotypic extremes

To identify strains that constituted extreme ends of the distributions for each trait one-sample t-tests were conducted comparing each strain mean (Table S1) to the “population” mean. At 6 weeks, all of the traits except perimeter₆ and density₆ had the same strains at the upper and lower extremes of the distributions. These were NZL/LtJ, KK/HIJ, and DDY/JcLSidSeyFrk/J, and LP/J, PWD/PhJ, and PL/J, respectively. For Perimeter₆, DDY/JcLSidSeyFrk/J, Qsi5, and NZL/LtJ were the top three and KK/HIJ was fourth. For density₆, the upper distribution extremes were PWK/PhJ, BUB/BnJ and NZO/HILtJ, respectively. At the lower end of the density₆ distribution, only DBA/2J was significantly less than the population mean. A third strain, CZECHII/EiJ, is worth mentioning since this strain not only had decreased ($p < 0.05$) perimeter₆ in comparison with the population mean, but also displayed a unique ductal pattern.

Visualization of ductal phenotype extremes

In addition to great variation in these global measurements of ductal development there were strains which displayed ductal patterning that could be considered extreme. Figure 4 shows examples of strains that were representative of each of the four development clusters and also illustrates examples of two extreme strains, KK/HIJ (A and B) and CZECHII/EiJ (O and P). Comparison of wholemount images of these strains with more commonly studied strains such as FVB/NJ (I and J) C57BL/6J (K and L), and BALBc (M and N) revealed striking differences (Figure 4). The most dramatic of these was observed in the CZECHII/EiJ strain. Ductal development for this strain was near the bottom of the distributions for all traits. Beyond this however, the mammary ductal system in these mice displayed pronounced elongation of 2 to 3 primary ducts with little to no bifurcation or side-branching beyond the region of the lymph node (Figure 4 O). By 12 wks of age this unique pattern manifest itself as unusually long primary branches with mostly secondary branching (Figure 4 P). This pattern was even more appreciable in rendered 3D reconstructions (Figure 5 A and D) of the mammary ductal trees obtained by wholemount staining for the epithelial marker, e-cadherin and imaged by optical projection tomography (Video S1). At the upper extreme of the developmental distribution in development cluster 1 were the KK/HIJ (Figure 4 A and B) and BUB/BnJ (Figure 4 C and D) strains. These mice not only had high values for all ductal development traits at 6 wks, but also displayed distinctly high amounts of branching near the ductal tips and higher amounts of TEB bifurcation or TEB which were lobular in appearance. This property was also better appreciated in rendered 3D reconstructions (Figure 5 D and E) of ductal trees from the KK/HIJ females taken at 6 weeks

of age (Video S2). Processing of the reconstructed tomographic data with TreeSurver produced a skeletonized representation that could be subsequently visualized as a phylogram (Figure 5 C and F). These phylograms further illustrated that the two strains exhibited not only global differences in mammary ductal development, but also local patterning differences. The phylogram for the CZECHII/EiJ was less open in early ductal generations as an indication of shorter duct length, while the reverse was true in the KK/HIJ phylogram. In addition, the KK/HIJ phylogram shows much more frequent branching at the ductal tips supporting the conclusion that strain-dependent differences exist in local patterning of the mammary ductal system.

Ductal development is modestly correlated to mammary tumorigenesis

To get an initial indication of the relationship between ductal branching and mammary tumorigenesis, the data collected in this study was combined with previously published data on mammary tumorigenesis and metastasis in F1 progeny resulting from a cross between different inbred mouse strains and the FVB/N-Tg(MMTV-PyVT)634Mul/J strain (Lifsted et al. 1998). Alignment of the strain means from the wholemount data with the means for tumor latency and metastatic index produced a set of coincident data points for 22 strains. Calculation of Pearson's correlations for these strains (Figure 6 A) revealed that several normal development traits including density_6, branches_12, density_12 branches_D and length_D had positive Pearson's correlations with tumor latency or metastatic index in the range of 0.33 ($P < 0.05$) to 0.40 ($P < 0.01$). Closer inspection of the scatter plot for density_12 and metastatic index revealed the appearance of two clusters of strains which were separated by differences in density (Figure 6 B). The R^2 for each of these was 0.73 and 0.76. This result supports the suggestion that tumor latency and metastatic index may be influenced by variations in the same factors that control normal mammary ductal development.

QTL for ductal development traits in the mouse

To identify genomic regions associated with the various ductal traits measured in this study, *in-silico* GWAS was conducted using the 400K SNP dataset. To set a reliable significance threshold EMMA was run on 1000 permutations of the phenotype data for each of the 15 traits. Taking the 5th percentile from the distribution of minimum p-values for each permutation analysis produced a genome-wide threshold for $\alpha = 0.05$ of $1.9 \times 10^{-6} \pm 3.6 \times 10^{-7}$. Based on this threshold, 20 mammary ductal QTL (*Mdq*) containing associated SNP were detected on chromosomes 1, 2, 5, 6, 7, 9, 13, 15, 16, and 18 (Figure 7, Table 1). The most significant ($p = 8.9 \times 10^{-12}$) of these (*Mdq1*) was detected for branches_D on MMU6 (Figure 7 M). This region was also associated with perimeter_D (Figure 7 L) and length_D (Figure 7 N). The second most highly significant ($p = 2.1 \times 10^{-9}$) locus, *Mdq2*, was on MMU1 and was associated with area_12 (Figure 7 F). This QTL was also associated with perimeter_12 (Figure 8 G), branches_12 (Figure 7 H) and length_12 (Figure 7 I), and was close to two additional loci (*Mdq7* and *Mdq20*) associated with the same traits (Table 1). The third most significant (3.2×10^{-9}) locus, *Mdq3*, was on MMU18. This QTL was associated with perimeter_6 (Figure 7 B), area_6 (Figure 7 A), branches_6 (Figure 7 C), and length_6 (Figure 7 D). There were three loci with p-values in the range of 10^{-8} found on MMU2, 13, and 16, respectively (Table 1). These were associated with density_6 (Figure 7 E), branches_D (Figure 7 M), and density_12 (Figure 7 J), respectively. In addition to *Mdq7*

and *Mdq20*, nine QTL had p-values in the range of 10^{-7} and three were in the range of 10^{-6} (Table 1). All significant QTL were tested for long-range LD, which has previously been suggested to be a confounding factor in GWAS (Cervino et al. 2005). Comparison of r^2 among all 20 loci (Table S3) demonstrated that *Mdq2* 7 and 20 displayed suggestive linkage ($r^2 > 0.5$) among themselves and with *Mdq17*, as did *Mdq3* and 15. Despite this, there was no evidence of strong long-range LD ($r^2 > 0.8$) among any of the loci.

Cross-referencing human breast cancer GWAS

To determine if any of the *Mdq* described here were syntenic with previously described human BrCa loci, the UCSC genome browser was used to perform a liftover of the murine coordinates for each locus from mm10 to hg19. Combining the results with a previously published summary of human BrCa risk loci (Ghoussaini et al. 2013) revealed that 5 of the 20 *Mdq* were syntenic with previously detected BrCa risk loci (Figure 7, Table 4). The potential significance of these loci as well as the remaining QTL will be addressed further below.

Defining LD Blocks and Identification of candidate genes

To identify potential candidate genes with the *mdq* loci, additional genotype calls were obtained from 9 MPD databases and included the calls for each of the 20 *mdq* lead SNP. For each of the lead SNP within a locus, both D and r^2 were calculated for surrounding SNPs at up to 10 Mbp on either side. An $r^2 > 0.8$ was used to set the size of the LD block relative to the lead SNP in each QTL (Table 1). Using this criteria a total of 6,650 high-LD SNP (high-LD SNP set) were identified, and 19 of the 20 loci had detected LD blocks. These blocks ranged in size from 43,727 bp for *Mdq12* to just over 4.1 Mbp for *Mdq20*. Intersection of these blocks with genomic data from the Mouse Phenome Database identified 157 potential candidate genes (Table 1). Local LD plots for *Mdq 1* and 2 illustrate the approximate boundaries and the genes contained in them (Figure 8 A & B). For *Mdq3*, although an associated SNP was contained within the last intron of the *Cdh2* gene, there was no strong evidence for an LD block containing other genes (Figure 8 C.). Both *Mdq11* and 15 were found to overlap with previously described human BrCa risk loci, and contain genes that have been found to play a role in normal mammary gland development (Figure 9 A & B, Table 1). In *Mdq11* the genes *Sox6*, *Plekha7*, and *Pik3c2a* have potential relevance to mammary biology or BrCa (Castellana et al. 2012; Kendrick et al. 2008; Kurita et al. 2013). While, *Mdq15* contained 19 genes including *Tbx3*, which plays an important role in embryonic mammary gland development and has also been implicated in human mammary ulnar syndrome (Bamshad et al. 1999; Bamshad et al. 1997). Furthermore, *Mdq19*, was interesting because the LD block for this region contained 12 genes including the *Pgr* gene, which is known to be indispensable for mammary ductal side branching (Mulac-Jericevic et al. 2003) (Figure 9C). Of the remaining QTL, noteworthy candidate genes within these loci include several keratin genes found in *Mdq17*, and the gene *Wnt2* found in *Mdq13* (Table 1). The fact that there was both overlap of these novel mammary ductal QTL with both BrCa risk alleles and genes with a known role in mammary ductal development supports the conclusion that both the measurements described in this manuscript, and the GWAS approach, were successful at discovering loci important to mammary gland development. Follow-up functional studies on these novel loci will be important.

Intersection of Gene Candidates with human breast tumor data

To identify links between the GWAS candidate genes and breast cancer outcomes in humans, the genes identified in Table 1 were cross-referenced to a compendium of breast tumor expression profile datasets (~1340 tumors) (Kessler et al. 2012) to determine if the gene set was associated with differences in tumor type or distant metastasis-free survival (DMFS). Of the 157 genes in Table 1, there were 82 human orthologs represented in the breast tumor compendium. Although the complete gene list as a signature was not associated with tumor type or survival, there were 30 genes in the list that were associated ($p < 0.05$, univariate Cox) with survival. The most significant of these (by Kaplan-Meier analysis, $p < 0.05$) were the orthologs for *Pcna*, *Pgr*, *Cacybp*, *Rgl1*, *Spbc25*, and *Rnft2* (Figure 10). Both *Pcna* and *Pgr* were highly associated ($p < 1 \times 10^{-6}$) with altered DMFS. For *Pcna*, *Spbc25*, *Cacybp*, and *Rnft2*, DMFS was lowest with high expression, while the opposite was true for *Pgr* and *Rgl1*. These results further support the notion that novel gene targets, relevant in the setting of human disease, can be discovered through the analysis of the genetics underlying normal mammary ductal development in mouse populations of varying genetic background.

Intersection of SNP associations with genomic features

The ultimate objective in GWAS is to identify the causal mutations for a particular disease or trait. The vast majority of disease-associated or trait-associated loci are found within non-coding regions of the genome (Li et al. 2012). Consequently, a crucial strategy to identifying the causal mutations underlying these loci has been to identify proxy SNPs within the QTL interval that are in high LD ($r^2 \geq 0.8$) to the associated SNP and then intersect these with regulatory regions identified by open chromatin (Faire-seq, Dnase-seq, Histone modifications) and other genomic features (TF binding, gene expression) (Chen et al. 2014; Maurano et al. 2012; Paul et al. 2013). Of the 6,650 SNP in the high-LD SNP-set, only 1% were in coding DNA. Sixtytwo Of the 157 genes encompassed in the QTL loci, contained SNPs in high LD ($r^2 \geq 0.8$) with the lead SNPs. Only 9 of these high-LD SNPs were found to produce non-synonymous (ns) substitutions in the amino acid sequence for the genes in which they resided. Examination of these substitutions (Table S4) (<http://www.russelllab.org/aas/aas.html>) revealed that 4 were predicted to potentially affect protein function, of *Apobec4*, *Gm5563*, and *Krt76* (Betts 2003). The remaining 5 nsSNPs were predicted to have minimal affect on protein function. In contrast, 99% of high-LD SNP were either intergenic, within introns, or within 5' or 3' untranslated regions. These results emphasize the importance of leveraging data on additional genomic features as a means to narrow the list of potentially causal SNPs in non-coding regions.

SNPs in 3' UTR microRNA target sites

MicroRNA (miR) are known to regulate gene expression by influencing mRNA stability or translation (Ambros 2004). Polymorphisms in miR targets have been associated with several diseases including BrCa (Brewster et al. 2012; Sethupathy and Collins 2008) and PolymiRTS is a useful tool to identify these (Bhattacharya et al. 2014). There were 60 high-LD SNP contained within 3' UTR. Of these, 12 overlapped with miR target sites in 9 genes (Table S5). In 5 of these, the mutant allele is predicted to alter miR target site. These

potentially function-altering UTR SNPs were found within the genes *Lamc1*, *St7*, *Dhrs9*, and *Colgalt*, suggesting that these genes could be candidates for future verification.

Overlapping SNPs with ChIP-seq data and HOMER motifs

To determine if the high-LD SNP within the *mdq* loci represented potential causal loci we intersected these SNP with other genomic features. We used ChIP-seq data for the modification Lysine 4 di-methylation on histone H3 (H3K4me2) as a mark for potential regulatory elements (Rijnkels et al. 2013), and ChIP-seq data for two transcription factors that have known importance in mammary gland biology, STAT5 and progesterone receptor (Lain et al. 2013; Yamaji et al. 2013). Of 6,650 high-LD SNP analyzed 444 were located in K4me2-enriched regions in the mammary gland (Table 2). While 23 and 71 high-LD SNP were located in STAT5a and PgR bound regions, respectively (Table 2). To identify SNP that overlap with predicted transcription factor (TF) binding motifs, we intersected the high-LD SNP set with HOMER motifs (Heinz et al. 2010). Of the 6,650 SNPs in the dataset 2995 overlapped with a HOMER motif (Table 2). A total of 172 of these were also found to overlap with K4me2-enriched regions, while 13 overlapped both with STAT5 and HOMER and 37 with PgR and HOMER. Lastly there were 6 high LD SNP that overlapped with open chromatin in the mammary gland (K4me2), STAT5 bound regions and Homer motifs and 18 overlapped with K4me2, PgR and homer motifs (Table S6). These results provide a more focused list of 208 high-LD, high-priority, SNPs for follow-up as causal loci. Of these, 23 are supported by both ChIP-seq and bioinformatics analysis (Table S6).

DISCUSSION

The heritable nature of BrCa (Peto and Mack 2000) has driven an enormous effort to identify loci within the human genome that modify the risk for this disease. Most recently, the use of genome-wide association studies (GWAS) in human populations has identified 72 loci associated with a modest (relative risk <1.5) increase in BrCa risk (Ghoussaini et al. 2013). The identification of these loci represent an enormous opportunity to further our understanding of BrCa, but this work also points to the complexity of the disease, since all of these common alleles have only modest effects and likely interact in complex ways. Even with all of the loci identified to date, a majority of the heritability for BrCa still remains to be identified (Ghoussaini et al. 2013). This highlights the need for novel approaches to identifying genetic loci that could contribute to BrCa risk. This work highlights the importance of using genetically diverse mouse populations as a means to link early mammary ductal development phenotypes to the potential risk alleles for BrCa in a way that could not be done in human studies.

Technical factors associated with variations in mammary ductal morphology

The analysis of mammary ductal development traits are complicated by two types of factors; technical and biological. From a technical standpoint, the traits are difficult to measure objectively and require sophisticated image analysis techniques (Fernandez-Gonzalez et al. 2004). The current study is one of only a few (Atwood et al. 2000; Fuseler et al. 2014; Robichaux et al. 2014) that uses image processing to segment the ductal tree away from the surrounding stroma and convert it into a structure can be directly measured. Even with this

approach, however, the measurements only describe ductal development on a global scale and do not capture regional variation within the ductal system. To do this with any degree of precision would require the ability to work in 3D. In other branched systems such as the kidney and lung, 3D analysis has facilitated the identification of regional differences in ductal patterning (Metzger et al. 2008; Short et al. 2013). For the lung, this has led to the conclusion that ductal patterning is “hardwired” (Metzger et al. 2008). In the kidney, 3D analysis has identified changes in ductal morphology in response to specific genetic perturbations that would not have been possible with 2D analysis (Short et al. 2013; Short et al. 2014). Although a more thorough 3D analysis of mammary ductal patterning has yet to be done, the imaging presented here represents a first step by comparing two strains, the CZECHII/EiJ and KK/HIJ, for which there are not only differences in global measurements of ductal development, but also regional differences in ductal patterning.

Biological factors associated with variations in mammary ductal morphology

The biological factors that contribute to the complexity of ductal development traits include age, diet, body mass, body composition, reproductive history, genetic background and in early studies, infection with MMTV (Gardner and Strong 1935; Gibson 1930; Huseby and Bittner 1946; Kamikawa et al. 2009; Olson et al. 2010). Although early studies did link MMTV infection to the presence of alveolar nodules, there was no connection to other aspects of ductal morphology (Huseby and Bittner 1946). In addition, the Jackson Laboratory mouse colonies have been free of MMTV since 1999, eliminating this factor from consideration in the current studies (Bean 2000). The moderately strong correlations observed between the ductal development traits and both body mass and age of vaginal patency supports the suggestion that all three of these traits are regulated to some degree by common factors. Variation due to estrus cycle stage was controlled using gonadotropins to synchronize females to estrus at the time of sampling (Hong et al. 2010; Owen et al. 2013). Because age of vaginal patency was not controlled for, some of the variation observed in ductal development can be attributed to the timing of sexual maturation. In addition, variation in responsiveness to gonadotropins for synchronization, as well as the magnitude of specific hormonal changes could also contribute to variations in ductal development. Despite this, ductal development, when analyzed as a series of quantitative traits, displayed moderately high heritability. This was illustrated by the fact that strain contributed significantly ($p < 0.0001$) to the proportion of the observed variance for all traits (Table S1). Along with this, the estimated broad-sense heritabilities for all 5 traits ranged from 0.32 to 0.65 depending on the age at which the traits were measured. Whether this heritability is the result of the above reproductive factors or the direct effects of factors within the mammary gland proper remains to be determined. Regardless of these caveats, the ability to quantitatively describe ductal development traits, estimate their heritabilities, and identify potential QTL has not previously been reported. In addition, opportunities exist in further analysis on strains showing more extreme variations in ductal development like the KK/HIJ and CZECHII.

Biological significance of detected QTL

The main goal of any GWAS study is to identify candidates for subsequent testing and confirmation. The candidate genes and SNP identified in this GWAS were prioritized by

several criteria. The first was effect size, while the second was based on the presence of nsSNPs with potential for modifying protein function. A third criteria was a priori knowledge of biological functions for the genes within any given QTL. Additional criteria included overlap with human BrCA GWAS loci, effects on DMFS (Figure 10), or location of SNP within miR target site, chromatin regulatory region or a transcription factor binding site.

From the standpoint of effect size, our most significant locus was *Mdq1*. Of the genes in this locus, *Plxn4* stands out because of its membership in a family of molecules that regulate axon guidance (Ohta et al. 1992; Peralá et al. 2012). These molecules are also expressed in tissues other than the nervous system and play varying roles in developmental and cancer (Peralá et al. 2012). Several *Plxn* and semaphorin family members are also expressed in the developing mammary gland (Morris et al. 2006). The second most significant association (*Mdq2*) contained the *Lamc1* and *Lamc2* genes, which are important because of their contribution to the extracellular matrix in which the epithelium develops (Aumailley 2013; Klinowska et al. 1999). The fact that *Rgl1* expression in tumors was inversely associated with aggressiveness (Figure 10) suggests that this *mdq2* candidate gene may also be important. Although nothing is known about the actions of *Rgl1* in the mammary gland, we do know that this protein interacts directly with *ras* and can reverse *ras*-dependent transformation in NIH-3T3 cells (Okazaki et al. 1996). Coupling this observation with the fact that *ras* expression and/or activation can influence both normal mammary development and cancer (Andres et al. 1987; Larive et al. 2012) makes *Rgl1* an attractive gene candidate within *Mdq2*. The presence of nsSNP causing potentially significant amino acid substitutions in *Apobec4*, which encodes for an RNA-editing enzyme (Rogozin et al. 2005), supports the suggestion that this *Mdq2* gene could also be an important candidate for functional verification.

Of the remaining QTL detected in this analysis (Table 1), *Mdq4*, *Mdq11*, 13, 14, 15, 17, 19, and 20 all contain gene candidate's which could function as causal loci in the regulation of mammary ductal development or BrCa. The *Mdq4*, 14, 15, and 20 loci contained the *Spc25*, *Pcna*, *Cacybp*, and *Rnft2* genes, respectively, which had among the most significant of affects in the Kaplan-Meier analysis (Figure 10 D). Of these, *Pcna* is essential for DNA replication and repair, *spc25* is essential in mitosis, and *Cacybp* a calcium-binding protein that also plays a potential role in the regulation of β -catenin turnover (Baple et al. 2014; Filipek 2006; McClelland et al. 2004). Both *Mdq11* and 15 are syntenic with BrCa-risk loci that have been detected in human GWAS studies. Of the genes in *Mdq11*, *Plekha7* encodes a protein that interacts with the E-cadherin-p120 complex to regulate tight-junction integrity (Kurita et al. 2013), *Sox6* is expressed in estrogen-receptor-negative mammary luminal progenitor cells and regulates lineage commitment (Kendrick et al. 2008), and *Pik3c2a* belongs to the most mutated signaling pathway in BrCa (Miller et al. 2011). The *Plekha7* gene has also been found to be overexpressed in invasive lobular carcinoma of the breast (Castellana et al. 2012). In *Mdq15*, *Tbx3* is the most interesting in that it encodes a protein that plays a major role in fetal mammary ductal development and has been implicated in a number of cancers including, pancreatic, melanoma, and breast. Importantly, *TBX3* has also been implicated in the dominantly inherited condition known as mammary-ulnar syndrome

(Bamshad et al. 1999; Bamshad et al. 1997). Studies in the mouse have demonstrated that *Tbx3* interacts with a number of important morphogenic signaling pathways including those for retinoic acid signaling, Wnt, FGFR (Cho et al. 2012; Davenport et al. 2003; Eblaghie et al. 2004). Along with this, haplo-insufficiency for this *Tbx3* results in altered placement or reduced numbers of mammary placodes and diminished ductal branching during the fetal period of development (Douglas and Papaioannou 2013). The last 2 QTL of interest, *Mdq13* and *19*, contained the genes *Wnt2*, and *Arhgap42* and *Pgr*, respectively. Expression of both *Wnt2* and *Pgr* are developmentally regulated within the mammary gland ductal epithelium (Buhler et al. 1993; Kouros-Mehr and Werb 2006). In addition, *Pgr* is required for normal mammary ductal side branching (Mulac-Jericevic et al. 2003) and targeted deletion of *Pgr* in the mouse results in delayed tumorigenesis in response to DMBA (Lydon et al. 1999). The protein encoded by *Arhgap42* has not been directly connected to mammary ductal development, but belongs to a family including P190A and B, which are required for normal mammary development P190A and B (Chakravarty et al. 2003; Heckman et al. 2007; Vargo-Gogola et al. 2006).

Intersection with regulatory elements

Because so few of the *mdq* loci contained genes with linked nsSNP that were capable of disrupting protein function, it was important to leverage data on other regulatory features in the genome to narrow the candidate list prior undertaking gene expression studies. There are now several striking examples of success with this approach, along with web-based tools available for the analysis of human regulatory SNP (Boyle et al. 2012; Macintyre et al. 2010; Thomas-Chollier et al. 2011; Ward and Kellis 2012). Unfortunately, application of this approach to mouse data is not as well developed and somewhat more challenging. In this regard while genomic overlaps in the high-LD SNP from this analysis were easily detected, high-throughput approaches for the appropriate statistical tests to predict the biological significance in these overlaps were not as readily available. With these limitations in mind, however, the 208 high-LD SNPs that were found to overlap with open chromatin containing either STAT5 or PgR binding sites or HOMER motifs is a more manageable number of causal loci to screen than the original 6,650. In addition, the presence of 3'UTR SNPs adds and addition 5 SNP to be screened for effects on gene expression.

Summary and conclusions

In summary, using publicly available SNP data in conjunction with quantitative measurements of mammary ductal development traits in strains from the MDP, new loci within the mouse genome have been identified that contain genes of potential importance not only to controlling normal mammary gland development, but also influencing the potential for breast cancer development in humans. As a results there are now some 208 SNPs that can serve as future targets for a potential role in regulating the expression or function of genes underlying mammary ductal development and BrCa. Although future studies are needed to better prioritize and confirm these SNP these initial results support the suggestion that they will further our understanding of both normal mammary ductal development and BrCa.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

The authors thank Ms. Elizabeth Lessels for her contribution to the initial phenotyping. Thanks also to Dr. Daniel Gatti for providing advice on the association analysis and for providing R scripts used in processing the SNP data set in this analysis. Thanks also to Mathew LaCourse for technical assistance related to optical projection tomography, to Dr. John Belmont for providing computing resources for the permutation analysis, and to Fengju Chen for technical assistance related to the mining the human tumor gene expression data. Thanks also to Dr. Jeff Rosen for helpful discussion involving the interpretation of the results, and for valuable suggestions on the manuscript.

FUNDING

This project was supported by NICHD grant number 5R21HD059746 (Darryl Hadsell), by USDA/ARS Cooperative Agreement No. 6250-51000-052 (Darryl Hadsell), and by NCI grant number P30 CA125123 (Chad Creighton). Ian Smyth holds a Future Fellowship from the Australian Research Council. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

- Ambros V. The functions of animal microRNAs. *Nature*. 2004; 431:350–355. [PubMed: 15372042]
- Andres AC, Schonenberger CA, Groner B, Hennighausen L, LeMeur M, Gerlinger P. Ha-ras oncogene expression directed by a milk protein gene promoter: tissue specificity, hormonal regulation, and tumor induction in transgenic mice. *Proceedings of the National Academy of Sciences of the United States of America*. 1987; 84:1299–1303. [PubMed: 3469670]
- Apolant H. Die epitheliale geschwulste des maus. *Arbeiten Koniglchn Ins Exp The Zu Frankfurt*. 1906; 1:61.
- Atwood CS, Hovey RC, Glover JP, Chepko G, Ginsburg E, Robison WG, Vonderhaar BK. Progesterone induces side-branching of the ductal epithelium in the mammary glands of peripubertal mice. *J. Endocrinol*. 2000; 167:39–52. [PubMed: 11018751]
- Aumailley M. The laminin family. *Cell Adh Migr*. 2013; 7:48–55. [PubMed: 23263632]
- Aupperlee MD, Drolet AA, Durairaj S, Wang W, Schwartz RC, Haslam SZ. Strain-specific differences in the mechanisms of progesterone regulation of murine mammary gland development. *Endocrinology*. 2009; 150:1485–1494. [PubMed: 18988671]
- Bamshad M, Le T, Watkins WS, Dixon ME, Kramer BE, Roeder AD, Carey JC, Root S, Schinzel A, Van Maldergem L, Gardner RJ, Lin RC, Seidman CE, Seidman JG, Wallerstein R, Moran E, Sutphen R, Campbell CE, Jorde LB. The spectrum of mutations in TBX3: Genotype/Phenotype relationship in ulnar-mammary syndrome. *Am J Hum Genet*. 1999; 64:1550–1562. [PubMed: 10330342]
- Bamshad M, Lin RC, Law DJ, Watkins WC, Krakowiak PA, Moore ME, Franceschini P, Lala R, Holmes LB, Gebuhr TC, Bruneau BG, Schinzel A, Seidman JG, Seidman CE, Jorde LB. Mutations in human TBX3 alter limb, apocrine and genital development in ulnar-mammary syndrome. *Nat. Genet*. 1997; 16:311–315. [PubMed: 9207801]
- Baple EL, Chambers H, Cross HE, Fawcett H, Nakazawa Y, Chioza BA, Harlalka GV, Mansour S, Sreekantan-Nair A, Patton MA, Muggenthaler M, Rich P, Wagner K, Coblentz R, Stein CK, Last JJ, Taylor AM, Jackson AP, Ogi T, Lehmann AR, Green CM, Crosby AH. Hypomorphic PCNA mutation underlies a human DNA repair disorder. *J. Clin. Invest*. 2014; 124:3137–3146. [PubMed: 24911150]
- Bean SLC, Swing SP, Macauley M, Neleski L. C3H strains free of exogenous mmtv. *JAX Notes*. 2000; 480:1.
- Betts, MJR.; R. B. Amino acid properties and consequences of substitutions. In: Barnes, MRG.; C. I., editors. *Bioinformatics for Geneticists*. Wiley; 2003.

- Bhattacharya A, Ziebarth JD, Cui Y. PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. *Nucleic Acids Res.* 2014; 42:D86–D91. [PubMed: 24163105]
- Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karczewski KJ, Park J, Hitz BC, Weng S, Cherry JM, Snyder M. Annotation of functional variation in personal genomes using Regulome DB. *Genome Res.* 2012; 22:1790–1797. [PubMed: 22955989]
- Brewster BL, Rossiello F, French JD, Edwards SL, Wong M, Wronski A, Whiley P, Waddell N, Chen X, Bove B, Hopper JL, John EM, Andrulis I, Daly M, Volorio S, Bernard L, Peissel B, Manoukian S, Barile M, Pizzamiglio S, Verderio P, Spurdle AB, Radice P, Godwin AK, Southey MC, Brown MA, Peterlongo P. Identification of fifteen novel germline variants in the BRCA1 3'UTR reveals a variant in a breast cancer case that introduces a functional miR-103 target site. *Hum. Mutat.* 2012; 33:1665–1675. [PubMed: 22753153]
- Brunschwig H, Levi L, Ben-David E, Williams RW, Yakir B, Shifman S. Fine-scale maps of recombination rates and hotspots in the mouse genome. *Genetics.* 2012; 191:757–764. [PubMed: 22562932]
- Brunskill EW, Georgas K, Rumballe B, Little MH, Potter SS. Defining the molecular character of the developing and adult kidney podocyte. *PloS one.* 2011; 6:e24640. [PubMed: 21931791]
- Buhler TA, Dale TC, Kieback C, Humphreys RC, Rosen JM. Localization and quantification of Wnt-2 gene expression in mouse mammary development. *Dev. Biol.* 1993; 155:87–96. [PubMed: 8416847]
- Burgess-Herbert SL, Tsaih SW, Stylianou IM, Walsh K, Cox AJ, Paigen B. An experimental assessment of in silico haplotype association mapping in laboratory mice. *BMC. Genet.* 2009; 10:81. [PubMed: 20003225]
- Castellana B, Escuin D, Perez-Olabarria M, Vazquez T, Munoz J, Peiro G, Barnadas A, Lerma E. Genetic up-regulation and overexpression of PLEKHA7 differentiates invasive lobular carcinomas from invasive ductal carcinomas. *Hum. Pathol.* 2012; 43:1902–1909. [PubMed: 22542108]
- Cervino AC, Li G, Edwards S, Zhu J, Laurie C, Tokiwa G, Lum PY, Wang S, Castellini LW, Lusis AJ, Carlson S, Sachs AB, Schadt EE. Integrating QTL and high-density SNP analyses in mice to identify *Insig2* as a susceptibility gene for plasma cholesterol levels. *Genomics.* 2005; 86:505–517. [PubMed: 16126366]
- Chakravarty G, Hadsell D, Buitrago W, Settleman J, Rosen JM. p190-B RhoGAP regulates mammary ductal morphogenesis. *Mol. Endocrinol.* 2003; 17:1054–1065. [PubMed: 12637587]
- Chang SH, Jobling S, Brennan K, Headon DJ. Enhanced Edar signalling has pleiotropic effects on craniofacial and cutaneous glands. *PloS one.* 2009; 4:e7591. [PubMed: 19855838]
- Chen CY, Chang IS, Hsiung CA, Wasserman WW. On the identification of potential regulatory variants within genome wide association candidate SNP sets. *BMC medical genomics.* 2014; 7:34. [PubMed: 24920305]
- Cho KW, Kwon HJ, Shin JO, Lee JM, Cho SW, Tickle C, Jung HS. Retinoic acid signaling and the initiation of mammary gland development. *Dev. Biol.* 2012; 365:259–266. [PubMed: 22387209]
- Churchill GA, Doerge RW. Empirical threshold values for quantitative trait mapping. *Genetics.* 1994; 138:963–971. [PubMed: 7851788]
- Davenport TG, Jerome-Majewska LA, Papaioannou VE. Mammary gland, limb and yolk sac defects in mice lacking *Tbx3*, the gene mutated in human ulnar mammary syndrome. *Development.* 2003; 130:2263–2273. [PubMed: 12668638]
- Davis RC, van Nas A, Bennett B, Orozco L, Pan C, Rau CD, Eskin E, Lusis AJ. Genome-wide association mapping of blood cell traits in mice. *Mamm. Genome.* 2013; 24:105–118. [PubMed: 23417284]
- Douglas NC, Papaioannou VE. The T-box transcription factors *TBX2* and *TBX3* in mammary gland development and breast cancer. *J. Mammary Gland Biol. Neoplasia.* 2013; 18:143–147. [PubMed: 23624936]
- Eblaghie MC, Song SJ, Kim JY, Akita K, Tickle C, Jung HS. Interactions between FGF and Wnt signals and *Tbx3* gene expression in mammary gland initiation in mouse embryos. *J. Anat.* 2004; 205:1–13. [PubMed: 15255957]

- Fernandez-Gonzalez R, Barcellos-Hoff MH, Ortiz-de-Solorzano C. Quantitative image analysis in mammary gland biology. *J. Mammary Gland Biol. Neoplasia*. 2004; 9:343–359. [PubMed: 15838604]
- Filipek A. S100A6 and CacyBP/SIP - two proteins discovered in ehrlich ascites tumor cells that are potentially involved in the degradation of beta-catenin. *Chemotherapy*. 2006; 52:32–34. [PubMed: 16340196]
- Finlay AY, Marks R. An hereditary syndrome of lumpy scalp, odd ears and rudimentary nipples. *Br. J. Dermatol*. 1978; 99:423–430. [PubMed: 708615]
- Flux DS. Growth of the mammary duct system in intact and ovariectomized mice of the CHI strain. *J. Endocrinol*. 1954; 11:223–237. [PubMed: 13211881]
- Frazer KA, Eskin E, Kang HM, Bogue MA, Hinds DA, Beilharz EJ, Gupta RV, Montgomery J, Morenzoni MM, Nilsen GB, Pethiyagoda CL, Stuve LL, Johnson FM, Daly MJ, Wade CM, Cox DR. A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. *Nature*. 2007; 448:1050–1053. [PubMed: 17660834]
- Fuseler JW, Robichaux JP, Atiyah HI, Ramsdell AF. Morphometric and fractal dimension analysis identifies early neoplastic changes in mammary epithelium of MMTV-cNeu mice. *Anticancer Res*. 2014; 34:1171–1177. [PubMed: 24596356]
- Gardner WU, Strong LC. The normal development of the mammary glands of virgin female mice of ten strains varying in susceptibility to spontaneous neoplasms. *Am. J. Cancer*. 1935; 25:285–290.
- Ghabrial AS, Levi BP, Krasnow MA. A systematic screen for tube morphogenesis and branching genes in the *Drosophila* tracheal system. *PLoS Genet*. 2011; 7:e1002087. [PubMed: 21750678]
- Ghazalpour A, Rau CD, Farber CR, Bennett BJ, Orozco LD, van Nas A, Pan C, Allayee H, Beaven SW, Civelek M, Davis RC, Drake TA, Friedman RA, Furlotte N, Hui ST, Jentsch JD, Kostem E, Kang HM, Kang EY, Joo JW, Korshunov VA, Laughlin RE, Martin LJ, Ohmen JD, Parks BW, Pellegrini M, Reue K, Smith DJ, Tetradis S, Wang J, Wang Y, Weiss JN, Kirchgessner T, Gargalovic PS, Eskin E, Lusis AJ, LeBoeuf RC. Hybrid mouse diversity panel: a panel of inbred mouse strains suitable for analysis of complex genetic traits. *Mamm. Genome*. 2012; 23:680–692. [PubMed: 22892838]
- Ghoussaini M, Pharoah PD, Easton DF. Inherited genetic susceptibility to breast cancer: the beginning of the end or the end of the beginning? *Am. J. Pathol*. 2013; 183:1038–1051. [PubMed: 23973388]
- Gibson LM. A comparative study of the life history of the female mammary gland in two strains of albino mice. *Cancer Res*. 1930; 14:31.
- Gjorevski N, Nelson CM. Integrated morphodynamic signalling of the mammary gland. *Nat Rev Mol Cell Biol*. 2011; 12:581–593. [PubMed: 21829222]
- Grupe A, Germer S, Usuka J, Aud D, Belknap JK, Klein RF, Ahluwalia MK, Higuchi R, Peltz G. In silico mapping of complex disease-related traits in mice. *Science*. 2001; 292:1915–1918. [PubMed: 11397946]
- Haaland, M. Spontaneous tumors in mice. In: Bashford, EF., editor. *Fourth Scientific Report on the Investigations of the Imperial Cancer Research Fund*. London: Imperial Cancer Research Fund; 1911. p. 1-113.
- Heckman BM, Chakravarty G, Vargo-Gogola T, Gonzales-Rimbau M, Hadsell DL, Lee AV, Settleman J, Rosen JM. Crosstalk between the p190-B RhoGAP and IGF signaling pathways is required for embryonic mammary bud development. *Dev. Biol*. 2007; 309:137–149. [PubMed: 17662267]
- Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*. 2010; 38:576–589. [PubMed: 20513432]
- Hojsgaard SHU, Robinson-Cox J, Leidi AA. doBy - Groupwise summary statistics, general linear contrasts, population means (least-squares-means), and other utilities. 2013
- Hong H, Yen HY, Brockmeyer A, Liu Y, Chodankar R, Pike MC, Stanczyk FZ, Maxson R, Dubeau L. Changes in the mouse estrus cycle in response to BRCA1 inactivation suggest a potential link between risk factors for familial and sporadic ovarian cancer. *Cancer Res*. 2010; 70:221–228. [PubMed: 20028858]

- Howard B, Panchal H, McCarthy A, Ashworth A. Identification of the scaramanga gene implicates Neuregulin3 in mammary gland specification. *Genes Dev.* 2005; 19:2078–2090. [PubMed: 16140987]
- Howard BA, Gusterson BA. The characterization of a mouse mutant that displays abnormal mammary gland development. *Mamm. Genome.* 2000a; 11:234–237. [PubMed: 10723730]
- Howard BA, Gusterson BA. Mammary gland patterning in the AXB/BXA recombinant inbred strains of mouse. *Mech. Dev.* 2000b; 91:305–309. [PubMed: 10704854]
- Hunter KW. Host genetics and tumour metastasis. *Br. J. Cancer.* 2004; 90:752–755. [PubMed: 14970848]
- Hunter KW. Mouse models of cancer: does the strain matter? *Nat Rev Cancer.* 2012; 12:144–149. [PubMed: 22257951]
- Huseby RA, Bittner JJ. A comparative morphological study of the mammary glands with reference to the known factors influencing the development of mammary carcinoma in mice. *Cancer Res.* 1946; 6:240–255. [PubMed: 21024836]
- Kamikawa A, Ichii O, Yamaji D, Imao T, Suzuki C, Okamatsu-Ogura Y, Terao A, Kon Y, Kimura K. Diet-induced obesity disrupts ductal development in the mammary glands of nonpregnant mice. *Dev. Dyn.* 2009; 238:1092–1099. [PubMed: 19384959]
- Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E. Efficient control of population structure in model organism association mapping. *Genetics.* 2008; 178:1709–1723. [PubMed: 18385116]
- Kendrick H, Regan JL, Magnay FA, Grigoriadis A, Mitsopoulos C, Zvelebil M, Smalley MJ. Transcriptome analysis of mammary epithelial subpopulations identifies novel determinants of lineage commitment and cell fate. *BMC Genomics.* 2008; 9:591. [PubMed: 19063729]
- Kessler JD, Kahle KT, Sun T, Meerbrey KL, Schlabach MR, Schmitt EM, Skinner SO, Xu Q, Li MZ, Hartman ZC, Rao M, Yu P, Dominguez-Vidana R, Liang AC, Solimini NL, Bernardi RJ, Yu B, Hsu T, Golding I, Luo J, Osborne CK, Creighton CJ, Hilsenbeck SG, Schiff R, Shaw CA, Elledge SJ, Westbrook TF. A SUMOylation-dependent transcriptional subprogram is required for Myc-driven tumorigenesis. *Science.* 2012; 335:348–353. [PubMed: 22157079]
- Kirby A, Kang HM, Wade CM, Cotsapas CJ, Kostem E, Han B, Furlotte N, Kang EY, Rivas M, Bogue MA, Frazer KA, Johnson FM, Beilharz EJ, Cox DR, Eskin E, Daly MJ. Fine Mapping in 94 Inbred Mouse Strains Using a High-density Haplotype Resource. *Genetics.* 2010
- Klinowska TC, Soriano JV, Edwards GM, Oliver JM, Valentijn AJ, Montesano R, Streuli CH. Laminin and beta1 integrins are crucial for normal mammary gland development in the mouse. *Dev. Biol.* 1999; 215:13–32. [PubMed: 10525347]
- Kohl M. MK misc: Miscellaneous functions from M. Kohl. 2013
- Kouros-Mehr H, Werb Z. Candidate regulators of mammary branching morphogenesis identified by genome-wide transcript analysis. *Dev. Dyn.* 2006; 235:3404–3412. [PubMed: 17039550]
- Kurita S, Yamada T, Rikitsu E, Ikeda W, Takai Y. Binding between the junctional proteins afadin and PLEKHA7 and implication in the formation of adherens junction in epithelial cells. *J. Biol. Chem.* 2013; 288:29356–29368. [PubMed: 23990464]
- Lain AR, Creighton CJ, Conneely OM. Research resource: progesterone receptor targetome underlying mammary gland branching morphogenesis. *Mol. Endocrinol.* 2013; 27:1743–1761. [PubMed: 23979845]
- Larive RM, Abad A, Cardaba CM, Hernandez T, Canamero M, de Alava E, Santos E, Alarcon B, Bustelo XR. The Ras-like protein R-Ras2/TC21 is important for proper mammary gland development. *Mol. Biol. Cell.* 2012; 23:2373–2387. [PubMed: 22535521]
- Li MJ, Wang P, Liu X, Lim EL, Wang Z, Yeager M, Wong MP, Sham PC, Chanock SJ, Wang J. GWASdb: a database for human genetic variants identified by genome-wide association studies. *Nucleic Acids Res.* 2012; 40:D1047–D1054. [PubMed: 22139925]
- Lifsted T, Le Voyer T, Williams M, Muller W, Klein-Szanto A, Buetow KH, Hunter KW. Identification of inbred mouse strains harboring genetic modifiers of mammary tumor age of onset and metastatic progression. *Int. J. Cancer.* 1998; 77:640–644. [PubMed: 9679770]

- Lydon JP, Ge G, Kittrell FS, Medina D, O'Malley BW. Murine mammary gland carcinogenesis is critically dependent on progesterone receptor function. *Cancer Res.* 1999; 59:4276–4284. [PubMed: 10485472]
- Macintyre G, Bailey J, Haviv I, Kowalczyk A. is-rSNP: a novel technique for in silico regulatory SNP detection. *Bioinformatics.* 2010; 26:i524–i530. [PubMed: 20823317]
- Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, Shafer A, Neri F, Lee K, Kutayavin T, Stehling-Sun S, Johnson AK, Canfield TK, Giste E, Diegel M, Bates D, Hansen RS, Neph S, Sabo PJ, Heimfeld S, Raubitschek A, Ziegler S, Cotsapas C, Sotoodehnia N, Glass I, Sunyaev SR, Kaul R, Stamatoyannopoulos JA. Systematic localization of common disease-associated variation in regulatory DNA. *Science.* 2012; 337:1190–1195. [PubMed: 22955828]
- McClelland ML, Kallio MJ, Barrett-Wilt GA, Kestner CA, Shabanowitz J, Hunt DF, Gorbisky GJ, Stukenberg PT. The vertebrate Ndc80 complex contains Spc24 and Spc25 homologs, which are required to establish and maintain kinetochore-microtubule attachment. *Curr. Biol.* 2004; 14:131–137. [PubMed: 14738735]
- McNally S, Martin F. Molecular regulators of pubertal mammary gland development. *Ann. Med.* 2011; 43:212–234. [PubMed: 21417804]
- Medina D. Of mice and women: A short history of mouse mammary cancer research with an emphasis on the paradigms inspired by the transplantation method. *Cold Spring Harb Perspect Biol.* 2010; 2:a004523. [PubMed: 20630995]
- Megarbane H, Cluzeau C, Bodemer C, Fraitaig S, Chababi-Atallah M, Megarbane A, Smahi A. Unusual presentation of a severe autosomal recessive anhydrotic ectodermal dysplasia with a novel mutation in the EDAR gene. *Am J Med Genet A.* 2008; 146A:2657–2662. [PubMed: 18816645]
- Metzger RJ, Klein OD, Martin GR, Krasnow MA. The branching programme of mouse lung development. *Nature.* 2008; 453:745–750. [PubMed: 18463632]
- Miller BH, Schultz LE, Gulati A, Su AI, Pletcher MT. Phenotypic characterization of a genetically diverse panel of mice for behavioral despair and anxiety. *PLoS. ONE.* 2010; 5:e14458. [PubMed: 21206921]
- Miller TW, Rexer BN, Garrett JT, Arteaga CL. Mutations in the phosphatidylinositol 3-kinase pathway: role in tumor progression and therapeutic implications in breast cancer. *Breast Cancer Res.* 2011; 13:224. [PubMed: 22114931]
- Morris JS, Stein T, Pringle MA, Davies CR, Weber-Hall S, Ferrier RK, Bell AK, Heath VJ, Gusterson BA. Involvement of axonal guidance proteins and their signaling partners in the developing mouse mammary gland. *J. Cell. Physiol.* 2006; 206:16–24. [PubMed: 15920758]
- Mulac-Jericevic B, Lydon JP, DeMayo FJ, Conneely OM. Defective mammary gland morphogenesis in mice lacking the progesterone receptor B isoform. *Proc. Natl. Acad. Sci. U. S. A.* 2003; 100:9744–9749. [PubMed: 12897242]
- Naylor MJ, Ormandy CJ. Mouse strain-specific patterns of mammary epithelial ductal side branching are elicited by stromal factors. *Dev. Dyn.* 2002; 225:100–105. [PubMed: 12203726]
- Neuwirth E. R Color Brewer: Color Brewer palettes. R package version 1.0-2. 2007
- Ochoa-Espinosa A, Affolter M. Branching morphogenesis: from cells to organs and back. *Cold Spring Harb Perspect Biol.* 2012; 4
- Ohta K, Takagi S, Asou H, Fujisawa H. Involvement of neuronal cell surface molecule B2 in the formation of retinal plexiform layers. *Neuron.* 1992; 9:151–161. [PubMed: 1632969]
- Okazaki M, Kishida S, Murai H, Hinoi T, Kikuchi A. Ras-interacting domain of Ral GDP dissociation stimulator like (RGL) reverses v-Ras-induced transformation and Raf-1 activation in NIH3T3 cells. *Cancer Res.* 1996; 56:2387–2392. [PubMed: 8625316]
- Olson LK, Tan Y, Zhao Y, Aupperlee MD, Haslam SZ. Pubertal exposure to high fat diet causes mouse strain-dependent alterations in mammary gland development and estrogen responsiveness. *Int J Obes (Lond).* 2010; 34:1415–1426. [PubMed: 20231845]
- Owen BM, Bookout AL, Ding X, Lin VY, Atkin SD, Gautron L, Klierer SA, Mangelsdorf DJ. FGF21 contributes to neuroendocrine control of female reproduction. *Nat. Med.* 2013; 19:1153–1156. [PubMed: 23933983]

- Paul DS, Albers CA, Rendon A, Voss K, Stephens J, van der Harst P, Chambers JC, Soranzo N, Ouwehand WH, Deloukas P. Maps of open chromatin highlight cell type-restricted patterns of regulatory sequence variation at hematological trait loci. *Genome Res.* 2013; 23:1130–1141. [PubMed: 23570689]
- Perala N, Sariola H, Immonen T. More than nervous: the emerging roles of plexins. *Differentiation.* 2012; 83:77–91. [PubMed: 22099179]
- Peto J, Mack TM. High constant incidence in twins and other relatives of women with breast cancer. *Nat. Genet.* 2000; 26:411–414. [PubMed: 11101836]
- Pletcher MT, McClurg P, Batalov S, Su AI, Barnes SW, Lagler E, Korstanje R, Wang X, Nusskern D, Bogue MA, Mural RJ, Paigen B, Wiltshire T. Use of a dense single nucleotide polymorphism map for in silico mapping in the mouse. *PLoS. Biol.* 2004; 2:e393. [PubMed: 15534693]
- Propper AY, Howard BA, Veltmaat JM. Prenatal morphogenesis of mammary glands in mouse and rabbit. *J. Mammary Gland Biol. Neoplasia.* 2013; 18:93–104. [PubMed: 23736987]
- Rasmussen, SBY.; L, JT.; Smith, GH. Preparing mammary gland whole mounts from mice. In: IP, BB.; M, M.; Asch, editors. *Methods in Mammary Gland Biology and Breast Cancer Research.* New York, NY: Kluwer Academic/Plenum Publishers; 2000.
- Richert MM, Schwertfeger KL, Ryder JW, Anderson SM. An atlas of mouse mammary gland development. *J. Mammary Gland Biol. Neoplasia.* 2000; 5:227–241. [PubMed: 11149575]
- Rijnkels M, Freeman-Zadrowski C, Hernandez J, Potluri V, Wang L, Li W, Lemay DG. Epigenetic modifications unlock the milk protein gene loci during mouse mammary gland development and differentiation. *PloS one.* 2013; 8:e53270. [PubMed: 23301053]
- Robichaux JP, Hallett RM, Fuseler JW, Hassell JA, Ramsdell AF. Mammary glands exhibit molecular laterality and undergo left-right asymmetric ductal epithelial growth in MMTV-cNeu mice*. *Oncogene.* 2014
- Rogozin IB, Basu MK, Jordan IK, Pavlov YI, Koonin EV. APOBEC4, a new member of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases predicted by computational analysis. *Cell Cycle.* 2005; 4:1281–1285. [PubMed: 16082223]
- Sakakura T, Suzuki Y, Shiurba R. Mammary stroma in development and carcinogenesis. *J. Mammary Gland Biol. Neoplasia.* 2013; 18:189–197. [PubMed: 23604977]
- Sethupathy P, Collins FS. MicroRNA target site polymorphisms and human disease. *Trends Genet.* 2008; 24:489–497. [PubMed: 18778868]
- Short K, Hodson M, Smyth I. Spatial mapping and quantification of developmental branching morphogenesis. *Development.* 2013; 140:471–478. [PubMed: 23193168]
- Short KM, Combes AN, Lefevre J, Ju AL, Georgas KM, Lamberton T, Cairncross O, Rumballe BA, McMahon AP, Hamilton NA, Smyth IM, Little MH. Global quantification of tissue dynamics in the developing mouse kidney. *Dev Cell.* 2014; 29:188–202. [PubMed: 24780737]
- Silberstein GB. Postnatal mammary gland morphogenesis. *Microsc. Res. Tech.* 2001; 52:155–162. [PubMed: 11169863]
- Sternlicht MD. Key stages in mammary gland development: the cues that regulate ductal branching morphogenesis. *Breast Cancer Res.* 2006; 8:201. [PubMed: 16524451]
- Stover BC, Muller KF. TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinformatics.* 2010; 11:7. [PubMed: 20051126]
- Tang PL, Cheung CL, Sham PC, McClurg P, Lee B, Chan SY, Smith DK, Tanner JA, Su AI, Cheah KS, Kung AW, Song YQ. Genome-wide haplotype association mapping in mice identifies a genetic variant in CER1 associated with BMD and fracture in southern Chinese women. *J. Bone Miner. Res.* 2009; 24:1013–1021. [PubMed: 19113921]
- Team RDC. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2010.
- Thomas-Chollier M, Hufton A, Heinig M, O'Keeffe S, Masri NE, Roeder HG, Manke T, Vingron M. Transcription factor binding predictions using TRAP for the analysis of ChIP-seq data and regulatory SNPs. *Nature protocols.* 2011; 6:1860–1869. [PubMed: 22051799]
- van Bokhoven H, Hamel BC, Bamshad M, Sangiorgi E, Gurrieri F, Duijf PH, Vanmolkot KR, van Beusekom E, van Beersum SE, Celli J, Merks GF, Tenconi R, Fryns JP, Verloes A, Newbury-Ecob RA, Raas-Rotschild A, Majewski F, Beemer FA, Janecke A, Chitayat D, Crisponi G,

- Kayserili H, Yates JR, Neri G, Brunner HG. p63 Gene mutations in eec syndrome, limb-mammary syndrome, and isolated split hand-split foot malformation suggest a genotype-phenotype correlation. *Am J Hum Genet.* 2001; 69:481–492. [PubMed: 11462173]
- van Genderen C, Okamura RM, Farinas I, Quo RG, Parslow TG, Bruhn L, Grosschedl R. Development of several organs that require inductive epithelial-mesenchymal interactions is impaired in LEF-1-deficient mice. *Genes Dev.* 1994; 8:2691–2703. [PubMed: 7958926]
- van Steensel MA, Celli J, van Bokhoven JH, Brunner HG. Probing the gene expression database for candidate genes. *Eur. J. Hum. Genet.* 1999; 7:910–919. [PubMed: 10602367]
- Vargo-Gogola T, Heckman BM, Gunther EJ, Chodosh LA, Rosen JM. P190-B Rho GTPase-activating protein overexpression disrupts ductal morphogenesis and induces hyperplastic lesions in the developing mammary gland. *Mol. Endocrinol.* 2006; 20:1391–1405. [PubMed: 16469769]
- Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* 2012; 40:D930–D934. [PubMed: 22064851]
- Warnes G, Bolker B, Lumley T. gplots: Various R programming tools for plotting data. R package version 2.6.0.
- Yamaji D, Kang K, Robinson GW, Hennighausen L. Sequential activation of genetic programs in mouse mammary epithelium during pregnancy depends on STAT5A/B concentration. *Nucleic Acids Res.* 2013; 41:1622–1636. [PubMed: 23275557]
- Yant J, Gusterson B, Kamalati T. Induction of strain-specific mouse mammary gland ductal architecture. *The Breast.* 1998; 7:4.
- Yuan R, Meng Q, Nautiyal J, Flurkey K, Tsaih SW, Krier R, Parker MG, Harrison DE, Paigen B. Genetic coregulation of age of female sexual maturation and lifespan through circulating IGF1 among inbred mouse strains. *Proc. Natl. Acad. Sci. U. S. A.* 2012; 109:8224–8229. [PubMed: 22566614]

Figure 1. Lattice plot showing scatter, histograms, and Pearson's correlations of the strain means for all traits measured

A total of 172 virgin female mice across 43 different inbred strains (4/strain) were imported to our animal facility from the Jackson Laboratory at 5 wks of age. Samples were collected at from 171 females at 6 weeks of age (A) and 158 females at 12 weeks (B) of age. Because sequential biopsies were taken from the same animal, the difference (C) between 6 and 12 week biopsies were used as an indicator of rate for each of the traits. Scatter plots in the lower left, histograms along the diagonal and Pearson's correlations in the upper right.

Figure 2. Distributions for quantitative ductal development traits among virgin female mice from inbred strains within the MDP

Ductal development traits were measured in inbred mouse strains (n 2–5/strain) from the MDP at either 6 (n=171) or 12 (n=158) wks of age. The trait distributions at 6 week (pink) or 12 weeks (blue) of age are presented as histograms for ductal area (A), ductal perimeter (B), total branch count (C), total duct length (D), and branch density (E). All five traits were increased ($P < 0.05$) in 12 week samples in comparison to 6 week samples.

Figure 3. Comparison of mammary ductal development traits among inbred strains from within the MDP show 4 development clusters
A heatmap displays the mean normal scores for six quantitative traits describing mammary ductal morphology at either 6 wks, 12 wks, or as the difference between 6 and 12 wks of age (A). All samples were collected from estrus-synchronized females. Strains were hierarchically clustered based the means for each trait-time point combination. Cells colored white have not data for the WSB/EiJ strain. A genetic distance matrix is shown (B) to indicate clustering of closely related strains. Strain names in blue indicate “classical” strains and strains names underlined indicate current common use strains. Detailed descriptive statistics for each of the strains along with ANOVA results and heritability estimates are provided in Table S1.

Figure 4. Images of mammary wholemounts prepared from strains representing each of the development clusters

The images shown are of the left #4 gland collected at 6wks (A,C,E,G,I, K, M, and O) of age and the right #4 gland collected at 12 wks (B,D,F,H,J, L, N, and P) of age from the same female within each strain. At the upper extreme of development in cluster 1 were the BUB/BnJ (C and D) and KK/HIJ (A and B) strains. In cluster 2 were the AKR/J (E and F) and PL/J (G and H) Strains. In cluster 3 were the FVB/NJ (I and J) and C57BL/6J (K and L) strains. At the lower extreme of ductal development were the BALB/cByJ (M and N), and CZECHII/EiJ (O and P) strains.

Figure 5. Imaging of the mammary ductal tree in 3D emphasizes the presence of local differences in mammary ductal topology

Whole mammary glands from females at 6 wks of age were immunostained with an antibody to e-cadherin, imaged by optical projection tomography and then reconstructed in 3D for the CZECHII/EiJ (A; supplemental video 1) and KK/HIJ (D; supplemental video 2) strains. Tree Surveyor allowed for the production of skeletons (B and E) and phylograms (C and F) for the visualization of local patterning variations.

Figure 6. Mammary branch density at 12 wks of age is positively correlated to mammary tumor metastasis

Strain means from the current study were combined with previously published observations on mammary tumor latency and the incidence of lung metastases in F1 offspring from inbred strains crossed with transgenic mice overexpressing polyomavirus middle T (FVB/N-Tg(MMTVPyVT)634Mul/J). Scatterplots, histograms, and Pearson's r are shown for those traits most highly correlated to tumor latency and metastatic index (A). Traits with a Pearson's $r > 0.3$ were significantly correlated ($p < 0.05$). Visualization of branch density at 12 wks of age suggests the presence of two groups of strains each with ductal density that is highly correlated to lung metastatic index (B). Tumor data are from (Lifstedt et al. 1998).

Figure 7. Significant genome-wide associations for mammary development traits

Genome-wide association analysis was conducted using EMMA. Manhattan plots are shown for ductal area (A, F, K) perimeter (B, G, L) branches (C, H, M), length (D, I, N), and density (E, J, O) at 6wk (A–E), 12wk (F–J) and for the delta between 6 and 12 wk (K–O). Regions with significantly associated SNPs (cyan) were identified using a Bonferroni-adjusted threshold of 2×10^{-6} . Significant associations were detected on MMU 1, 2, 5, 6, 7, 13, 15, 16, and 18. Regions of synteny between mouse and human were identified using the UCSC genome browser. These regions were then cross referenced with a list of recently summarized human BrCa risk loci. Chromosomal locations of ductal development loci are shown as squares. Each square is labeled with the corresponding location in the human genome. Mouse loci overlapping human BrCa loci are colored red. Locations of the human BrCa loci were obtained from (Ghoussaini et al 2013).

Figure 8. Visualization of local LD in the top 3 candidate regions

LD was assessed by calculating r^2 for SNP within a region that was 1 MBP on either side of the lead SNP for the top 3 associations based on P-value. These associations were detected on MMU 6 (A), 1 (B), and 18 (C). The associated SNP are plotted along with their respective P-values, as well as recently derived estimates of local recombination rates (Brunschwig et al 2012). Potential candidate genes are also shown as well as potential causal SNP contained within and around genes. For each plot, the lead SNP is shown as the largest symbol, colored blue along with its respective ID and P-value. Symbols colored red are SNPs in

high LD ($r^2 > 0.8$) with the lead SNP, Symbols colored orange are in moderate LD ($0.6 < r^2 < 0.8$). For green symbols, r^2 is between 0.4 and 0.6. Cyan symbols have low LD ($0.2 < r^2 < 0.4$). Blue symbols have no linkage ($r^2 < 0.2$).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 9. Visualization of local LD in candidate regions with overlap to human BrCa risk alleles or containing known mammary ductal development genes

LD was assessed by calculating r^2 for SNP within a region that was 1 MBP on either side of the lead SNP for the top QTL that overlapped with known human BrCa loci (*Mdq11* and *15*) or contained genes with strong evidence supporting a direct role in mammary ductal development (*Mdq19*) based on p-value. These QTL were detected on MMU5 (A), 6, (B), and 9 (C). The associated SNP are plotted along with their respective P-values, as well as recently derived estimates of local recombination rates (Brunschwig et al 2012). Potential candidate genes are also shown as well as potential causal SNP contained within and around

genes. For each plot, the lead SNP is shown as the largest symbol, colored blue along with its respective id and p-value. Symbols colored red are SNPs in high LD ($r^2 > 0.8$) with the lead SNP, symbols colored orange are in moderate LD ($0.6 < r^2 < 0.8$). For green symbols, r^2 is between 0.4 and 0.6. Cyan symbols have low LD ($0.2 < r^2 < 0.4$). Blue symbols have no linkage ($r^2 < 0.2$).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 10. Select GWAS candidate genes correlate with breast cancer aggressiveness and distant metastasis-free survival

Kaplan-Meier plots in breast cancer patients from the Kessler *et al.* dataset, demonstrating six of the 43 candidate genes underlying QTL associated with mammary ductal development in the mouse were also linked to alterations in distant metastasis-free survival (DMFS).

Kaplan-Meier plots are shown for *Pcna* (A), *Pgr* (B), *Rgl1* (C), *Tmem230* (D) (human ortholog: *C20orf30*), *Lrp2* (E), and *Rassf2* (F). Results are from analysis of public gene

expression data from 1340 primary breast cancers, compiled from published studies (Kessler et al. 2012).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Genomic regions associated with variations in postpubertal mammary ductal development.

QTL	Traits	SNP	P	Effect Size ^a	Chr	Location	Mouse Locus	Human Locus	BrCa Locu ^a	Block Location	Block Size ^b	Genes ^c
<i>Mdq1</i>	PerimeterD, BranchesD, LengthD	rs29972830	8.91E-12	4.36	6	32,739,715	6qA3.3	7q32.3	NO	32,546,507	429,737	Plxna4, Plsna4os, Chchd3
<i>Mdq2</i>	Area12, Perimeter12, Branches12, Length12	rs32398376	2.08E-09	3.24	1	153,235,430	1qG2	1q25.3	NO	152,506,397	1,520,391	Edem3, 1700025G04Rik, Tsen15, Colgalt2, Rgl1, Apobec4, Arpc5, NCF2, Sng7, Nmnat2, Lamec2, Lamec1
<i>Mdq3</i>	Area6, Perimeter6, Branches6, Length6	rs134832228	3.18E-09	3.26	18	16,786,338	18qa1	18q12.1	NO	No Block		Cdh2
<i>Mdq4</i>	Density6	rs33051600	2.56E-08	2.95	2	69,387,396	2qC2	2q24.3	NO	69,182,183	244,961	Nostrin, Spe25, G6pc2, Abcb11, Dhrs9, Lrp2
<i>Mdq5</i>	BranchesD	rs30095733	6.33E-08	2.58	13	116,444,425	13qD2.2	5q11.2	YES	116,444,425	186,800	
<i>Mdq6</i>	Density12	rs46013670	7.65E-08	2.61	16	35,694,403	16qB3	3q21.1	NO	35,533,911	831,340	Senna5b, Dirc2, Hspbap1, Parp14, Kpna1, Wdr5b, Fam162a, Ccdc58, Csta1, Sfia211, Gm5483, 2010005H15Rik, Sfta1, bc117090
<i>Mdq7</i>	Area12, Perimeter12, Length12	rs50444438	1.05E-07	2.53	1	174,976,036	1qH3	1q43	NO	174,929,222	111,022	
<i>Mdq8</i>	DensityD	rs27176282	1.24E-07	2.51	2	18,571,206	2qA3	10p12.3/12.1	YES	17,953,218	656,706	H2atb1, A930004D18Rik, Gm17762, Skida1, Mliti10, Mir7655, Dnajc1, Bmi1
<i>Mdq9</i>	PerimeterD, BranchesD	rs36588132	1.40E-07	2.43	9	114,428,637	9qF3	3p22.3/3p23	NO	114,033,461	450,841	Susd5, Bcl2a1c, 4930525D188Rik, 4930520O04Rik, Ctrap, Tmppe, Glb1
<i>Mdq10</i>	BranchesD	rs29830755	2.06E-07	2.36	6	53,318,046	6qB3	7p15.1	NO	53,206,173	374,149	Gm4872, 9430076C15Rik, Creb5
<i>Mdq11</i>	Density12	rs3696903	3.81E-07	2.45	7	115,893,951	7qF1	11p15.2/11p15.1	YES	115,346,337	1,072,227	Sox6, 1700003G18Rik, 1110004F10Rik, Plekha7, Rps13, Pik3c2a
<i>Mdq12</i>	AreaD	rs33753649	5.27E-07	2.35	5	112,763,278	5qF	22q12.1	YES	112,763,278	43,727	Myo18b, 1000798b10Rik
<i>Mdq13</i>	Branches12, Length12	rs30305626	6.76E-07	2.14	6	18,014,442	6qA2	7q31.2	NO	17,639,977	780,274	Capza2, Sl7, Wnt2, Asz1, Cftr, Cttnp2
<i>Mdq14</i>	BranchesD	rs27275156	8.29E-07	2.19	2	131,981,644	2qF2	20p13/20p12	NO	131,980,060	427,064	Rassf2, Slc23a2, Tmem230, Pcna, Cds2, Prokr2
<i>Mdq15</i>	Perimeter6	rs29778211	8.38E-07	2.07	5	119,440,195	5qF	12q24.21	YES	117,633,361	2,324,170	Ksr2, Nos1, Fbxo21, Tesc, Fbxw8, Hrk, Rnf2, 2410131K14Rik, Med13l, 4930413E156Rik, 1700081H04Rik, Gm5563, Gm7538, AW549542, Gm16063, Tbx3, Tbx3os2, Tbx5
<i>Mdq16</i>	DensityD	rs31558147	8.43E-07	2.15	5	76,686,563	5qC3.3	4q12	NO	76,621,475	124,930	Cep135

QTL	Traits	SNP	P	Effect Size ^a	Chr	Location	Mouse Locus	Human Locus	BrCa Locus ^e	Block Location	Block Size ^b	Genes ^c
<i>Mdq17</i>	Perimeter12, Length12	rs46415121	1.06E-06	2.11	15	101,879,006	15qF2	12q13.13	NO	101,937,324	70,768	Krt76, Krt4, Krt79, Krt78
<i>Mdq18</i>	LengthD	rs27336088	1.21E-06	2.17	2	159,752,856	2qH2	20q12	NO	159,363,719	531,385	
<i>Mdq19</i>	PerimeterD	rs29782631	1.52E-06	2.04	9	9,141,014	9qA1	11q22.1	NO	7,901,285	3,294,490	Yap1, 9230110c19Rik, AK129341, 1700128F08Rik, Mir1899, Trpc6, LOC102634431, Pgr, Arhgap42, Gm16833m Cntn5, Mir6237
<i>Mdq20</i>	Perimeter12, Length12	rs36274122	1.66E-06	2.03	1	162,458,462	1qH2.1	1q24.3	NO	160,122,637	4,140,515	Tnn, Mrps14, Cacybp, Rabgap11, Gpr52, Rchl1, Serpinc1, Zbtb37, Gas5, Snord47, Dars2, Cenpl, Klhl20, Ankrd45, 4930469G21Rik, Prdx6, Tnfrsf4, Tnfrsf18, Fasl, Suco, 4930558K02Rik, Pige, Dnm3, 2810442N19Rik, Dnm3os, Mir199a-2, Mir214, Mettl13, Vamp4, Myoc, 7420461p10Rik, Prrc2c, Fmo4, Fmo2, Fmo6, Fmo3, Mroah9, Prnx1, Gorab, Mettl11gb, Kifap, Scyl3, BC055324, Mettl18, Sele, Sell, Selp, Gm16548, F5, Slc19a2

^a Cohen's D,
^b Based on SNP with an r2 0.8 relative to the lead SNP

Table 2

Intersection of high-LD SNP with genomic regulatory features.

Feature	# SNP	# SNP in HOMER-motifs
High-LD SNP	6650	2995
H3K4me2 mammary gland	444	172
Stat5a binding mammary gland	23	13
H3K4me2 and Stat5a	13	6
PgR binding mammary gland	71	37
K4me2 and PgR	29	18
H3K27me3 mammary gland	2408	