Accumulation of multipotent progenitors with a basal differentiation bias during aging of human mammary epithelia

James C Garbe[1], Francois Pepin[1], Fanny Pelissier[1,4], Klara Sputova[1], Agla J Fridriksdottir[2], Diana E Guo[1], Rene Villadsen[2], Morag Park[3], Ole W Petersen[2], Alexander D. Borowsky[5], Martha R Stampfer[1], and Mark A LaBarge[1]*

[1]Life Science Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA
[2]Department of Cellular and Molecular Medicine, University of Copenhagen, Copenhagen, Denmark, DK-2100
[3]School of Medicine, McGill University, Montreal, 687 Pine Avenue West, H3A 1A1, Quebec, Canada
[4]Institute of Bioengineering, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland.
[5]Department of Pathology, University of California Davis, Sacramento, CA 95817, USA

The authors declare no conflicts of interest.

*Contact: Mark A LaBarge; 1 Cyclotron Road, MS977, Berkeley CA 94720; PH (510)-486-4544 FAX (510)486-5568; Email at MALabarge@lbl.gov

Running title: Aging introduces a basal bias in human mammary progenitors

Keywords: Aging, mammary gland, multipotent progenitors, HMEC

ABSTRACT

Women over 50 years of age account for 75% of new breast cancer diagnoses, and the majority of these tumors are of a luminal subtype. Although age-associated changes, including endocrine profiles and alterations within the breast microenvironment, increase cancer risk, an understanding of the molecular mechanisms that underlie these observations is lacking. In this study, we generated a large collection of normal human mammary epithelial cell strains from women aged 16 to 91 years, derived from primary tissues, to investigate the molecular changes that occur in aging breast cells. We found that in finite-lifespan cultured and uncultured epithelial cells, aging is associated with a reduction of myoepithelial cells and an increase in luminal cells that express keratin 14 and integrin α6, a phenotype that is usually expressed exclusively in myoepithelial cells in women under 30. Changes to the luminal lineage resulted from age-dependent expansion of defective multipotent progenitors that gave rise to incompletely differentiated luminal or myoepithelial cells. The aging process therefore results in both a shift in the balance of luminal/myoepithelial lineages and to changes in the functional spectrum of multipotent progenitors, which together increase the potential for malignant transformation. Together, our findings provide a cellular basis to explain the observed vulnerability to breast cancer that increases with age.

INTRODUCTION

The frequency of human carcinomas increases with age, and the dominant paradigm suggests that acquired mutations and epigenetic modifications account for that increase [1]. The vast majority of women (>75%) diagnosed with breast cancer are over 50 years of age [2, 3], and those tumors are largely of a luminal subtype [4]. Although a few mutations are specifically associated with breast cancers, there does not appear to be a strong age-dependent association with particular mutations [5]. Age-related changes in epigenetic modifications, such as DNA methylation, have been reported in a number of different tumors [6, 7]. An age-dependent, estrogen receptor-independent gene expression signature identified in type-matched breast tumors [5], also suggests epigenetic and/or microenvironmental changes are involved in the pathogenesis of age-related breast cancers. However, at present there is not striking genetic or epigenetic evidence that fully explains the increased incidence of breast cancer with advanced age.

Changes in endocrine regulation and in the cellular and molecular composition of breast microenvironment are the foremost factors that a majority of women aged over 50 years have in common. The timing of changes in the endocrine system associated with menopause is correlated with the physical changes in breast composition, although mechanisms are unknown. Estrogen and progesterone can directly impact the epithelium by causing fluctuations in the stem cell pool [8, 9], and are required for gland morphogenesis [10-12]. The effects of cyclic and prolonged exposures of the normal epithelium to sex hormones over a lifespan are not well defined, although it is known that hormone replacement therapy can induce a higher risk of breast cancer [13, 14]. Aging also is associated with physical changes to the breast such as increases in adipose cells and decreased overall density [15] [16]. Changes in protease expression [17] could be related to age-related discontinuities in laminin-111 of the basement membrane [18], which could impact tissue polarity [19]. These changes occur gradually, thus slowly evolving the molecular constituency of the breast microenvironment. Microenvironmental composition was shown to skew the fate decision process in human mammary multipotent progenitors [20]. Thus age-related changes to microenvironment could affect the differentiation and proportion of epithelial lineages.

Changes to the balance of human mammary epithelial cell (HMEC) lineages in the epithelium may presage susceptibility to breast cancer. Loss of the myoepithelial lineage is linked to breast cancer progression, possibly because they produce laminin-111, which is a key basement membrane component that maintains normal polarity [19, 21]. Humans and mice bearing BRCA-1 mutations exhibited increased proportions of putative luminal progenitors that correlated with increased cancer risk [22-24]. In MMTV-PyMT transgenic mice, which are models of Her2 positive luminal-type tumors, the numbers of progenitors increased in a tumor-stage specific manner, suggesting they may play a role in luminal breast cancer pathogenesis as well [25]. Tumors that are thought to arise due to mutations in BRCA-1 account for only about 5% of all breast cancers, whereas age-related breast cancers account for >75% of all cases. Given growing support for an etiological relationship between changes in the progenitors of normal mammary epithelia and cancer progression, it is important to determine the impact of aging on progenitors and more committed lineages of the human mammary epithelium. Herein we reveal fundamental age-dependent transcriptional and functional changes to HMEC that are associated with changes in the distributions of luminal, myoepithelial, and multipotent progenitors, which together provides a cellular basis for increased vulnerability to breast cancer with age.

MATERIALS AND METHODS

Cell culture – HMEC strains were established and maintained according to previously reported methods [26] using M87A medium with oxytocin and cholera toxin [27]. For 3D cultures, a feeder layer of unsorted primary epithelial cells was made in 24-well plates. 250µl of growth factor reduced Matrigel™ (Becton Dickinson) was then layered on top of the feeder layer, and allowed to polymerize at 37ºC. 10,000 FACS-enriched cells were resuspended in 300-350µl of Matrigel™ and then layered on top of the cell-free gel, and then polymerized at 37ºC. Gels were cultured with H14 medium.

Flow cytometry – HMEC at 4th passage were trypsinized and resuspended in their media. For enrichment or identification of LEP and MEP lineages, anti-CD227-FITC (Becton Dickinson, clone HMPV) or anti-CD10-PE, -PE-Cy5 or –APC (BioLegend, clone HI10a), respectively, or of LBP, anti-CD117-Alexa647 (BioLegend, clone 104D2), EpCam-BV421 (BioLegend, clone 9C4) were added to the media at 1:50 for 25 minutes on ice, washed in PBS and sorted or analyzed. Anti-CD49f-PE (Chemicon, clone CBL-458P) was used at 1:100 dilution. Results were consistent across multiple instrumentation platforms: at LBNL, FACS Calibur (Becton Dickenson) for analysis only and FACS Vantage DIVA (Becton Dickinson) for sorting at LBNL and a FACS Aria (Becton Dickinson) at University of Copenhagen, Denmark.

Immunofluorescence and Immunohistochemistry– Cultured HMEC were fixed in methanol:acetone (1:1) at -20ºC for 15 minutes, blocked with PBS/5% normal goat serum/0.1% Triton X-100, and incubated with anti-K14 (1:1000, Covance, polyclonal rabbit) and anti-K19 (1:20, Developmental Studies Hybridoma Bank, clone Troma-III) overnight at 4ºC, then visualized with fluorescent secondary antibodies (Invitrogen). EdU was added to culture media 4 hours prior to harvesting cells for immunofluorescence, and was imaged with A647 click reagents according to manufacturer specifications (Invitrogen). Cells were imaged with a 710LSM microscope (Carl Zeiss). Four-color image analysis of K14, K19, EdU, and DAPI was conducted using a modified watershed method in MatLab software (Mathworks).

Paraffin embedded tissue sections were deparaffinized and antigen retrieved (Vector Labs). For immunofluorescence, sections were blocked and stained as above. Primaries antibodies were K14 as above, K19 (1:100, AbCAM, AAH07628) was K8 was visualized with anti K8 (1:100, AbCAM, clone HK-8). For immunohistochemistry, a

Dako Autostainer was used together with all Dako reagents for staining. Samples were blocked in Dual Endogenous Enzyme, EnVision+ Dual Link System-HRP was used as the secondary antibody (Dako). Liquid DAB+ Substrate Chromogen System was used as the substrate (Dako). Nuclei were stained with Hematoxylin. Sections were incubated with the following primary antibodies for 20-30 min at room temperature: CD10 (ready to use (RTU), Leica, clone56C6), CD117 (1:600, Dako, polyclonal), K5/6 (RTU, Dako, D5/16B4), K19 (RTU, Dako, RCK108), and HMA (1:200, Dako, HHF35).

Gene expression analysis – Expression profiles of morphologically normal human epithelial cells were obtained as described previously[28]. Expression profiles for pre-stasis HMEC at multiple passages were obtained as described previously[27].

Differential expression was performed using R/Bioconductor using the Limma method[29]. A gene is considered differentially expressed if the p-value adjusted by the Benjamini Hochberg[30] false discovery rate (FDR) method is < 0.01 for the LCM samples and 0.1 for the cell lines. A more stringent threshold is used in the case of the cell strains due to the increased power associated with the larger sample size. In the case of the microdissected profiles, cells from women less than 45 years old (n=17) were compared to cells from women 60 years old or older. In the case of the cell strains, the earliest passage (passage 2) for each strain was used. The 3 cell strains from younger women less than 30y (184D,48R and 240L) were compared to 3 strains from women older than 55y (122L,153L,96R).

Cross-platform comparison was done using the HUGO gene symbols contained in the annotations contained in the R/bioconductor packages hgu133a2.db and hgug4112a.db version 2.4.5. Enrichment was performed using Fisher's exact test based on the total number of differentially expressed probes associated with a HUGO gene symbols. In order to facilitate the cross-platform comparison in Figures 1D and 1E and S1, only the top 100 unique genes from the LCM represented on both platforms were used. In the Agilent platform, only the probe with the lower p-value was used if several probes mapped to the same gene. In the Affymetrix platform, the probeset with the highest variance is selected. Accession number for GEO database pending.

Statistical Analysis – Graphpad Prism 5.0 was used for all statistical analysis, with exception of gene expression analysis (see above). One-way ANOVA were used for all

data sets. Linear regression was used to determine changes as a function of age, significance was established when $p<0.05$. Grouped analyses were performed with Bonferonni's test for multiple comparisons and Bartlett's test for equal variance, with significance established when $p<0.05$.

RESULTS

Myoepithelial cells decline, and luminal cells become more numerous and more basal-like with age

We generated a diverse cohort of finite lifespan HMEC strains to facilitate functional analysis of the aging process in humans. This HMEC Aging Resource was derived from a collection of uncultured organoids from >200 individuals aged 16(y)ears-91y. Organoids from 36 reduction mammoplasty (RM) and peripheral non-tumor mastectomy (P) tissues (Table 1) were cultured using low stress medium, M87A with oxytocin [27]. M87A medium supports growth of multiple mammary epithelial lineages for up to 40 to 60 population doublings prior to stasis, a stress-associated senescence arrest [27]. In contrast, the defined serum-free MCDB170-type media, such as the commercially sold MEGM, support $p16^{INK4A}$(-) post-stasis HMEC with basal or myoepithelial (MEP) phenotypes, and a greatly reduced pre-stasis proliferative potential [31, 32] (Fig S1A).

Comparative phenotypic and molecular analyses revealed that pre-stasis HMEC strains retained the same lineages present in vivo. Flow cytometry (FACS) analyses of dissociated mammary epithelial organoids using antibodies that recognized luminal epithelial (LEP) and MEP lineage markers (CD227 and CD10, respectively) [33], demonstrated the presence of both lineages (Fig S1B left panel). Pre-stasis cultures were initiated from undissociated organoids attached to tissue culture plastic [26]. Immunofluorescence analysis of keratin intermediate filament proteins (K)14 and K19 verified that $K14^+/K19^-$ MEP, $K14^-/K19^+$ LEP, and $K14^+/K19^+$ putative progenitors [34, 35] were present in HMEC populations that migrated onto and proliferated on the culture plastic (Fig S1B right panel). All experiments reported here that involved cultured strains were conducted on 4th passage pre-stasis HMEC because they were found to be heterogeneous; LEPs (0.5% to 53% of total) and MEPs (14% to 77%) were present in every pre-stasis strain at 4th passage (for example Fig S1C). Comparison of HMEC lineage markers in two 4th passage strains and two uncultured dissociated organoids using seven-parameter flow cytometry revealed that the $CD10^-/CD227^+$ LEPs correspond to the $EpCam^{hi}/CD49f^{low}$ population, and the $CD10^+/CD227^-$ MEPs correspond to the $EpCam^{low}/CD49f^{hi}$ population (Fig S3). When using flow cytometry, LEP and MEP

lineages were principally defined using CD10 and CD227 in this study because the distribution of the cells expressing those two markers in 4th passage HMEC more closely reflected the distribution observed in primary dissociated organoids, whereas the distributions of EpCam and CD49f expression changed slightly due to culture adaptation.

Lobules involute with advancing age leaving behind ducts and residual lobules with changed morphology [15, 36], suggesting that representation of LEPs, MEPs, and progenitors may change with age. FACS analyses of the 36 HMEC strains revealed that CD227$^-$/CD10$^+$ MEP decreased ($p=0.0019$) and that CD227$^+$/CD10$^-$ LEP increased ($p<0.0001$) as proportions of the total population with age (Fig 1A and 1B). We hypothesized that the age-dependent changes arose either through age-dependent shifts in the proportions of LEPs and MEPs in the organoids used to establish the strains, and/or through intrinsic changes to functional properties of LEPs or MEPs that would aid survival or propagation of one lineage. Examination by FACS of eight uncultured organoids, which were dissociated in parallel, revealed that MEP proportions decreased ($p<0.05$) and proportions of LEPs trended upward with age (Fig 1C and 1D). Evidence of age-dependent functional changes in LEPs that could alter their ability to bind ECM and survive in culture, was observed by FACS measurement of CD49f (integrin $\alpha6$) protein expression, which is used by MEPs to attach to the basement membrane. CD49f was increasingly expressed in LEPs from uncultured organoids as a function of age ($p=0.05$), reaching a level on a par with the MEPs from the same specimen (Fig 1E and 1F). Whereas LEPs and MEPs from women <30y showed unimodal distributions of CD49f expression, bimodal distributions of CD49f expression were measured in MEPs in 3 out of 4, and in 1 out of 4 LEP specimens in the >55y group, suggesting that aging was associate with evolution of additional lineage subsets not present in younger women (Fig 1E). Thus, a decline of MEPs, and an increase of LEPs that exhibited molecular features usually ascribed to MEPs, were measured with increasing age in cultured pre-stasis strains and in cells from uncultured dissociated organoids.

Age-dependent gene expression hallmarks from *in vivo* are preserved in pre-stasis HMEC strains

To determine whether molecular hallmarks of aging *in vivo* were preserved in cultured HMEC strains, gene expression patterns from HMEC strains (three <30y and three >55y) were compared to gene expression patterns from laser capture microdissected (LCM) morphologically normal breast epithelium from 59 individuals aged 27y to 77y undergoing reduction mammoplasty or cancer removal surgery [28]. In the LCM dataset, 3013 unique genes were differentially expressed (FDR<0.05, Limma [29]) between women <45y and ≥60y. The top 100 differentially expressed genes stratified the entire collection of LCM samples as a function of age (Fig 2A). Analysis of six HMEC strains identified 121 unique differentially expressed genes between young and old. There is a significant overlap of 18 genes (*C2orf55, CCDC47, CLDN8, CREBBP, GFER, GRIN1, HCG26, IGF1, LEF1, MAN1A2, NPEPPS, PLEKHA1, PSD, SOCS3, SRSF10, STRN3, TCP11L1 and TF*) between both sets of genes (p=0.04, Fisher's exact test). The 100 genes from the *in vivo* tissues clustered the six HMEC strains, with representatives of early and late passages for each, based on the age of the women (Fig 2B). Consistent with increased proportions of LEPs in strains from women >55y, the gene profiles from EpCam-enriched LEPs from a 19y (strain 240L) and from milk-derived luminal cells (strain 250MK) clustered with the major branch that contained all the strains from women >55y. The gene profiles from CD10-enriched cells from strain 240L clustered with some later passage strains, consistent with the tendency for HMEC strains to become enriched for basal cells with extended culture. Thus, despite the different sources of RNA and gene array platforms, cultured pre-stasis HMEC appear to retain a molecular signature of aging that was identified in vivo.

cKit$^+$ HMEC are putative multipotent progenitors that are more numerous with age

Changes in lineage proportions in 4th passage strains and in organoids were consistent across the strains, although it also was evident that cell culture on plastic caused a selection bias for cells with basal features (i.e. MEPs). Whereas, the proportion of LEPs increased with donor age when viewed at 4th passage or in organoids, it decreased with passage in culture, the decrease becoming pronounced by 8th passage (Fig 3A). The receptor tyrosine kinase cKit was postulated to be a marker of luminal progenitors in human from gene array analyses [22], and empirical evidence in mice

demonstrated its expression on putative luminal progenitors [37]. The proportion of cKit-expressing (cKit$^+$) HMEC, measured with FACS, decreased with passage in pre-stasis strains (Fig 3A). However, cKit$^+$ HMEC increased as a function of age when measured at 4th passage (Fig 3B) and in 11 dissociated primary reduction mammoplasty samples (Fig 3C).

To determine whether cKit$^+$ HMEC were capable of self-maintenance, 4th passage HMEC were FACS-enriched for cKit-expression (Fig 3D) and then cultured for three additional passages to assess the resultant culture composition. FACS analysis of the 8th passage cultures started from cKit$^+$ HMEC (Fig 3F) revealed 20-fold enrichment for CD227$^+$/CD10$^-$ LEP ($p<0.05$) and 2-fold for cKit-expressing cells ($p<0.05$) compared to parallel cultures started with unsorted HMEC (Fig 3E).

To investigate morphogenic capacity, primary cKit$^+$ HMEC from dissociated organoids were embedded in laminin-rich ECM (lrECM) at low density. Compared to cKit$^-$ cells, cKit$^+$ were 6-fold enriched in their ability to form terminal-duct lobular-like units (TDLU) (Fig 3G): 3% of cKit$^+$ cells formed TDLU compared to 0.5% of cKit$^-$ cells (n=3 individuals). The cKit$^+$-derived TDLU were comprised of K19$^+$ LEP and K14$^+$ MEP (Fig 3H). That cKit$^+$ showed a limited ability to self-maintain, gave rise to CD227$^+$/CD10$^-$/K19$^+$ LEP and CD227$^-$/CD10$^+$/K14$^+$ MEP, and were capable of clonal and robust morphological activity in 3D lrECM supported the hypothesis that cKit$^+$ cells were progenitors capable of multilineage differentiation.

Similar behavior was exhibited by RM- and P-derived HMEC

The population of women who undergo reduction mammoplasty procedures is skewed towards younger ages. Much of the material used to establish strains and for organoid analyses from women >60y were from P-derived samples (Table 1). Although P-tissues were normal appearing, there may have been field effects or microtumors that were not detected. Therefore, age-dependent lineage distributions were compared from RM- and P-derived HMEC strains as independent groups. The 21 RM-derived strains showed significantly decreased proportions of MEP ($p<0.05$), and increased LEP ($p<0.003$) and cKit$^+$ HMEC ($p<0.05$) with age (Fig S2A). The 15 P-derived strains showed significantly increased LEP ($p<0.0007$), a trend for increasing cKit$^+$ HMEC

(p=0.0507) and no change in MEPs with age (Fig S2B). Given that the 36 HMEC strains analyzed in this report were derived from a genetically diverse collection of individuals with unknown parity and estrous status, the observed R-square values (from 0.246 to 0.606) and p-values (from 0.05 to <0.0001) suggests a remarkable relationship with effects from aging. Comparison of proportions of the three lineages in RM- and P-derived strains grouped by similar age (24-29y RM vs 24y-30y P, and 41-62y RM vs 45-65y P) also revealed no significant differences within the age groups (Fig S2C-F). Thus, we detected no statistically significant differences in age-grouped RM- vs P-derived strains, and lineage distributions in strains derived from either tissue source followed similar trends with age.

cKit$^+$ progenitors exhibit an age-dependent basal activity bias

To determine whether a differentiation bias arose in HMEC with age, unsorted HMEC and FACS enriched cKit$^+$ cells (Fig 4A) from ten women (five<30y and five>55y) were subjected to lineage-forming assays. Lineage analyses were performed with markers that were different from the ones used to FACS-enrich the cells, plus a marker of DNA synthesis, in order to gain additional information. The ratios of K14 to K19 proteins expressed, and EdU incorporation were measured in each cell using automated quantitative image analysis. After 48 hours of culture, unsorted HMEC from women <30y exhibited pronounced K14$^+$/K19$^-$ MEP populations, and minor K14$^-$/K19$^+$ LEP and K14$^+$/K19$^+$ progenitor populations (Fig 4B and 4D top left). The <30y cKit-enriched population gave rise to three distinct populations corresponding to LEP, progenitors, and MEP, which is consistent with our interpretation of multipotent activity (Fig 4B and 4D top right). Unsorted HMEC from women >55y exhibited a minor K14$^-$/K19$^+$ LEP population and pronounced K14$^+$/K19$^+$ progenitor and K14$^+$/K19$^-$ MEP populations (Fig 4B and 4D bottom left). Surprisingly, >55y cKit-enriched population showed little evidence of differentiation into K14$^-$/K19$^+$ LEPs, instead, exhibiting mainly a K14$^+$/19$^+$ phenotype (Fig 4B and 4D bottom right). Thus, >55y cKit$^+$ progenitors have a defect in differentiation, which makes them unable to produce CD227$^+$/K14$^-$/K19$^+$ LEPs in any significant proportion.

To determine whether HMEC lineages differed in their rate of proliferation as a function of age, the percentage of cells that incorporated EdU in unsorted and cKit$^+$ HMEC was measured. In unsorted cultures, K14$^+$ MEPs from women <30y incorporated significantly more EdU than <30y K19$^+$ LEPs or K14$^+$/K19$^+$ progenitor cells, and more than MEPs from >55y HMEC strains ($p<0.05$) (Fig 4C top). All three lineages in strains from women >55y exhibited similar levels of EdU incorporation, suggesting that there was no proliferative advantage for one lineage over another in the older specimens (Fig 4C). The K14$^+$/K19$^+$ HMEC from the >55y group incorporated 5-fold more EdU than K14$^+$/K19$^+$ cells from the <30y group ($p<0.05$)(Fig 4C), and >55y LEP exhibited a trending increase of EdU incorporation compared to those from women <30y. That the three lineages derived from cKit$^+$ HMEC after only 48h of culture exhibited patterns of EdU incorporation similar to the unsorted cells, but without significant differences, suggests that they were proliferating and in an early stage of differentiation (Fig 4C bottom). That comparable proliferation is exhibited by HMEC with MEP, LEP and progenitor phenotypes from women >55y, but not <30y, may help explain the age-related increase in those populations in strains and organoids.

A small cohort of mammary epithelia exhibit increased basal characteristics with age in vivo

To determine the status of keratin expression in vivo, K8, K14, and K19 protein expression were evaluated in paraffin embedded tissue sections of normal breast tissue from three women <45y and three >65y. In addition to markers of lineage, MEPs and LEPs can also be identified in vivo by their positioning: MEPs are adjacent to basement membrane on the basal side of the gland, and LEPs are surrounded by MEPs on the luminal interior side adjacent to the luminal space. Basally-located K14-expressing MEPs were observed in every specimen (Fig 5A-D). Luminally-located LEPs in the women <45y expressed more K19 than LEPs in samples >65y (Fig 5A and 5B), whereas there was more K8 expressed in LEPs from the women >65y than in <45y (Fig 5C and 5D). In the women >65y luminally-located LEP layers exhibited more heterogeneity of K19 expression, with several K19$^-$ cells adjacent to K19$^+$ cells. K14$^+$/K19$^+$ progenitors also appeared more frequently in the samples from women >65y (Fig 5B inset). Though

qualitative, the samples could be compared to one another because they were stained in parallel and imaged in a single session using the same microscope settings. Moreover, sweat glands within the same sections also stained for K8, K14, and K19, but age-dependent changes in K8 and K19 intensity were not detected (data not shown), thus fixation artifacts were unlikely.

Similar changes were observed in immunohistochemical analysis of 4μm serial sections from three women <45y and three >65y; serial sections that were representative of the analysis are shown from lobules of a 37y and a 76y woman (Fig 5E). In the 37y lobules, K19 was intense and uniformly present in all LEP and was polarized at the luminal surface, cKit^{+} cells were infrequent and scattered, MEP marker smooth muscle actin (SMA) was intense in the MEPs, and basal keratin K5/6 exhibited light expression in the MEP layer and no expression in LEPs (Fig 5E top row). In contrast, the lobule of the 76y woman showed light K19 expression without polarity in LEPs, strong and frequent expression of cKit, comparable SMA, and strong K5/6 in LEPs as well as the MEPs (Fig 5E bottom row). These in vivo observations of changes in K19 and K14 expression and the acquisition of more basal phenotypes across the lineages of the mammary epithelium are consistent with the conclusions drawn from the primary and cultured HMEC studies.

DISCUSSION

Aging-related changes to the breast are well known to occur, but their impact on the mammary epithelium and their relationship to the increased frequency of breast cancer in women >55y are not well understood. Molecular signatures of aging at the level of gene expression were reported in human kidney and muscle [38, 39], and now in mammary gland (Fig 2), which suggests that functionally the tissues also should differ in young and old. A major challenge that is faced in studying human aging at the cell and molecular levels is the lack of model systems that facilitate a functional understanding of the consequences of molecular changes. We have addressed this problem by using new methodologies that allow long-term growth of HMEC of multiple lineages from women of all ages, enabling examination of normal cell strains in controlled contexts. Culture systems are imperfect replicas of in vivo, however the biochemical and functional phenotypes of aging that were revealed upon examination of the cell strains were corroborated to large extent by observations of the same phenotypes in vivo (e.g. in dissociated uncultured organoids, paraformaldehyde fixed breast tissue sections, and gene expression from LCM normal breast epithelia).

Functional and molecular interrogation of HMEC strains juxtaposed with analyses of primary organoids and normal breast tissue sections revealed that the proportion of MEPs declined, whereas LEPs increased with age. The LEPs from women >55y were surprisingly distinctive compared to their younger counterparts. Through the aging process, LEPs unexpectedly acquired some myoepithelial-like characteristics, which were consistent with age-dependent changes in proportions and activity of cKit$^+$ HMEC. Cultured cKit$^+$ HMEC from women <30y exhibited functional properties of multipotent progenitors that gave rise to LEPs and MEPs, but their activity changed with age, exposing a tendency in cKit$^+$ HMEC from donors >55y to produce LEPs that frequently expressed K14 in addition to K19 and CD227 (Muc1). MEPs derived from cKit$^+$ HMEC of women >55y also exhibited less intense expression of K14 relative to K19, indicating a tendency for the multipotent progenitors to incompletely differentiate into either LEP or MEP lineages, as they were defined in HMEC from younger women. Importantly, our conclusion that cKit$^+$ HMEC exhibited activity of multipotent progenitors was made possible by functional evaluation of cells enriched from many women in both age groups,

in addition to the use of organotypic 3D culture and quantitative single cell analyses of cell fate decisions. The conclusion would potentially have been different had cKit$^+$ HMEC from fewer individuals in only one age group been evaluated. The changes to the epithelium described herein could make older women more vulnerable to malignant progression. Myoepithelial cells are thought to be tumor-suppressive [19, 21], and progenitors are putative etiological roots of some mammary tumors [22-25]. Thus during the aging process a putative target population of cells is increased and there is simultaneous decrease in the cells thought to suppress tumorigenic activity.

The factors that lead to the expansion of progenitors are likely to encompass the sum of aging-related changes that occur over an individual's lifespan. For example, endocrine changes from menarche through menopause are well documented, and exert system-wide effects. Early menarche and late menopause correlates with increased breast cancer risk [40], as does use of progestin in hormone replacement therapy [41]. Dissociated primary human mammary organoids exhibited increased mammosphere-forming units following progesterone exposure, suggesting an increase in multipotent progenitors [42]. In mice, progesterone at the luteal dioestrus phase [9], and estrogen plus progesterone during pregnancy [8] were shown to cause amplification of the mammary progenitor pool. Perhaps lifelong repeated hormone exposures during menstrual cycles and pregnancies are the basis for a gradual net gain of cKit$^+$ HMEC.

Age-associated changes to the local breast microenvironment also may account for some of the observed changes in gene expression patterns and lineage distributions. Known age-related changes include: increased adipose, decreased connective tissue [15, 17], decreased overall density [16], disruptions in the basement membrane [18], and changes in protease expression [43]. These types of changes may alter distributions of the different epithelial lineages because human mammary progenitor cell fate decisions can be influenced by changes in microenvironment [20]. Embryonic microenvironments [47], and tumor core versus periphery regions [48] were shown to correlate with microenvironment-specific epigenetic modifications to tumor cells. Extending these concepts to normal adult tissues, changes to the breast microenvironment also may help explain age-dependent gene expression patterns and lineage distributions. The fact that an age-dependent gene expression signature persists in early to late passage cultured

strains argues strongly for metastable epigenetic forms of regulation as important mechanistic components of the age-dependent phenotypes measured here.

Cell culture has a strong selection bias for basal phenotypes, defined here as cells that express CD10, integrins, K14, K5, and bears no markers of LEPs. In vivo, the majority of cells with basal phenotypes are located on the basal surface of the gland in contact with the basement membrane. This study was made possible because the in vitro basal selection problem was partly solved by using the M87A medium; nevertheless MEP phenotypes dominated the cultures in late passages regardless of age, indicating that long-term maintenance of the luminal phenotype in culture remains a challenge. The proportional reduction of $CD10^+$ MEPs with age is puzzling because cells with basal phenotypes are better able to thrive in ECM-rich microenvironments, such as adjacent to the basement membrane. Moreover, the reduction was obvious by FACS analysis with MEP-specific surface markers in a diverse and large cohort of cultured strains and primary organoids, but was difficult to discern in histological sections. It is unlikely that age-related loss of MEPs measured by FACS is due to adaptation to culture causing loss of CD10 expression, because culture conditions tend to drive more basal phenotypes and the decrease in MEPs was observed in dissociated uncultured organoids. One explanation could be that aging-related loss of lobules [15, 36] created an enrichment of ductal structures, thus the strains established from tissues of older women had an intrinsically different distribution of lineages to begin with. Indeed, analyses of 19 primary organoids by FACS (Fig 1E and F, 3C) indicated there were changes in lineage distributions with age in the absence of culture. A second explanation, which is not mutually exclusive with the first, is that the changes in lineage were proportional, and the LEPs and $cKit^+$ progenitors proliferate as well as MEPs in HMEC from women >55y, whereas in women <30y there was a distinct proliferative advantage in the MEP (Fig 4C). Finally, the reduction in MEPs also could be related to the global changes in gene expression that were observed in the LCM epithelia. Reduced *LAMA1* expression, the alpha-chain component of laminin-111 that is crucial for normal polarity and is normally expressed by MEPs [19, 21], indicated modification of the MEPs genetic program. Although uncultured LEPs gained protein expression of integrin α6 with age, gene expression of integrins α6 and β1 were globally reduced with age. Both integrins have

been shown to be components of a feed-back circuit that regulates the MEP phenotype in mammary epithelial cells from humans and mice [20, 49], suggesting that the basal-regulatory machinery may be disrupted in MEPs, and inappropriately engaged in LEPs, during the aging process. Use of FACS was emphasized in this study because the technology enabled quantification of lineage distributions in heterogeneous tissues. However, doing so necessitated removing the cells from their in vivo context and the small number of histological sections examined did not allow statistical analysis. Thus histological studies of large cohorts of normal breast tissues using the markers described herein will be required to completely reconcile age-related changes that were measured with FACS with histological changes.

The age-dependent epithelial changes described herein, combined with microenvironmental, endocrine, genetic and epigenetic changes may presage, age-dependent vulnerability to breast cancer. To adequately address potential preventive and therapeutic interventions, it is important to understand how the aging process is linked to changes in the balance of lineages and to the functioning of progenitors and their progeny, and whether there is a direct link between these age-related phenotypes and cancer progression.

Acknowledgements

We thank Drs. Mina Bissell and James Lorens for evaluating this manuscript, and Batul Merchant for technical support of this work. MAL is supported by the National Institute on Aging, R00AG033176 and R01AG040081, and by the U.S. Department of Energy Laboratory Directed Research and Development (LDRD) and the Low Dose Radiation Research Program Contract Contract No. DE-AC02-05CH11231. JCG and MRS are supported by Department of Defense grant BCRP BC060444.

References

1. DePinho, R.A., *The age of cancer.* Nature, 2000. **408**(6809): p. 248-54.
2. Jemal, A., et al., *Cancer statistics, 2007.* CA Cancer J Clin, 2007. **57**(1): p. 43-66.
3. Smigal, C., et al., *Trends in breast cancer by race and ethnicity: update 2006.* CA Cancer J Clin, 2006. **56**(3): p. 168-83.
4. Carey, L.A., et al., *Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study.* JAMA, 2006. **295**(21): p. 2492-502.
5. Yau, C., et al., *Aging impacts transcriptomes but not genomes of hormone-dependent breast cancers.* Breast Cancer Res, 2007. **9**(5): p. R59.
6. Issa, J.P., et al., *Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon.* Nat Genet, 1994. **7**(4): p. 536-40.
7. Waki, T., et al., *Age-related methylation of tumor suppressor and tumor-related genes: an analysis of autopsy samples.* Oncogene, 2003. **22**(26): p. 4128-33.
8. Asselin-Labat, M.L., et al., *Control of mammary stem cell function by steroid hormone signalling.* Nature, 2010. **465**(7299): p. 798-802.
9. Joshi, P.A., et al., *Progesterone induces adult mammary stem cell expansion.* Nature, 2010. **465**(7299): p. 803-7.
10. Brisken, C., et al., *A paracrine role for the epithelial progesterone receptor in mammary gland development.* Proc Natl Acad Sci U S A, 1998. **95**(9): p. 5076-81.
11. Lydon, J.P., et al., *Mice lacking progesterone receptor exhibit pleiotropic reproductive abnormalities.* Genes Dev, 1995. **9**(18): p. 2266-78.
12. Mulac-Jericevic, B., et al., *Defective mammary gland morphogenesis in mice lacking the progesterone receptor B isoform.* Proc Natl Acad Sci U S A, 2003. **100**(17): p. 9744-9.
13. Ross, R.K., et al., *Effect of hormone replacement therapy on breast cancer risk: estrogen versus estrogen plus progestin.* J. Nat. Cancer Inst., 2000. **92**(4): p. 328-32.
14. Schairer, C., et al., *Menopausal estrogen and estrogen-progestin replacement therapy and breast cancer risk.* JAMA : the journal of the American Medical Association, 2000. **283**(4): p. 485-91.
15. Milanese, T.R., et al., *Age-related lobular involution and risk of breast cancer.* J Natl Cancer Inst, 2006. **98**(22): p. 1600-7.
16. McCormack, V.A., et al., *Changes and tracking of mammographic density in relation to Pike's model of breast tissue aging: a UK longitudinal study.* Int J Cancer, 2010. **127**(2): p. 452-61.
17. Well, D., et al., *Age-related structural and metabolic changes in the pelvic reproductive end organs.* Semin Nucl Med, 2007. **37**(3): p. 173-84.
18. Howeedy, A.A., et al., *Differential distribution of tenascin in the normal, hyperplastic, and neoplastic breast.* Lab Invest, 1990. **63**(6): p. 798-806.
19. Gudjonsson, T., et al., *Normal and tumor-derived myoepithelial cells differ in their ability to interact with luminal breast epithelial cells for polarity and basement membrane deposition.* J Cell Sci, 2002. **115**(Pt 1): p. 39-50.
20. LaBarge, M.A., et al., *Human mammary progenitor cell fate decsions are products of interactions with combinatorial microenvironments.* Integrative Biology, 2009. **1**(1): p. 70-79.

21. Hu, M., et al., *Regulation of in situ to invasive breast carcinoma transition.* Cancer Cell, 2008. **13**(5): p. 394-406.
22. Lim, E., et al., *Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers.* Nat Med, 2009. **15**(8): p. 907-13.
23. Molyneux, G., et al., *BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells.* Cell Stem Cell, 2010. **7**(3): p. 403-17.
24. Proia, T.A., et al., *Genetic Predisposition Directs Breast Cancer Phenotype by Dictating Progenitor Cell Fate.* Cell Stem Cell, 2011. **8**: p. 149-163.
25. Kouros-Mehr, H., et al., *GATA-3 links tumor differentiation and dissemination in a luminal breast cancer model.* Cancer Cell, 2008. **13**(2): p. 141-52.
26. Stampfer, M., R.C. Hallowes, and A.J. Hackett, *Growth of normal human mammary cells in culture.* In Vitro, 1980. **16**(5): p. 415-25.
27. Garbe, J.C., et al., *Molecular distinctions between stasis and telomere attrition senescence barriers shown by long-term culture of normal human mammary epithelial cells.* Cancer Res, 2009. **69**(19): p. 7557-68.
28. Finak, G., et al., *Gene expression signatures of morphologically normal breast tissue identify basal-like tumors.* Breast Cancer Res, 2006. **8**(5): p. R58.
29. Smyth, G.K., *Limma: linear models for microarray data*, in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, R.C. Gentleman, et al., Editors. 2005, Springer: New York. p. 397-420.
30. Benjamini, Y. and Y. Hochberg, *Controlling the false discovery rate: a practical and powerful approach to multiple testing.* J. Roy. Statist. Soc. Ser, 1995. **B 57**(1): p. 289-300.
31. Taylor-Papadimitriou, J., et al., *Keratin expression in human mammary epithelial cells cultured from normal and malignant tissue: relation to in vivo phenotypes and influence of medium.* J Cell Sci, 1989. **94 (Pt 3)**: p. 403-13.
32. Brenner, A.J., M.R. Stampfer, and C.M. Aldaz, *Increased p16 expression with first senescence arrest in human mammary epithelial cells and extended growth capacity with p16 inactivation.* Oncogene, 1998. **17**(2): p. 199-205.
33. Raouf, A., et al., *Transcriptome analysis of the normal human mammary cell commitment and differentiation process.* Cell Stem Cell, 2008. **3**(1): p. 109-18.
34. Gudjonsson, T., et al., *Isolation, immortalization, and characterization of a human breast epithelial cell line with stem cell properties.* Genes Dev, 2002. **16**(6): p. 693-706.
35. Villadsen, R., et al., *Evidence for a stem cell hierarchy in the adult human breast.* J Cell Biol, 2007. **177**(1): p. 87-101.
36. Hutson, S.W., P.N. Cowen, and C.C. Bird, *Morphometric studies of age related changes in normal human breast and their significance for evolution of mammary cancer.* J Clin Pathol, 1985. **38**(3): p. 281-7.
37. Regan, J.L., et al., *c-Kit is required for growth and survival of the cells of origin of Brca1-mutation-associated breast cancer.* Oncogene, 2011.
38. Rodwell, G.E., et al., *A transcriptional profile of aging in the human kidney.* PLoS Biol, 2004. **2**(12): p. e427.

39. Zahn, J.M., et al., *Transcriptional profiling of aging in human muscle reveals a common aging signature.* PLoS Genet, 2006. **2**(7): p. e115.
40. Bernstein, L., *Epidemiology of endocrine-related risk factors for breast cancer.* J Mammary Gland Biol Neoplasia, 2002. **7**(1): p. 3-15.
41. Rossouw, J.E., et al., *Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial.* JAMA, 2002. **288**(3): p. 321-33.
42. Graham, J.D., et al., *DNA replication licensing and progenitor numbers are increased by progesterone in normal human breast.* Endocrinology, 2009. **150**(7): p. 3318-26.
43. Imai, S., et al., *Dissociation of Oct-1 from the nuclear peripheral structure induces the cellular aging-associated collagenase gene expression.* Mol Biol Cell, 1997. **8**(12): p. 2407-19.
44. Sternlicht, M.D., et al., *The stromal proteinase MMP3/stromelysin-1 promotes mammary carcinogenesis.* Cell, 1999. **98**(2): p. 137-46.
45. Radisky, D.C., et al., *Rac1b and reactive oxygen species mediate MMP-3-induced EMT and genomic instability.* Nature, 2005. **436**(7047): p. 123-7.
46. Sternlicht, M.D., M.J. Bissell, and Z. Werb, *The matrix metalloproteinase stromelysin-1 acts as a natural mammary tumor promoter.* Oncogene, 2000. **19**(8): p. 1102-13.
47. Costa, F.F., et al., *Epigenetically reprogramming metastatic tumor cells with an embryonic microenvironment.* Epigenomics, 2009. **1**(2): p. 387-398.
48. Jie, G., et al., *Relationship between expression and methylation status of p16INK4a and the proliferative activity of different areas' tumour cells in human colorectal cancer.* Int J Clin Pract, 2007. **61**(9): p. 1523-9.
49. Deugnier, M.A., et al., *Cell-extracellular matrix interactions and EGF are important regulators of the basal mammary epithelial cell phenotype.* J Cell Sci, 1999. **112 (Pt 7)**: p. 1035-44.

Figure Legends

Figure 1 – Epithelial lineages change as a function of age. (A) Representative FACS analyses of CD227 and CD10 expression in 4th passage HMEC strains isolated from one woman <30years (195L) and one >55years (805P). FACS plots are shown as 5% contour plots with outliers identified, at left are isotype antibody controls, and at right are the CD10 and CD227 stained samples. Gates identifying luminal epithelial (LEP) and myoepithelial (MEP) are shown. (B) Linear regression showing changes in proportions of LEPs (green) and MEPs (red) in HMEC strains at 4th passage as a function of age (n=36 individuals). LEPs and MEPs from reduction mammoplasty (RM)-derived strains are shown with filled circles or boxes, and from peripheral to tumor (P)-derived strains with open circles or boxes, respectively. (C) Representative FACS analyses from the corresponding uncultured dissociated epithelial organoids. FACS plots are shown as 5% contour plots with outliers identified, at left are isotype antibody controls, and at right are the CD10 and CD227 stained samples. (D) Linear regression of proportions of LEPs (green) and MEPs (red) in dissociated uncultured organoids as a function of age (n=8 individuals). LEPs and MEPs from RM-derived organoids are shown with filled circles or boxes, and from P-derived organoids with open circles or boxes, respectively. (E) Histograms of CD49f (integrin α6) expression by flow cytometry on CD227+ LEPs (green lines) and CD10+ MEPs (red lines) from dissociated organoids. The gray-colored shade boxes indicate the threshold at which there is little or no CD49f expression as determined in isotype negative control stains (gray lines). (F) Regression analysis of Log_2 change in mean expression of CD49f in LEPs normalized to the levels in MEPs from dissociated organoids as a function of age (n=8 individuals).

Figure 2 – A 100 gene signature stratifies human mammary epithelial cells by age. (A) The 100 most variable age-dependent genes identified from a set of 59 laser captured micro-dissected (LCM) phenotypically normal human mammary epithelium samples stratified the gene expression profiles by age. (B) The same signature clustered gene expression profiles of multiple passages and replicates of HMEC strains 184, 48RT, and 240L (<30y) and 122L, 153, and 96R (>55y), in an age-dependent manner. Heat maps

represent Z-scores for each gene, where red represents higher expression and green represents lower expression. A positive-fold change represents a higher expression in samples from younger women. Specimen names are shown just below the heat maps. Profiles from FACS-enriched EpCam$^+$ (LEP) and CD10$^+$ (MEP) 240L HMEC, and luminal 250MK HMEC from isolated milk (MILK) are indicated. Samples from multiple passages of each HMEC strain are shown (passage is denoted with 'p'), and most were analyzed with biological replicates (denoted either with '.1 vs .2' or 'A vs B').

Figure 3 – Proportions of cKit$^+$ HMEC, putative multipotent progenitors, increase with age. (A) Changes in proportions of LEPs and cKit$^+$ HMEC in three representative strains as a function of passage. (B) Linear regression of proportions of cKit$^+$ HMEC in strains at 4th passage as a function of age (n=36 individuals). cKit$^+$ HMEC from RM-derived strains are shown with filled triangles and from P-derived strains with open triangles. (C) Linear regression of proportions of cKit$^+$ cells in dissociated uncultured organoids as a function of age (n=11). (D) FACS plot showing the gating logic used for sorting cKit$^+$ HMEC from strain 122L at 4th passage. Inset, shows the LEP and MEP distribution at 4th passage. (E) FACS analysis of strain 122L at 8th passage. (F) FACS analysis of cKit-enriched-derived cultures at 8th passage. (G) Phase images of representative structures derived from cKit$^+$ (left) and cKit$^-$ (right) cells cultured in laminin-rich basement membrane for 14 days. (H) Immunofluorescence of a transverse frozen section that shows keratin (K)14 (red) and K19 (green) protein expression in a duct of a cKit$^+$-derived TDLU-like structure from 3D culture. Nuclei were stained with DAPI (blue), the three-color merged image is shown at right.

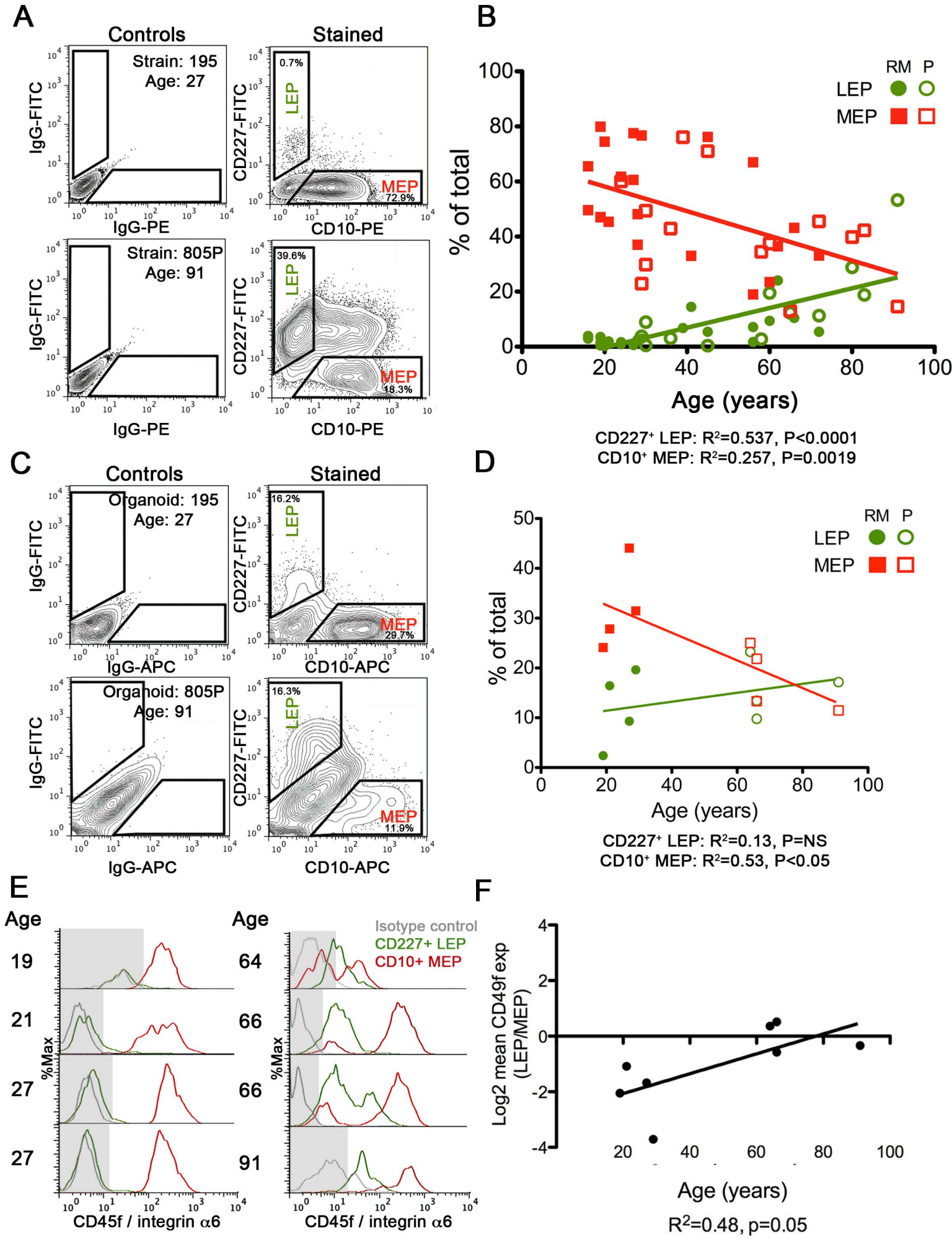
Figure 4 – cKit$^+$ progenitors exhibit age-dependent differentiation defects. (A) A representative contour FACS plot from strain 353P, showing the gating logic used to enrich cKit$^+$ from 4th passage HMEC strains. (B) Histograms representing average lineage distributions from five individuals <30y (strains 240L, 407P, 168R, 123, and 124) or five >55y (strains 122L, 881P, 353P, 464P, and 451P) in unsorted HMEC (left column), and cKit$^+$ progenitors (right column) after 48h of culture on tissue culture plastic. Histograms represent $\log_2$ transformed ratios of K14 to K19 protein expression in

single cells, histograms are heat mapped to indicate cells with the phenotypes of $K14^-$ /$K19^+$ LEPs (green) $K14^+$/$K19^+$ progenitors (yellow), and $K14^+$/$K19^-$ MEPs (red), error bars represent SD (n=2500 cells/histogram). (C) Scatter plot representing EdU incorporation in the different lineages, as defined by K14 and K19 expression, in (top) unsorted 4th passage HMEC strains and (bottom) $cKit^+$-derived cells after 48h culture. Lines indicate average and error bars represent SEM (n=2500 cells per age group). (D) Representative images of unsorted, CD227- and cKit-enriched HMEC from two individuals, after 48h of culture; Protein expression of K14 (red), K19 (green) are shown, and nuclei are stained with DAPI (blue). Scale bar represents 20μm.

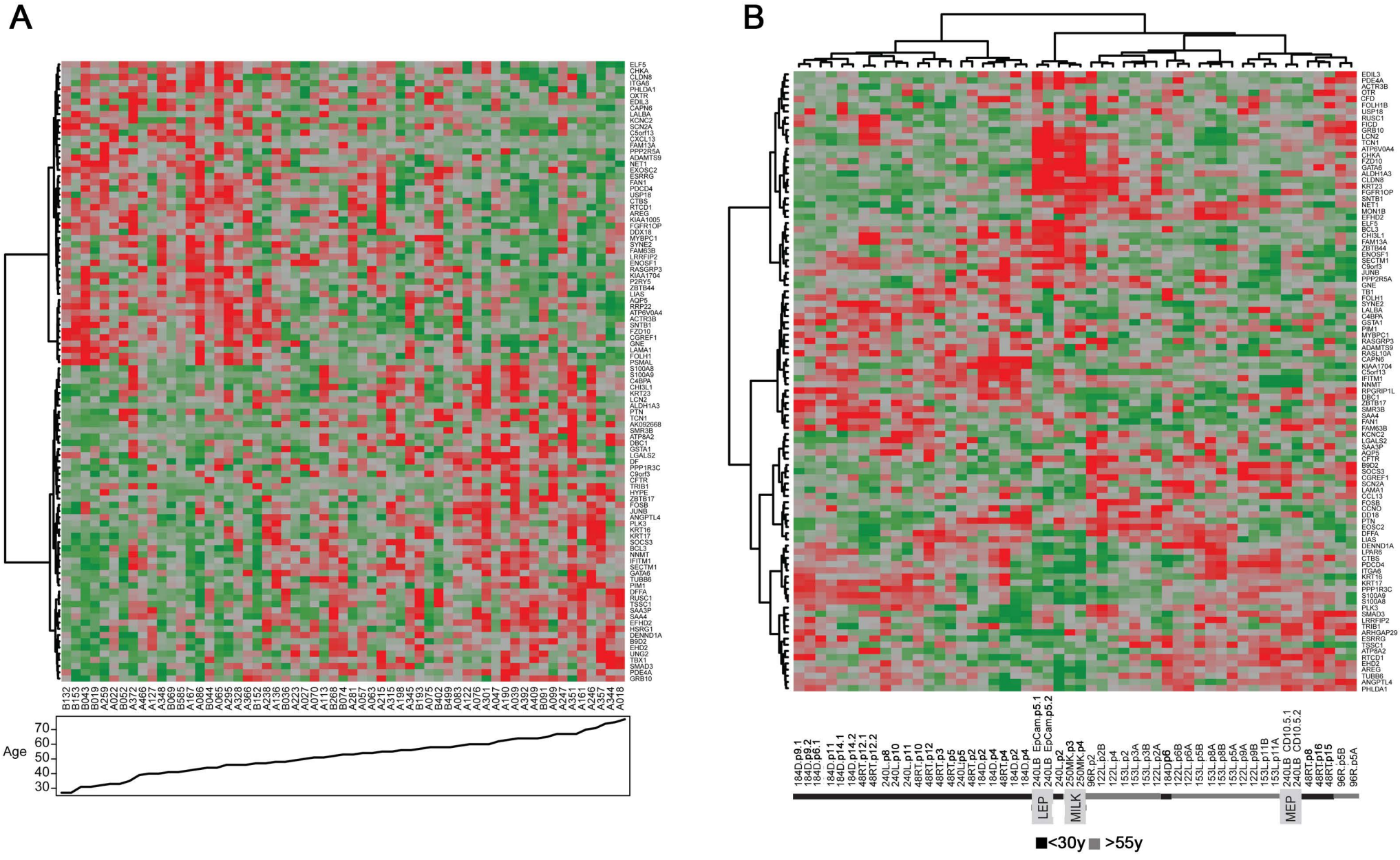
Figure 5 – Protein expression patterns in vivo are consistent with findings in cultured HMEC strains. Immunofluorescence images of normal mammary glands from six individuals from two age groups: (A and B) are stained to show K14 (red) and K19 (green) expression, (C and D) are stained to show K14 (red) and K8 (green) expression; DAPI nuclear stain shown in all images (blue). Scale bar represents 100μm. Insets show a higher magnification view of the indicated locations within each image, the arrows in 'B' point to two clusters of $K14^+$/$K19^+$ cells. (E) 4μm serial sections shows a lobule from representative 37y and 76y women immunohistochemically stained to show expression of K19, cKit/CD117, smooth muscle actin (SMA), and K5/6 (brown), nuclei are stained blue. Scale bar represents 50μm.

Table 1 – List of specimens.

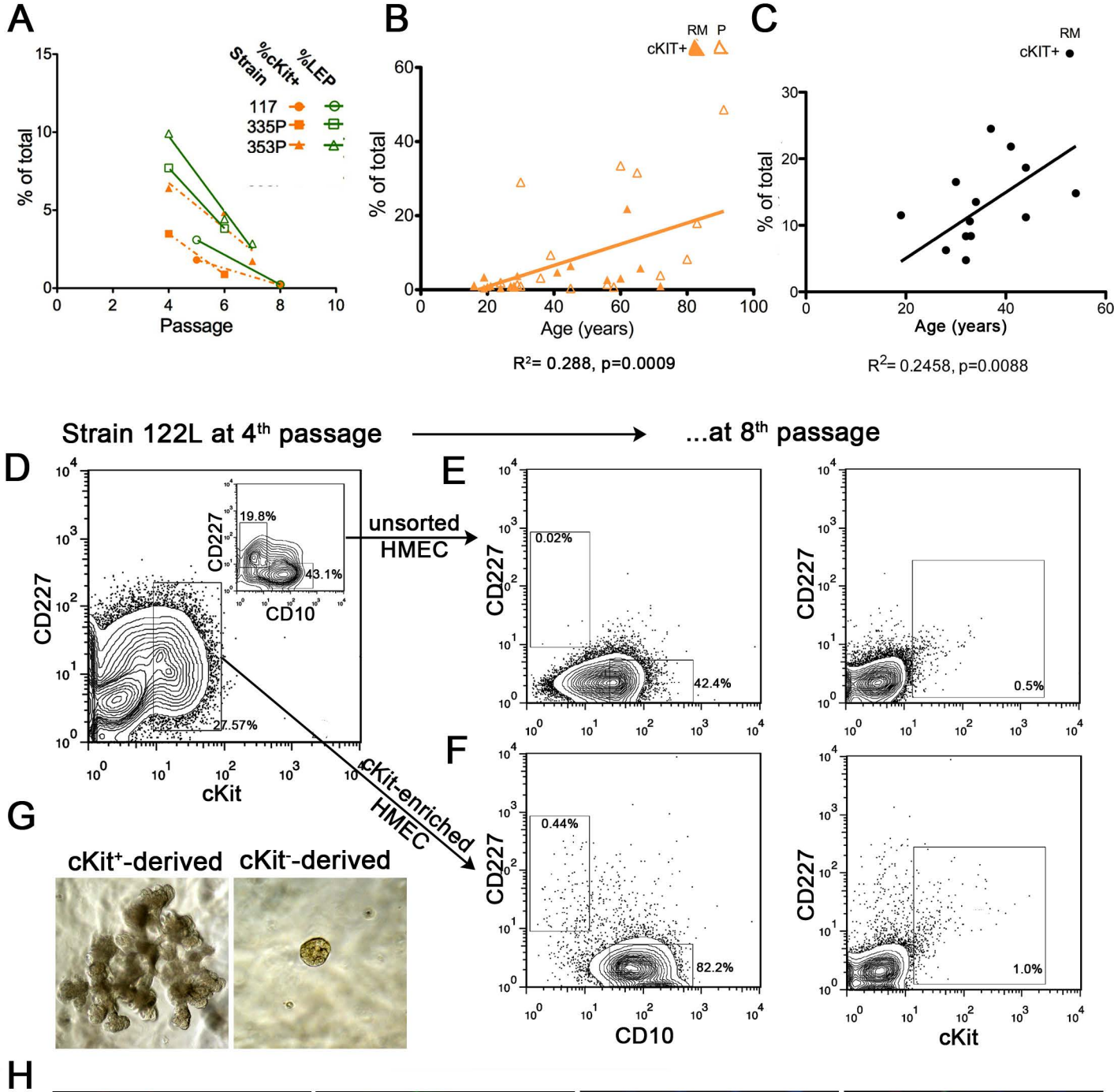
Garbe etal Figure 1



Garbe et al Figure 2



Garbe etal Figure 3



Garbe etal Figure 4

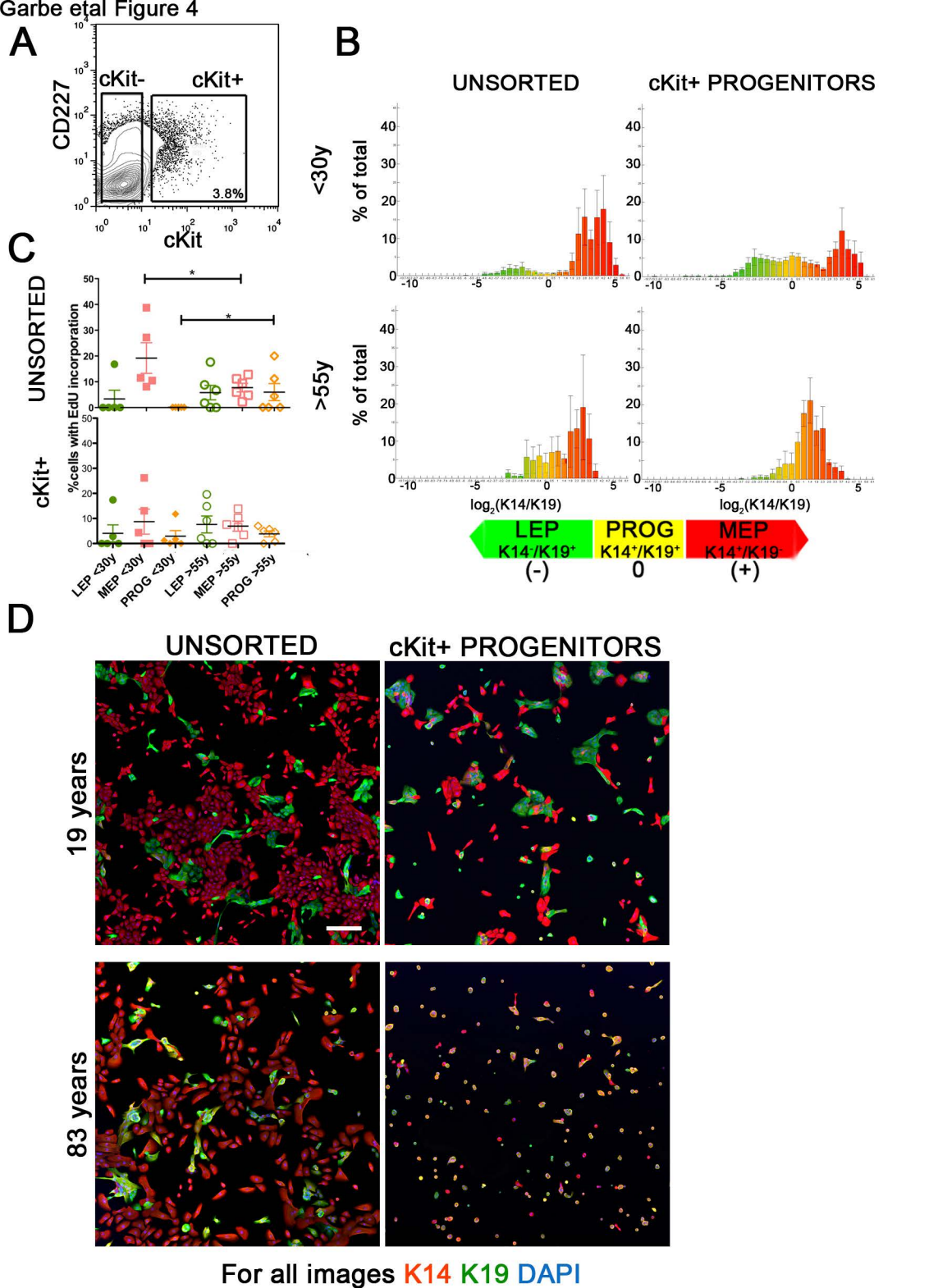
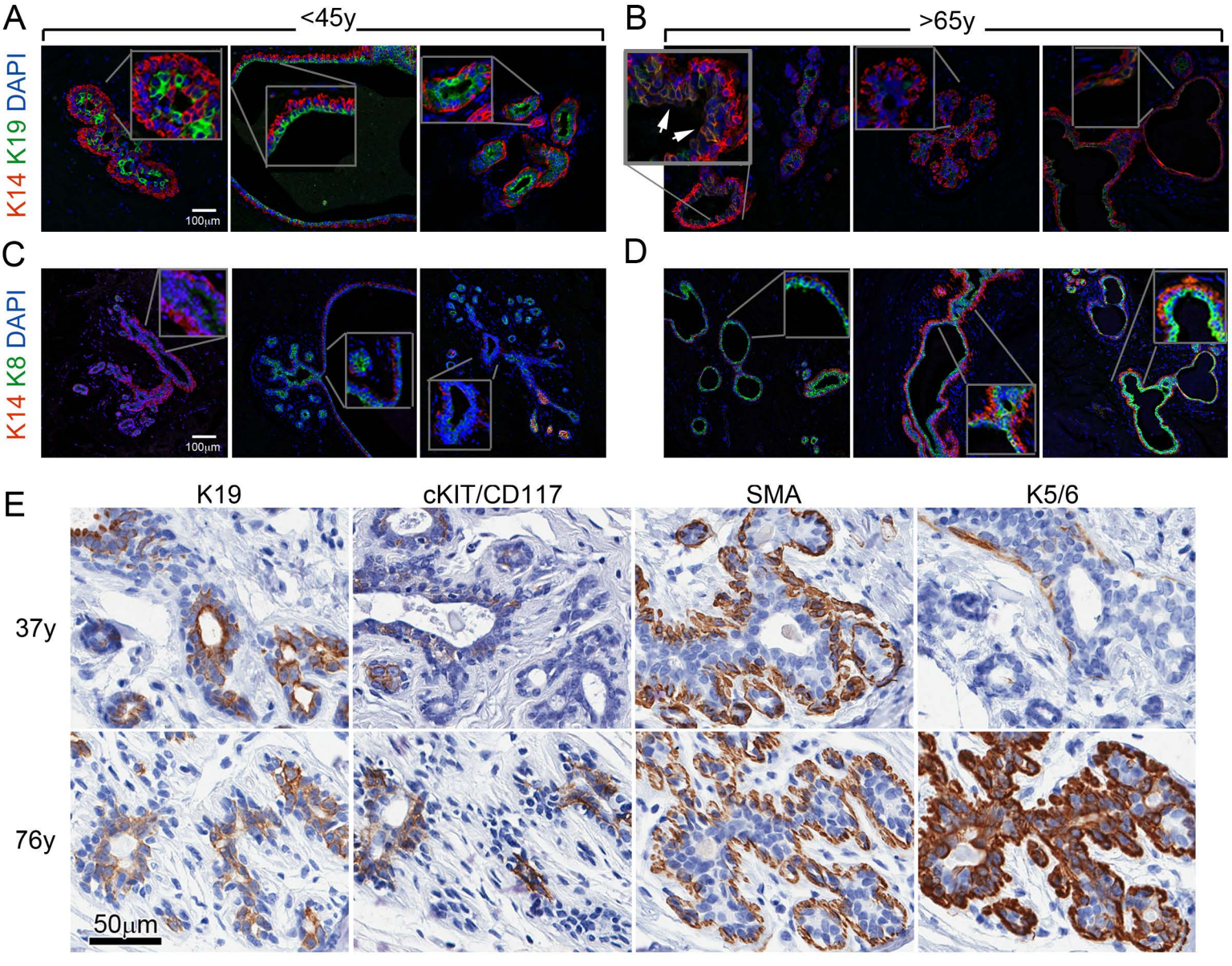


Figure 5 Garbe etal



Garbe etal Table 1 - Characteristics of HMEC strains and uncultured organoids

Sample	Age (years)	Source	Country of surgery	Est prestasis strain	Organoid analysis	Known characteristics/pathology notes
48R	16	RM	USA	yes		African American
160	16	RM	USA	yes		
240	19	RM	USA	yes		mildly hyperplastic
404	19	RM	USA	no	CD10,CD227,CD49f	minimal fibrocystic change
407P	19	P	USA	yes		IDC, lymphnode+, ER+, PR+
168R	19	RM	USA	yes		African American
D1	19	RM	DK	no	cKit	
399	20	RM	USA	yes		benign
356	21	RM	USA	yes	CD10,CD227,CD49f	normal
184	21	RM	USA	yes		
1P	24	P	USA	yes		IDC, lymphnode-
97	27	RM	USA	yes		
195L	27	RM	USA	yes	CD10,CD227,CD49f	
123	27	RM	USA	yes		African American, mammary hyperplasis
400	27	RM	USA	no	CD10,CD227,CD49f	mild ductal ectsia and mastitis
51L	28	RM	USA	yes		mild periductal mastitis (R+L), focal microcalcification (R.)
172L	28	RM	USA	yes		minimal phase of fibrocystic disease
D2	28	RM	DK	no	cKit	
124	29	RM	USA	yes		
676P	29	P	USA	yes		
42P	30	P	USA	yes		IDC, lymphnode-
1030P	30	P	USA	yes		
D2	30	RM	DK	no	cKit	
D4	32	RM	DK	no	cKit	
D5	33	RM	DK	no	cKit	
D6	34	RM	DK	no	cKit	
90P	36	P	USA	yes		BRCA1 mut (185delAG)
D7	37	RM	DK	no	cKit	
100P	39	P	USA	yes		non-invasive ductal carcinoma, ER-, PR-
D8	41	RM	DK	no	cKit	
245AT	41	RM	USA	yes		ATM heterozyote, tissue was clinically normal tissue
D9	44	RM	DK	no	cKit	
D10	44	RM	DK	no	cKit	
173P	45	RM	USA	yes		IDC, ER-, PR-
D11	54	RM	DK	no	cKit	
191	56	RM	USA	yes		slight fibrocystic disease, hypertrophy, stromal fibrosis and adenosis present in mammary parenchyma
117	56	RM	USA	yes		patchy stromal fibrosis (R.), fibrocystic disease (L.)
335P	58	P	USA	yes		Infiltrating adenocarcinoma, ER+, PR+/-
153L	60	RM	USA	yes		benign fibrocystic disease
639P	60	P	USA	yes		
122L	62	RM	USA	yes		fibrocystic disease, hypertrophy, apocrine metaplasia of ductal epithelium, cystic dilatia of ducts and focal areas of intraductal hyperplasia
394P	64	P	USA	no	CD10,CD227,CD49f	IDC
881P	65	P	USA	yes		
355P	66	P	USA	no	CD10,CD227,CD49f	Infiltrating adenocarcinoma
374P	66	P	USA	no	CD10,CD227,CD49f	infiltrating adenocarcinoma, Padget disease
96R	66	RM	USA	yes		slight focal fibrocystic change
429	72	RM	USA	yes		
353P	72	P	USA	yes		Colloid (mucinous) carcinoma, ER+/-, PR-
464P	80	P	USA	yes		
451P	83	P	USA	yes		
805P	91	P	USA	yes	CD10,CD227,CD49f	

Soure: RM = Reduction mammoplasty tissue, P = Peripheral non-tumor containing mastectomy tissue.
Country of surgery: USA = Peralta Cancer Center, Oakland CA USA; DK = University of Copenhagen, Copenhagen, Denmark
Organoid analysis: anti CD10 clone HI10a, anti CD227 #559774, anti CD49f clone GoH3, anti cKit clones 104D2 and K45.
Known characteristics/pathology notes: Information is incomplete, but was provided here cases in which some outstanding feature was known to the authors or was noted in pathology reports. (ER=estrogen receptor, PR=progesterone receptor, IDC=invasive ductal carcinoma.)

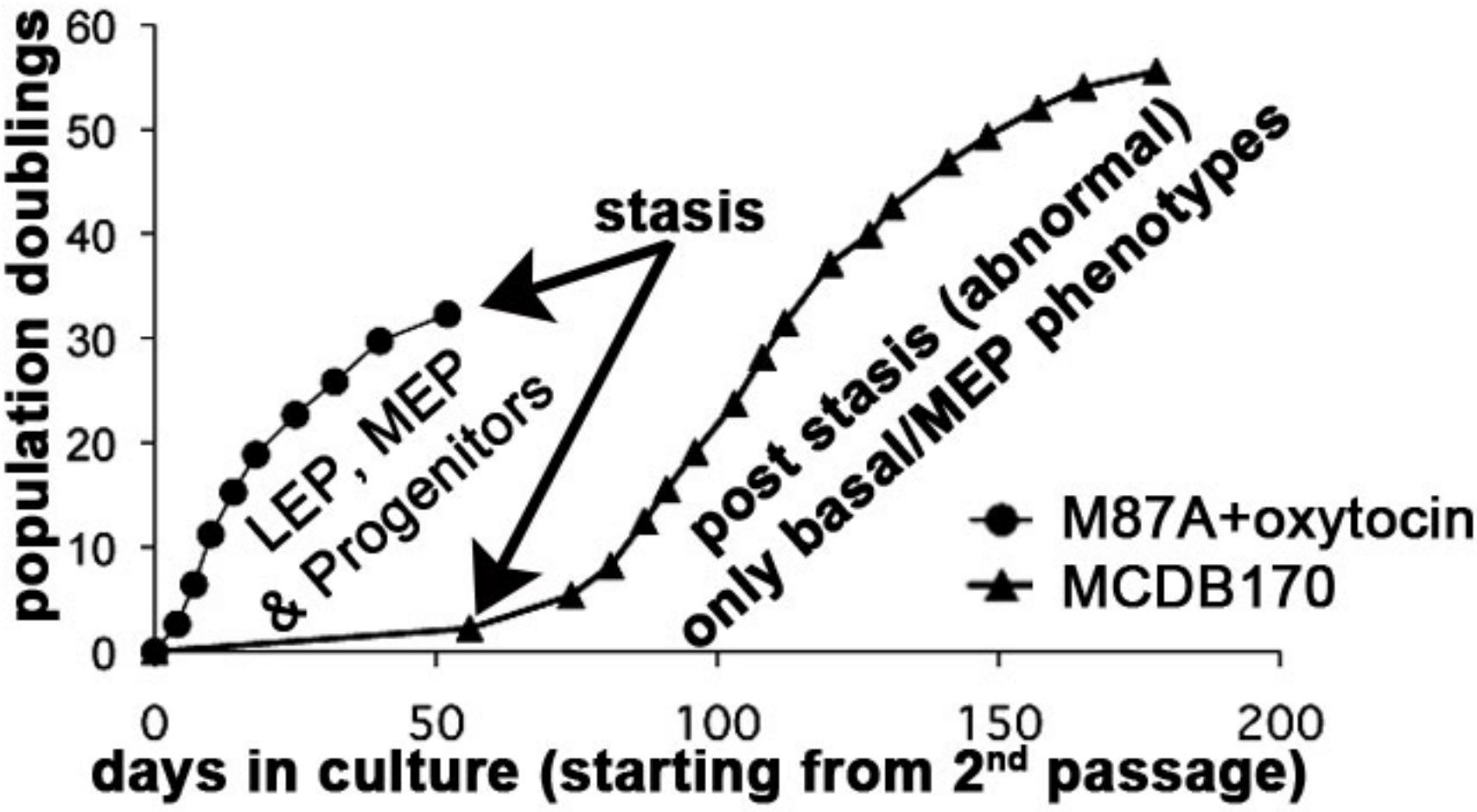
Figure S1 – Cultured pre-stasis finite lifespan HMEC at early passage preserve key features of HMEC *in vivo*. (A) Representative growth curves of HMEC strain 240 from 2nd passage grown in M85+oxytocin (a formulation similar to M87A) versus MCDB170+supplements (commercially MEGM). Growth in M87A stops at stasis (stress-associated senescence) [27, 31]; however MCDB170 induces epigenetic changes leading to abnormal post-stasis HMEC growth [31, 48] (B) Uncultured organoids; (left) FACS analysis for CD227 (Muc-1) and CD10 (CALLA) on cells of dissociated uncultured organoids from 51L mammoplasty tissue shows presence of LEP ($CD10^-$/$CD227^+$) and MEP ($CD10^+$/$CD227^-$) lineages. (right) Epithelial lineages in primary organoid outgrowths in M87A+oxytocin identified by staining with antibodies to K14 (red) and K19 (green); nuclei were stained with DAPI (blue). LEP ($K14^-$/$K19^+$, green), MEP ($K14^+$/$K19^-$, red), and progenitors ($K14^+$/$K19^+$, yellow) are visible. Unstained cells are observed in the organoid core due to incomplete antibody penetration. (C) 4th passage HMEC; (left) Typical FACS and (right) immunofluorescence analyses of a pre-stasis culture at 4th passage.

Figure S2 – Analysis of HMEC lineages in strains derived from reduction mammoplasty (RM) versus peripheral non-tumor regions from mastectomy (P) tissues. (A) Linear regression showing changes in proportions of LEP (filled circles), MEP (filled squares), and $cKit^+$ HMEC (filled triangles) in RM-derived HMEC strains at 4th passage as a function of age (n=21 individuals). (B) Linear regression showing changes in proportions of LEP (open circles), MEP (open squares), and $cKit^+$ HMEC (open triangles) in P-derived HMEC strains at 4th passage as a function of age (n=15 individuals). Associated statistics are shown at the bottom of the regression plots. Dot graphs showing comparisons of the proportions of (C) LEP, (D) MEP, and (E) $cKit^+$ HMEC in 24-29y RM versus 24-30y P and in 41-62y RM versus 45-65y P age groups. Group averages and SE are shown, RM-derived samples are denoted with filled symbols and P-derived with open symbols. (F) A summary of statistics showing the ANOVA and t-tests that were used to compare the groups in C-E.

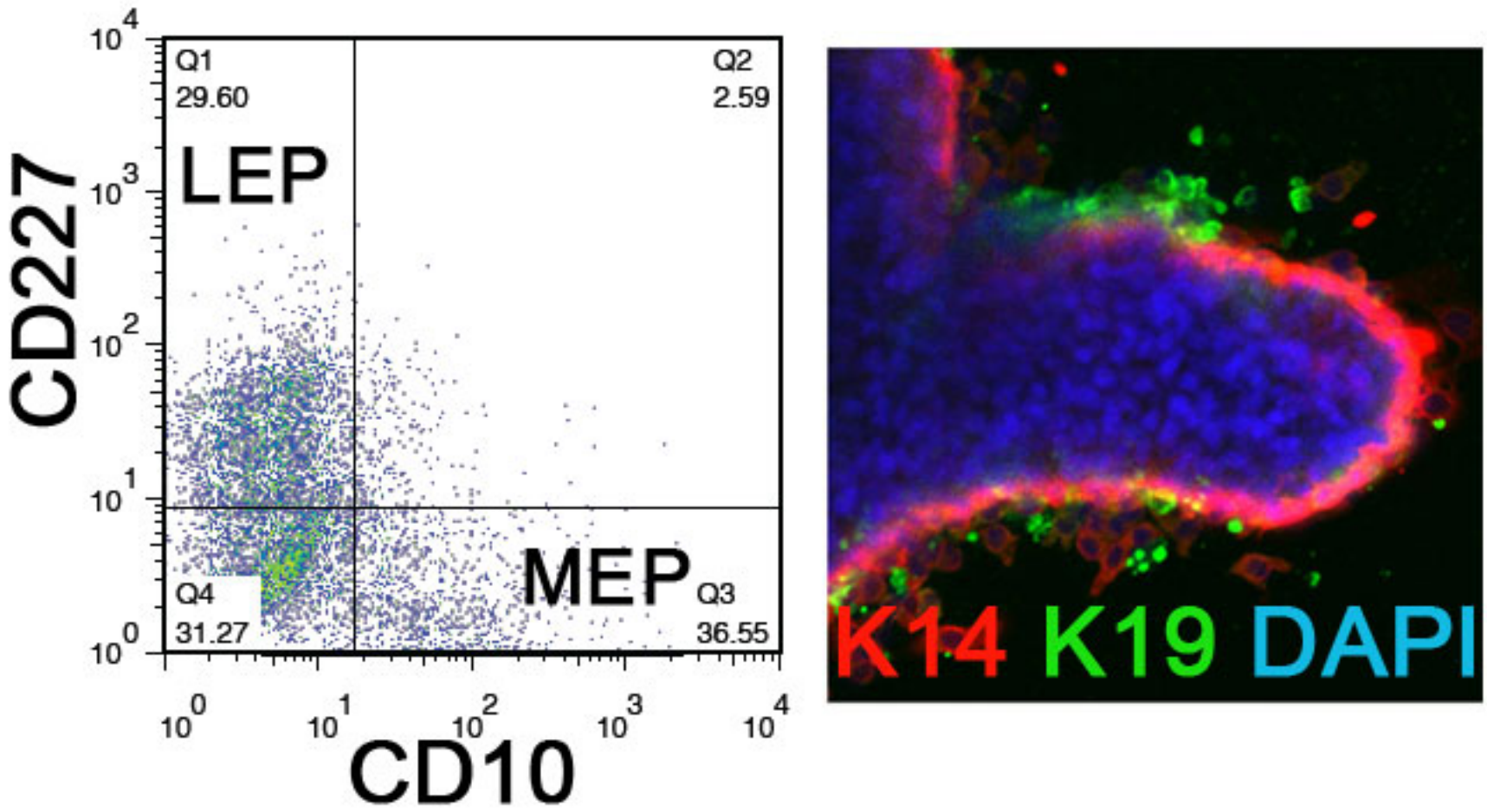
Figure S3 – Lineage marker distributions in cultured and uncultured HMEC. Seven-parameter flow cytometry analyses of 4th passage HMEC strains (A) 240L and (B) 122L, and of dissociated organoid specimens (C) 53 and (D) 29. Forward and side scatter parameters are not shown; CD10, CD227, CD49f, EpCAM, and CD117/cKit expression parameter plots are shown. Gates demarcating the regions that correspond to CD10^{-}/CD227^{+} LEP, CD10^{+}/CD227^{-} MEP, and cKit^{+} HMEC were determined using unstained controls (gray-colored shade boxes). For each strain or specimen, CD10/CD227 and CD227/cKit expression profiles are shown as pseudocolor heatmaps with the LEP, MEP, and cKit regions identified. Multicolor overlays show the EpCam/CD49f expression files of the cells that fall within the LEP (green), MEP (red), and cKit (orange) gates. In the multicolor overlay plots, all the cells that are do not fall within the gated regions appear black.

Figure S1 Garbe etal

A



B



C

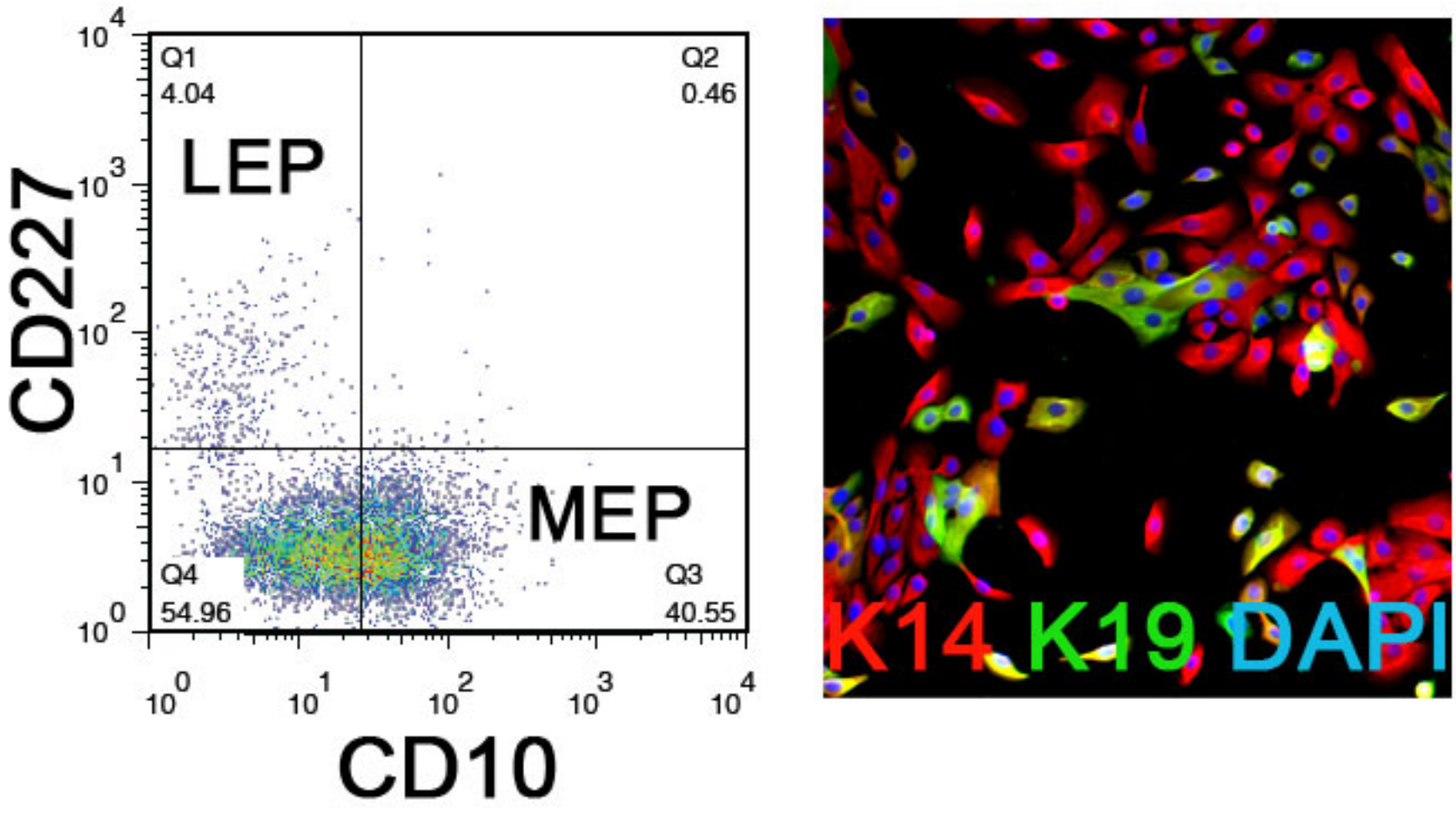
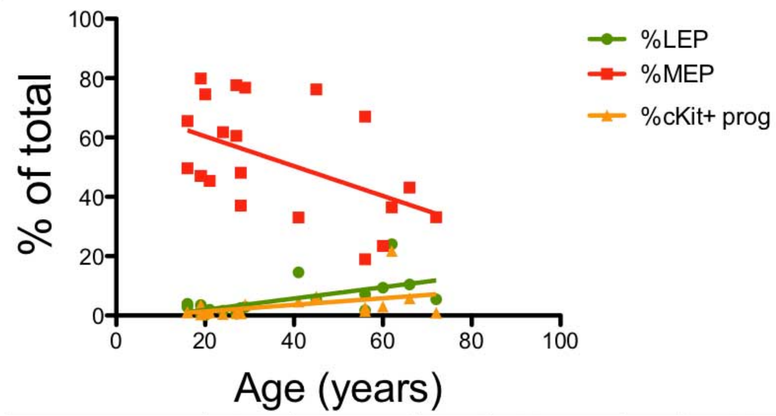


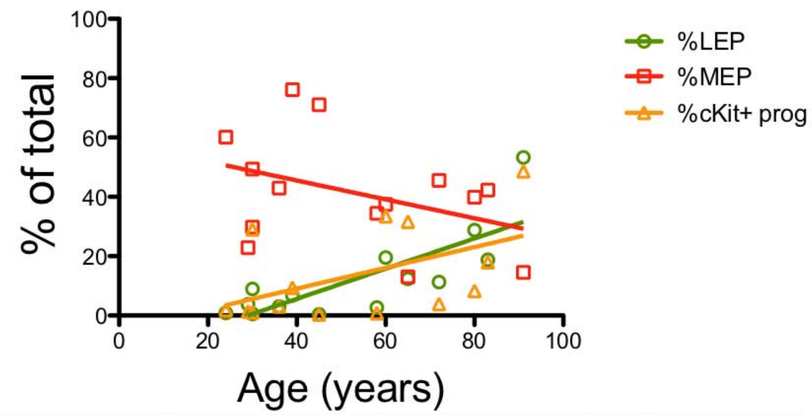
Figure S2 Garbe etal

A



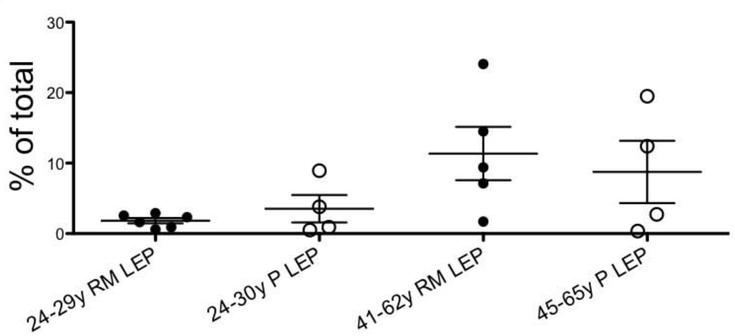
	CD227+ LEP	CD10+ MEP	cKit+ PROG
1/slope	5.256	-1.993	9.005
R^2	0.385	0.251	0.199
P value	0.0027	0.0205	0.0421
Deviation from zero	Significant	Significant	Significant

B

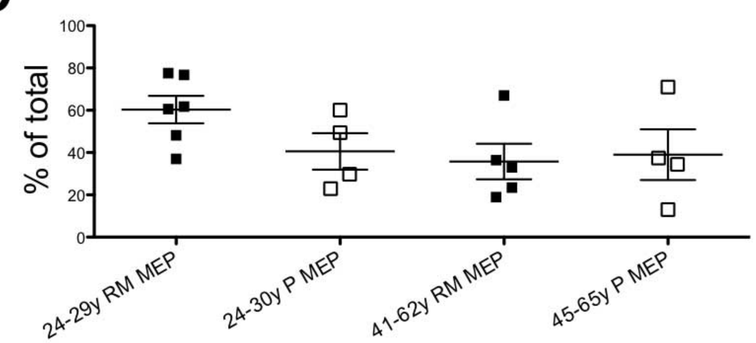


	CD227+ LEP	CD10+ MEP	cKit+ PROG
1/slope	2.020	-3.354	2.797
R^2	0.6058	0.1302	0.2629
P value	0.0006	0.1863	0.0507
Deviation from zero	Significant	Not significant	Not significant

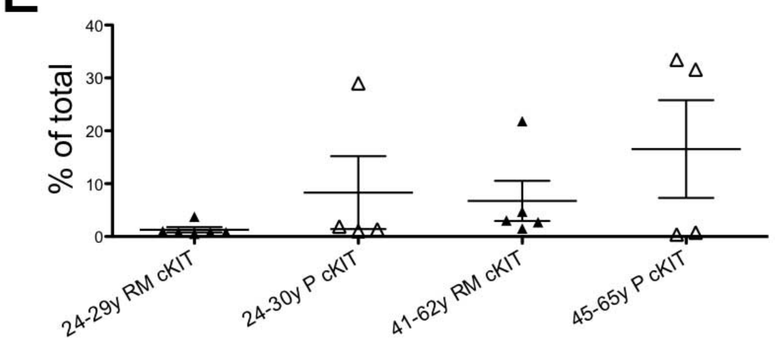
C



D



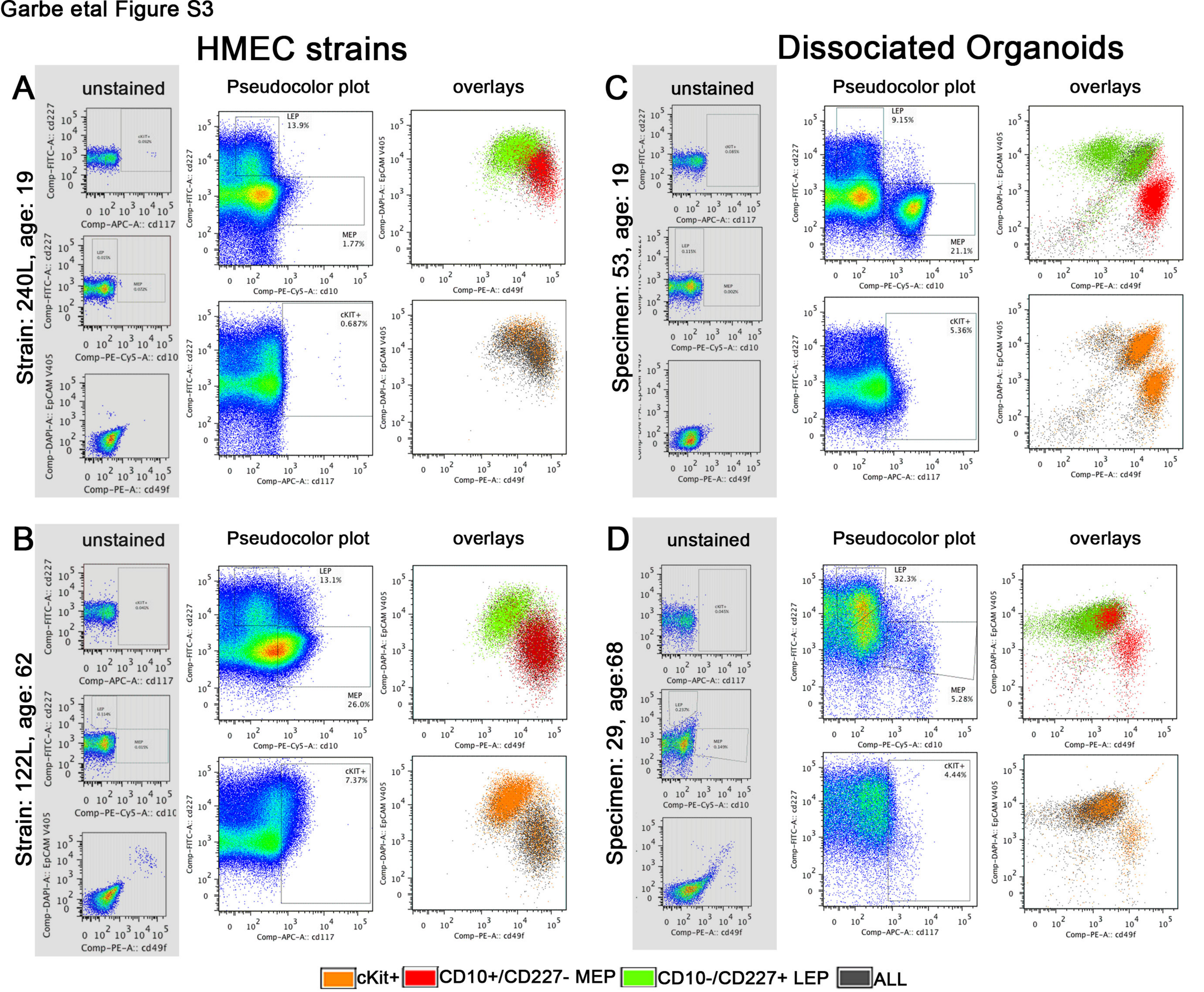
E



F

Table Analyzed	GROUPED analyses				
One-way analysis of variance					
P value	< 0.0001				
P value summary	***				
Are means signif. different? (P < 0.05)	Yes				
Number of groups	12				
F	11.42				
R square	0.7362				
ANOVA Table	SS	df	MS		
Treatment (between columns)	21632	11	1967		
Residual (within columns)	7750	45	172.2		
Total	29382	56			
Bonferroni's Multiple Comparison Test	Mean Diff.	t	Significant? P < 0.05?	Summary	95% CI of diff
24-30y P LEP vs 24-29y RM LEP	1.703	0.2011	No	ns	-21.68 to 25.08
24-30y P MEP vs 24-29y RM MEP	-19.76	2.333	No	ns	-43.14 to 3.620
24-30y P cKIT vs 24-29y RM cKIT	7.030	0.8299	No	ns	-16.35 to 30.41
45-65y P LEP vs 41-62y RM LEP	-2.610	0.2964	No	ns	-26.91 to 21.69
45-65y P MEP vs 41-62y RM MEP	3.200	0.3634	No	ns	-21.10 to 27.50
45-65y P cKIT vs 41-62y RM cKIT	9.789	1.112	No	ns	-14.51 to 34.09

Garbe etal Figure S3



DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.