Contrasting roles of different mismatch repair proteins in basal-like breast cancer

Jiao Mo[1,2,†], Nicholas Borcherding[3,4,†], Sung Jo[3,5,†], Tanzia Islam Tithi[1], Edward Cho[3,6], Kailey E Cash[7], Masayoshi Honda[7], Kawther K. Ahmed[3,8], Ryan Kolb[1], Ronald Weigel[6], Maria Spies[7], Weizhou Zhang[1,9 *]

1. Department of Pathology, Immunology and Laboratory Medicine, College of Medicine, University of Florida, Gainesville, FL 32610, USA;
2. Current: R & D, Thermo Fisher Scientific, Alachua, FL 32615, USA;
3. Department of Pathology, the University of Iowa Carver College of Medicine, Iowa City, IA, 52242
4. Current: Department of Pathology and Immunology, Washington University School of Medicine, St Louis, MO 63110, USA. Electronic address: borcherding.n@wustl.edu.
5. Current: R & D, Carbon Biosciences, Waltham, MA 02451
6. Department of Surgery, the University of Iowa Carver College of Medicine, Iowa City, IA, 52242
7. Department of Biochemistry and Molecular Biology, the University of Iowa Carver College of Medicine, Iowa City, IA, 52242
8. Current: Department of Pharmaceutics, the University of Baghdad College of Pharmacy, Bab-almoadham, PO Box 14026, Baghdad, Iraq
9. University of Florida Health Cancer Center (UFHCC), the University of Florida, Gainesville, FL 32610, USA;

[†] These authors contributed equally.

*Correspondence: zhangw@ufl.edu

Abstract

The Mismatch repair (MMR) pathway is known as a tumor suppressive pathway and genes involved in MMR are commonly mutated in hereditary colorectal or other cancer types. However, the function of MMR genes/proteins in breast cancer progression and metastasis are largely unknown. We found that MSH2, but not MLH1, is highly enriched in basal-like breast cancer (BLBC) and that its protein expression is inversely correlated with overall survival time (OS). *MSH2* expression is frequently elevated due to genomic amplification or gain-of-expression in BLBC, which results in increased MSH2 protein that pairs with MSH6 (collectively referred to as MutSα). Genetic deletion of MSH2 or MLH1 results in a contrasting phenotype in metastasis, with MSH2-deletion leading to reduced metastasis and MLH1-deletion to enhanced liver or lung metastasis. Mechanistically, MSH2-deletion induces the expression of a panel of chemokines in BLBC via epigenetic and/or transcriptional regulation, which leads to an immune reactive tumor microenvironment (TME) and elevated immune cell infiltrations. MLH1 is not correlated with chemokine expression and/or immune cell infiltration in BLBC, but its deletion results in strong accumulation of neutrophils that are known for metastasis promotion. Our study supports the differential functions of MSH2 and MLH1 in BLBC progression and metastasis, which challenges the paradigm of the MMR pathway as a universal tumor suppressive mechanism.

Introduction

Basal-like breast cancer (BLBC) is an aggressive molecular subtype of breast cancer, accounting for 12.3-36.7% of invasive ductal carcinomas in different patient cohorts[Toft and Cryns, 2011]. The absence of approved targeted therapy and the variable responses to chemotherapy in BLBC patients represent vital and unmet clinical need. BLBC is also defined by high levels of genomic instability, with p53 pathway alterations seen in 85-95% of tumors[Toft and Cryns, 2011]. This instability in BLBC is compounded by the characteristic loss in chromosome 5q, consisting of DNA repair genes, *RAD17*, *RAD50*, *MSH3*, and X*RCC4[Rakha et al., 2008].* In addition,15-20% of BLBC possess mutations in *BRCA1* or *BRCA2,* which function in double-stranded DNA repair*.* Half of the 15-20% of BLBC tumors with *BRCA1* or *BRCA2* are a result of germline mutations and are associated with a 70% lifetime risk of breast or ovarian cancer to individuals[Rakha et al., 2008]. Recent success of clinical trials using inhibitors for Poly-ADP-ribose polymerase 1 (PARP1), a single-strand DNA repair sensor, in germline *BRCA*-mutated ovarian and breast cancers have led to Food and Drug Administration approval[Zimmer et al., 2018]. PARP1 plays multiple role in genome maintenance, signaling and repair of DNA damage and stabilization of stalled DNA replication forks [Lord and Ashworth, 2017; Ray Chaudhuri and Nussenzweig, 2017]. Its pharmacological inhibition leads to cell death in cells carrying biallelic BRCA mutations [Ashworth, 2008; Hengel et al., 2017; Zimmer et al., 2018]. Despite the positive successes of PARP inhibitors in trials, the translation of PARP therapy to the majority of BLBC and ovarian cancers with intact *BRCA1/2* have been unsuccessful. Even in patients with BRCA-mutated breast cancers, 5-year survival with Olaparib is about 15%, and that the rest develop resistance [Incorvaia et al., 2017; Ledermann et al., 2016; Lupo and Trusolino, 2014; Murata et al., 2016].

The success of PARP inhibitors for a subset of BLBC/TNBC patients has motivated the field to investigate the potential of other DNA repair-based targeted therapy. Mismatch repair (MMR)

genes have traditionally been thought of as tumor suppressors, with germline mutations identified in the autosomal dominant Lynch Syndrome [Bronner et al., 1994; Fishel, 2015; Fishel et al., 1993; Leach et al., 1993]. Closely associated with hereditary non-polyposis colorectal cancer, Lynch syndrome also increases risk in the development of cancers along the gastrointestinal tract, endometrial cancer, and ovarian cancer[Modrich, 1994]. Defective MMR (dMMR) genes, including MLH1, MSH2, MSH6 and PMS2, lead to the increase in mutations and are commonly associated with microsatellite instability (MSI)[Leach et al., 1993; Parsons et al., 1993]. This increased mutational rate is thought to underlie the 25-80% therapeutic response rate of anti-PD-1 immune checkpoint inhibitor (ICI) in patients with Lynch-Syndrome-associated tumors[Geoerger et al., 2020; Le et al., 2020; Le et al., 2015; Marabelle et al., 2020]. The result of this trial led to the unprecedented approval of anti-PD-1 therapies for metastatic dMMR or MSI cancers, irrespective of the site of origin. Interestingly, patients with MSI or dMMR exhibit very different responses to ICIs, depending on the extend of insertion-deletion mutations[Mandal et al., 2019] or the status of cGAS-STING-dependent DNA sensing and interferon β (IFN-β) activation [Guan et al., 2021; Lu et al., 2021], leading to the recruitment of $CD8^+$ T cells by CCL5 and CXCL10 production [Mowat et al., 2021]. Though most of these studies are based on genetic silencing or knockout of key MMR genes *MLH1* or *MSH2*, those two act epistatically and are commonly considered equal and have similar impacts on cancer progression or immune regulations via the MMR pathway.

Despite acting epistatically in the mismatch recognition, and performing tumor suppressor functions, we found that MSH2 and MLH1 have contrasting roles in the metastasis of BLBC. MSH2 protein is associated with decreased immune infiltration and increased metastasis; whereas MLH1 suppresses metastasis. MSH2-deletion promotes immune cell infiltrations via inducing a systematic increase in chemokines by releasing the epigenetic suppression pathways. This study identifies the opposite roles of different MMR proteins in cancer

progression and metastasis and establishes the rationale to target MSH2 or MutSα for BLBC therapy.

Results

MSH2 and MSH6 are elevated in BLBC and predict poor overall survival.

We previously developed analysis workflow to identify survival outcomes using reverse-phase protein array (RPPA)[Li et al., 2013b] quantification available in the Cancer Genome Atlas (TCGA)[Borcherding et al., 2018]. We applied the system to examine prognostic markers for overall survival in BLBC (Fig 1A), with a focus on DNA repair proteins available on the RPPA (highlighted in red). Of the 6 DNA repair proteins that had significant ($P < 0.05$) prognostic value, elevated expression of MSH2, RAD50, KU80, CHK2, and MSH6 predicted poor overall survival with Cox proportional hazard ratios greater than 1 (Fig. 1B). The division of samples in high versus low was based on maximal standardized two-sample linear rank statistic to find an optimal cutpoint, providing a range proportional comparison (Fig. 1C). Notably, MSH2 evenly split at 50% of samples (Fig. 1C) and predicted the worse survival for any DNA repair protein (Fig 1D, upper panel, hazard ratio=7.08). MSH6, which forms a heterodimer with MSH2, had a different optimal cutpoint at 21.8% MSH6-high versus 78.2% MSH6-low, but still predicted poor overall survival (Fig. 1D, lower panel, hazard ratio=3.26). Since the TCPA dataset includes limited proteins, we also expanded the analysis using the TCGA RNA sequencing data where we found that *MLH1* and *PMS1* (collectively referred to as MutLα) expression predicts better prognosis with HR of 0.65 and 0.42 respectively (Supplementary Table 1). Other poor prognosis predictors in the DNA repair pathways include *MSH3* in the MMR pathways, as well as *MPG* and *NTHL1* in the base excision repair (BER) pathway (Supplementary Table 1). We also performed Cox regressions in the other molecular subtypes and found that there is no significant predictive value for MSH2 and MSH6 in HER2+ or Luminal A breast cancers

(Supplementary Fig. 1A,B). Interestingly, in Luminal B tumors, MSH2 protein level predicted improved overall survival (Supplementary Fig. 1C). Across breast cancer samples, MSH2 and MSH6 protein levels were directly correlated with one another (Fig. 1E), supporting the established literature on the need of heterodimer formation for MSH6 stabilization[Arlow et al., 2021]. Accompanying the survival predictions, we also found MSH2 (Fig. 1F) and MSH6 (Fig.1G) are significantly elevated in BLBC compared to the other breast cancer molecular subtypes. MSH3 pairs with MSH2 to form MutSβ, but MSH3 is excluded in these bioinformatics analysis because gene locus of MSH3 is frequently deleted in upto 70% TNBC/BLBC [Johannsdottir et al., 2006; Natrajan et al., 2009; Turner et al., 2010].

The contrasting roles of MSH2 and MLH1 in breast cancer metastasis.

The differential impacts of MSH2 and MLH1 on BLBC prognosis indicate the potential different functions of them in BLBC progression. We further expanded the analysis among the TCGA pan-cancer datasets and found that *MSH2* is frequently amplified with the gain of mRNA expression in breast cancer (BRCA), lung cancers (LUAD and LUCA), ovarian cancer (OV) *etc* (Supplementary Fig. 2A). MLH1, in contrast, exhibits frequent mutations and deletions but with very limited amplification or gain of expression (Supplementary Fig. 2B), further supporting the potentially different functions of MSH2 and MLH1 among different cancer types. To determine the roles of MSH2 and MLH1 in BLBC, we used protein-gRNA (RNP) transfection-based CRISPR/Cas9 system to knock out their expression in several BLBC cells, including murine 4T1 (Supplementary Fig. 3A-B) and Py8119 (Supplementary Fig. 3C-D) breast cancer cells, murine MC38 colon cancer cells (Supplementary Fig.4A-B), human MDA-MD-231 (Supplementary Fig. 4C-D) basal B breast cancer cells. All single clones were validated for the deletion of target genes and several clones were combined (at least 5 knockout clones, referred to as either control (Con) or knockout (KO)) for each cell type (Supplementary Fig. 3-4) for tumor and metastasis studies (Fig. 2, Supplementary Fig. 5). MSH2 KO did not significantly influence

primary tumor growth from Py8119 (Fig. 2A), 4T1 (Fig. 2C), MC38 (Supplementary Fig. 5A) or MDA-MB-231 (Supplementary Fig. 5B), but led to a reduction of lung metastasis in Py8119 (Fig. 2B, P = 0.009) or 4T1 (Fig. 2D, 3/7 mice with 4-5 metastatic nodules in control group while no mice in the MSH2 KO group) orthotopic models, or lung and liver experimental metastasis via *i.v.* injection of Py8119 cells in immune competent C57BL/6J (Py8119) or Balb/C mice (4T1 model) (Fig. 2E-F). MLH1 KO, however, led to an increase in primary tumor growth and metastasis in the Py8119 and 4T1 models (Fig. 2B, 2D), but not in the MC38 colon cancer model (Supplementary Fig. 5A) . The *i.v.* experimental metastasis model further validates the significant increase in metastasis due to *MLH1* KO, indicating that tumor size increase may not be the primary reason of increased metastasis due to MLH1 deficiency. It is worth noting that the MSH2-mediated metastasis relies on the intact immune system as *i.v.* injection of Py8119 cells into the NOD/Scid/IL-2R$\gamma^{-/-}$ (NSG) mice – lacking functional T, B and nature killer (NK) cells – had equal amount of metastasis between control and MSH2 KO groups (Fig. 2G). When injecting the human MDA-MB-231 cells into NSG mice, MLH1 KO also led to increase in tumor growth and tendency of shorter overall survival of the mice (Supplementary Fig. 5B-C), suggesting that the MLH1-mediated tumor/metastasis-suppressing phenotypes are likely via cancer-cell intrinsic pathways and are independent of T, B or NK cells.

The differential roles of MSH2 and MLH1 in tumor-infiltrating (TI) immune cells in BLBC

As deletion of cancer cell-intrinsic MSH2 or MLH1 has opposite phenotypes in metastasis and that MSH2-dependent tumor/metastasis promotion relies on an intact immune system, we determined the immune cell compositions as we established before[Kim et al., 2021; Kolb et al., 2021], including tumors or spleens (Supplementary Fig. 6, Fig. 3) of the Py8119-tumor bearing mice (Fig.2A). MSH2 KO in Py8119 cells did not significantly change the peripheral immune profiles (spleen) in general (Fig. 3A, 3C, 3E, 3G, 3I, 3K, 3M), except for the slight decrease in dendritic cells (DC, CD11c^{+}MHCII^{+}) and active DCs (CD80^{+}) (Fig. 3I), or monocytes

($CD11b^+Ly6C^+Ly6G^-$) (Fig. 3M). In contrast, MSH2 KO led to significantly increase in majority of TI immune cells, including total leukocyte or lymphocytes (Fig. 3B), NK ($CD3^-NK1.1^+$, active NK (Granzyme B^+), NKT ($CD3^+NK1.1^+$) (Fig. 3D), conventional $CD4^+$ T cells (Tconv, $CD4^+Foxp3^-$), regulatory T cells (Treg, $CD4^+CD25^+Foxp3^+$) (Fig. 3F), $CD8^+$ and active $CD8^+$ (Granzyme B^+/$Perforin^+$) T cells (Fig. 3H), neutrophils (Fig. 3L) or monocytes (Fig. 3N), without significant changes in tumor associated macrophages (TAM, Fig. 3N).

MLH1 KO, interestingly, has an impact more on peripheral immune profiles than on TI immune profiles (Fig. 3A, Fig. 3K), with a marked decrease in peripheral lymphocytes (Fig. 3A) and a dramatic increase in neutrophils (Fig. 3K, an increase from 9% in the control group to 21% in the KO group, P=0.003) and monocytes (Fig. 3N, P= 0.003). Other altered cells in the periphery included a slight increase in NKT cells (Fig. 3C) and minor decrease in active $CD8^+$ T cells (Fig. 3G). In the TME, MLH1 KO led to significant decreases in total leukocytes and lymphocytes (Fig. 3B), DCs, active DCs (Fig. 3J), and neutrophils (Fig. 3L), without significant influences on NK, T, or TAM subpopulations (Fig. 3D, 3F, 3H). These immunological changes induced by deletion of cancer cell intrinsic MSH2 or MLH1 agree with the tumor/metastasis phenotypes observed in BLBC tumors (Fig. 2).

MSH2 KO leads to extensive expression of chemokine/cytokines and immune activation in BLBC

To understand the differential roles of MSH2 and MLH1 in immune modulation, we performed RNA sequencing of total RNAs from tumors in Fig. 2a and Fig. 2c. Gene Set Enrichment Analysis (GSEA) identified that the chemokine pathways are the only consistent pathways that are enriched in MSH2 KO Py8119 (Fig. 4A) and 4T1 (Fig. 4B) tumors, relative to both Con and MLH1 KO tumors (Fig. 4C), in agreement with the general increase in TI immune cell populations (Fig. 3). Other pathways involved in immune modulations include a panel of MHC

molecules in the Py8119 MSH2 KO tumors (Fig. 4D), in agreement with immune active phenotype and T cell recognition. MSH2 KO in BT549 cells led to a similar MHC upregulation, indicating an activation of antigen presentation pathways via cancer-intrinsic mechanisms (Supplementary Fig. 7). We collected conditioned medium from WT 4T1 cells, or MSH2 KO or MLH1 KO 4T1 cells, and found that several cytokine/chemokine proteins were also selectively elevated in the MSH2 KO cells relative to WT and MLH1 KO cells (Fig. 4E, Supplementary Fig. 8), suggesting the cancer-cell intrinsic changes in chemokine/cytokine production mediated by MSH2.

Datamining the TCGA dataset, we found that MSH2 protein is significantly and inversely correlated with chemokines such as CCL3, CCL4, CCL5 (Fig. 4F), MHC molecules (Fig. 4G-H) and markers for monocytes, T cells or B cells (Fig. 4I) among the human BLBC patient specimens. In addition, we separated human BLBC specimens by MSH2 protein level (high versus low) and estimated the immune profiles using ESTIMATE and CIBERSORT algorithms. Consistent with the correlation studies (Fig. 4I, we found increased estimated immune cell infiltration (Fig. 4J), with lineage specific increases in B cells (P=0.0084), monocytes/macrophages (P=0.021), and NK cells (P=0.04), with CD8 T cells approaching significance (P=0.062) (Fig. 4K). GSEA analysis further revealed that MSH2-low BLBC specimens have significantly enriched signaling pathways involving in immune regulations (Supplementary Table 2); whereas MSH2-high BLBC specimens have elevated pathways including RNA processing, damaged DNA binding, microtubule organization, DNA strand elongation, chromosomal segregation *etc*. that are important for cancer cell proliferation and progression (Supplementary Table 3). In contrast, MLH1-low BLBC specimens have elevated pathways involving in ribosomal functions, translation, mitochondrial functions *etc* (Supplementary Table 4, Supplementary Fig.9A-D), indicating more active and aggressive cancers, which are in line with the results shown in Fig. 2. MLH1-high BLBC specimens did not exhibit any distinct pathways related to cancer progression (Supplementary Table 5).

We performed differential gene expression compared MSH2-high (n=60, red) to MSH2-low (n=59, black) using estimated counts (Supplementary Fig. 10A). As expected, we found MSH2-high associated with increased RNA levels of *MSH2* and *MSH6*, in addition to increased *PMS1*, which participates in the repair downstream of the MutSα heterodimer (Supplementary Fig. 10B). MSH2-low BLBC samples had increased levels of immune checkpoints, *IDO1*, *CD274*, and *HAVCR2* (Supplementary Fig. 10C), indicating an immune active but exhausted tumor microenvironment. This increased levels of immune markers in MSH2-low samples was seen across chemokines/cytokines (Supplementary Fig. 10D) and chemokine/cytokine receptors (Supplementary Fig. 10E). We also performed Ingenuity Pathway Analysis (IPA) for upstream regulators of the transcriptional differences observed (Supplementary Fig. 10F). Quadrant I refers to upstream regulators that have increased RNA expression and activation in MSH2-high BLBC samples, while quadrant III refers to regulators with increased expression and activation in MSH2-low samples. In MSH2-low BLBC samples, we observed a general increase in cytokine/chemokine signaling, including the IL-1, IFN-γ, IL5, WNT5A, LIF, CXCL12 etc (Supplementary Fig. 10F, quadrant III and bar graph).

Long-term MSH2 KO leads to exhaustion and sensitizes BLBC to immunotherapy

To study the passage-dependent effect of MSH2 in BLBC, we used LentiCRISPRV2 system to knockout MSH2 within MDA-MB-231 cells and culture them up to 30th passages. We confirmed MSH2 KO in these cells even after 30th passage (Supplementary Fig. 11) and performed RNAseq of earlier passages (P5) and late passages (P30) (Fig. 5A). Interestingly, MSH2 KO in the basal B MDA-MB-231 cells led to a passage-dependent shift from an early immune activation at P5 – evidenced by elevation of CD40, CD27, ICOSLG etc – to an exhaustion phenotype supported by elevated expression of immune checkpoints PVR, PDCD1LG2 (PD-L2) and CD274 (PD-L1) (Fig. 5A), in agreement with the human BLBC data showing reverse correlation of MSH2 with PVR and CD274 (Supplementary Fig. 10C). Using the 4T1 BLBC

model, we established a cell lysate vaccine model to boost the immune activation within 4T1 tumors (Fig. 5B), otherwise known as an unresponsive tumor model to immune checkpoint inhibitors [Sagiv-Barfi et al., 2015]. We found that after vaccination, *Msh2*-KO led to an improved survival (Fig. 5C) and reduced tumor growth (Fig. 5D), suggesting vaccination sensitizes *Msh2*-KO effect even in the primary tumors. Vaccination + anti-PD-1 improved survival (Fig. 5C) and inhibited tumor growth (Fig. 5D) in the WT 4T1 tumor bearing mice, but further extended survival and reduced tumor growth in *Msh2*-KO tumor bearing mice (Fig. 5C,5D). Flow cytometric analysis supported that *Msh2*-KO led to increased PD-1^+TIM-3^+ exhausted T cells and anti-PD-1 treatment reduced that population (Fig. 5E-F).

Long-term MSH2 KO leads to epigenetic activation of immune-related gene expression

Since long-term MSH2 KO leads to significant changes in several immune relevant genes including chemokines/cytokines, MHC (Fig. 4) and immune checkpoints (Fig. 5), we reason that those long-term gene regulation could likely be due to epigenetic regulations via interactions with DNA methyltransferase (DNMT) as reported before [Ding et al., 2016; Ding et al., 2019; Li et al., 2013a]. We performed bioinformatics analysis of the DNA methylation status on the *PVR* or *CD274* (PD-L1) locus in human BLBC specimens deposited in the TCGA and found that expression of PVR or CD274 was negatively correlated with the promoter methylation status (Fig. 6A), indicative of an epigenetic mechanism. Using real-time PCR, we confirmed that late passage of MSH2-KO cells (P30) exhibit increased PVR (2.42 fold) and CD274/PD-L1 (3.03 fold) expression relative to control cells (Fig. 6B, P values shown between Con and P30). Five days of 5-Azacytidine treatment – an FDA-approved DNA methyltransferase (DNMT) inhibitor, induced significant increase in *PVR* and *CD274* expression in the control cells (Fig. 6B). 5-Azacytidine further induced *CD274* expression in the MSH2-KO cells (Fig. 6B). Another level of gene regulation of exhaustion markers comes from transcriptional regulation by the interferon (IFN) pathway. We did find that IFNγ was able to induce the strong expression of *PVR* or *CD274*

mRNA in the MSH2-KO cells, to a much lesser extend in the control cells (Fig. 6C-D). At the mechanistic level, we performed DNMT1 activity assay as illustrated (Fig. 6E). We found that addition of MutSα to the enzymatic mix enhanced DNMT1 activity in CpG methylation (Fig. 6F-G). Replication foci targeting sequence (RFTS) is part of an N-terminal domain from DNMT1 and can specifically inhibit DNA-binding and catalytic activity of DNMT1. We found that the RFTS domain inhibited DNMT1 activity with the presence MutSα, supporting that MutSα specifically enhances the activity of DNMT1 (Fig. 6H-I). These results support that MSH2 is involved in epigenetic regulation of these genes via DNA methylation.

Discussion

Known as a tumor suppressive pathway, MMR genes are mostly considered equal related to cancer initiation, hypermutator phenotype, high level of mutational load and the sensitivity to ICIs. Based on several important trials[Geoerger et al., 2020; Le et al., 2020; Le et al., 2015; Marabelle et al., 2020], FDA approved pembrolizumab for first-line treatment of patients with microsatellite instability-high (MSI-H) or mismatch repair deficient (dMMR) cancers, irrelevant of cancer types. Interestingly, we found that *MSH2* expression is frequently elevated in human BLBC and several other cancer types and its protein expression is correlated with worse prognosis in human BLBC. Mouse models also support the tumor/metastasis-promoting function of MSH2, in contrast to the role of MLH1 in tumor/metastasis suppression. This is the first report showing opposite functions of MMR genes in metastasis.

Methods

Reverse-phase Protein Analysis

RPPA data was downloaded from the TCPA Portal located at http://tcpaportal.org/. Data was processed as previously described and clinical data was attached by Patient/Sample

ID[Borcherding et al., 2018]. Protein-based survival analysis was performed using the previously-developed TRGAted R Shiny application, code and processed data for all TCGA cohorts are located at https://github.com/ncborcherding/TRGAted. [Borcherding et al., 2018] Molecular-subtype designation was based on the TCGA Analysis Working Group pipeline, fitting PAM50 subtypes based on RNA-seq data. Protein versus mRNA correlations was performed using log2(x+1) mRNA quantification downloaded from the cBioPortal [Cerami et al., 2012; Gao et al., 2013]. Correlations for both protein-mRNA and protein-protein comparisons utilized the rank-based Spearman approach.

Differential Gene and Pathway Analysis

Gene-level HTSeq count data was downloaded from the UCSC Xena Browser at http://xena.ucsc.edu/. Using the DESeq2 R package (v1.16.1), count data were separated in half by MSH2 RPPA quantification and converted into negative binomial distributions. A parametric model was fitted for the data and significance and log2 fold-change was computed using Wald testing. P-values for the differential analysis was corrected for multiple hypothesis testing using the Benjamini and Hochberg method. Significance threshold was set at adjust P-value ≤ 0.05 and log2-fold change ≥ |0.5|. A total of 2,094 genes met the criteria, with 650 genes significantly increased in MSH2-high BLBC samples and 1,444 genes significantly decreased. For visualization of the expression of specific genes by MSH2 grouping, the expression matrix was regularized log transformed. Ingenuity Pathway Analysis was performed using the above significance cut-off. DESeq-based log2-fold change was incorporated with the bias-corrected Z-score for upstream regulator analysis in order to visualized activation relative to mRNA fold-change.

RNA-immune cell estimates

Immune and stromal ESTIMATE scores for all breast cancer samples in the TCGA were downloaded (http://bioinformatics.mdanderson.org/estimate/). BLBC samples with MSH2 protein quantifications were extracted and split by mean MSH2 value. In addition to general immune cell difference, log2 RSEM RNA values for BLBC samples were imported into the CIBERSORT portal, in order to estimate the contribution of individual immune lineages. Absolute estimates of immune cell populations were combined for total immune cell lineages and then were then divided by mean MSH2 protein values.

Cell lines and cell culture

MC38 mouse colon carcinoma cells (Kerafast Inc., Boston, MA) were maintained in DMEM supplemented with 10% FBS, 1mM glutamine, 0.1M non-essential amino acids (Thermo Fisher, Waltham, MA), 1 mM sodium pyruvate, 10 mM HEPES, 100 U/ml penicillin and 100 µg/ml streptomycin. 4T1 mouse mammary carcinoma cells were purchased from American Type Culture Collection (ATCC) and cultured in RPMI medium supplemented with 10% fetal bovine serum (FBS) 100 U/ml penicillin and 100 µg/ml streptomycin. Py8119 mouse mammary carcinoma cells were purchased from ATCC and cultured in DMEM supplemented with 10% fetal bovine serum (FBS), 100 U/ml penicillin and 100 µg/ml streptomycin. All cells were cultured at 37 °C and 5% CO2 in a humidified incubator.

CRISPR/CAS9-mediated knockout and viability assay

The knockout clones of MC38, Py8119 and 4T1 cells were generated by transient transfection of Cas9 protein-guide RNA complex into parental cells, followed by single clone section and immunoblot to confirm KO of MSH2 or MLH1 following the instruction (Synthego). The knockout clones MDA-MB-231 and BT549 cells were constructed either by transient transfection of Cas9 protein-guide RNA complex or by lentiviral particles co-expressing Cas9 and guide RNA

(LentiCRISPRv2), following by puromycin selection for stable pooled culture. Cells were passages every 3 days and passage numbers were recorded.

Animal studies

All animal studies were approved by the University of Florida Institutional Animal Care and Use Committee (IACUC) under protocol 202110399. All animals were housed in a pathogen-free Association for Assessment and Accreditation of Laboratory Animal Care accredited facility at the University of Florida and performed in accordance with IACUC guidelines. The room temperature is 21.1-23.3 °C with an acceptable daily fluctuation of 2°C. Typically the room is 22.2 °C all the time. The humidity set point is 50% but can vary +/-15% daily depending on the weather. The photoperiod is 12:12 and the light intensity range is 15.6-25.8 FC. C57BL/6 and Balb/C mice were purchased from Charles River Laboratories (San Diego, CA). NOD-SCID interleukin-2 receptor gamma null (NSG) mice were purchased from Jackson Laboratories (Bar Harbor, ME). For syngeneic tumor models, 500 cells (for Py8119) were resuspended in 50% phosphate buffered saline (PBS) + 50% Matrigel (Corning Inc., Corning, NY) and implanted into 6-8 week old C57BL/6 (MC38, *s.c*; and Py8119, intra mammary fatpad), Balb/C (4T1, intra mammary fatpad), or NSG mice (MDA-MB-231, intra mammary fatpad). Tumor growth was monitored daily and measured twice weekly with calipers and volume was determined using the formula ½ (L x W2).

Flow cytometry

Mouse tumors were excised and approximately 200 mg of tumor tissue were enzymatically and mechanically digested using the mouse Tumor Dissociation Kit (Miltenyi Biotec) to obtain a single cell suspension. Red blood cells were lysed using ACK lysis buffer and mononuclear cells were isolated by density gradient using SepMate Tubes (StemCell Technologies) and

Lymphoprep density gradient media (StemCell Technologies). Mouse cells were then washed and incubated with combinations of the following antibodies: anti-mouse CD62L-BV785 (clone MEL-14), anti-mouse MHCII I-A/I-E-BB515 (Clone 2G9, BD Biosciences), anti-mouse CD11B-PEdazzle (clone M1/70), anti-mouse CD45-AF532 (clone 30F.11), anti-mouse CD3-APC (clone 17A2), anti-mouse CD8-BV510 (clone 53-6.7), anti-mouse CD4-BV605 (clone GK1.5), anti-mouse NK1.1-AF700 (clone PK136), anti-mouse CD69-SB436 (clone H1.2F3, eBioscience), anti-mouse CD279 (clone-PerCP-EF710, eBioscience Inc), anti-mouse CD366-PacBlue (clone B8.2c12), anti-mouse CD11C-PE-Cy7 (clone N418), anti-mouse Ly6G-FITC (clone IA8), anti-mouse Ly6C-BV711 (clone HK1.4) anti-mouse F4/80-BV650 (clone BM8), anti-mouse CD80-BV480 (clone 16-10A1, BD Biosciences), anti-mouse CD25-PE-Cy5 (clone PC61) plus FVD-eFluor-780 (eBioscience) and mouse FcR blocker (anti-mouse CD16/CD32, clone 2.4G2, BD Biosciences). After surface staining, cells were fixed and permeabilized using the FOXP3/Transcription Factor Staining Buffer Set (eBioscience). Cells were stained with a combination of the following antibodies: anti-mouse FOXP3-APC (clone FJK-16S, eBioscience), anti-mouse Granzyme B-Pacific Blue (clone GB11), anti-mouse Perforin-PE (clone S16009B), anti-mouse Ki-67-PerCP-Cy5.5 (clone 16A8). Flow cytometry was performed on a 3 laser Cytek Aurora Cytometer (Cytek Biosciences, Fremont, CA) and analyzed using FlowJo software (BD Biosciences). All antibodies are from Biolegend unless otherwise specified.

Immunoblotting

Cells or tumors samples were lysed with RIPA buffer (150 mM NaCl, 5 mM EDTA, 50 mM Tris pH8.0, 1% sodium deoxycholate, 1% NP-40, 0.5% SDS) supplemented with 1mM dithiothreitol and protease inhibitors. Cell lysates were then separated by SDS-PAGE and analyzed by standard western blotting protocol as we previously published[Kolb et al., 2021].

Statistical Analysis

Two sample statistical testing principally utilized Welch's t-test, allowing for unequal variance and sample size, unless otherwise indicated. Multiple comparisons across three or more groups utilized ANOVA with the Tukey honest significant difference adjustment for multiple comparisons.

Conflicts of Interest:

All the authors claim no conflict of interest.

Acknowledgements

The work was mainly supported by DOD/CDMRP grant BC180227 (W.Z.) and BC180227P (M.S.), and partially supported by NIH grants CA200673 (W.Z.), CA203834 (W.Z.), CA260239 (W.Z.), and DOD/CDMRP grant BC200100 (W.Z.). W.Z. was also supported by an endowment fund from the Dr. and Mrs. James Robert Spenser Family.

References

Arlow T, Kim J, Haye-Bertolozzi JE, Martinez CB, Fay C, Zorensky E, Rose MD, Gammie AE: MutSalpha mismatch repair protein stability is governed by subunit interaction, acetylation, and ubiquitination. G3 (Bethesda) 2021;11.

Ashworth A: A synthetic lethal therapeutic approach: poly(ADP) ribose polymerase inhibitors for the treatment of cancers deficient in DNA double-strand break repair. J Clin Oncol 2008;26:3785-3790.

Borcherding N, Bormann NL, Voigt AP, Zhang W: TRGAted: A web tool for survival analysis using protein data in the Cancer Genome Atlas. F1000Res 2018;7:1235.

Bronner CE, Baker SM, Morrison PT, Warren G, Smith LG, Lescoe MK, Kane M, Earabino C, Lipford J, Lindblom A, et al.: Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. Nature 1994;368:258-261.

Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C, Schultz N: The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2012;2:401-404.

Ding N, Bonham EM, Hannon BE, Amick TR, Baylin SB, O'Hagan HM: Mismatch repair proteins recruit DNA methyltransferase 1 to sites of oxidative DNA damage. J Mol Cell Biol 2016;8:244-254.

Ding N, Miller SA, Savant SS, O'Hagan HM: JAK2 regulates mismatch repair protein-mediated epigenetic alterations in response to oxidative damage. Environ Mol Mutagen 2019;60:308-319.

Fishel R: Mismatch repair. J Biol Chem 2015;290:26395-26403.

Fishel R, Lescoe MK, Rao MR, Copeland NG, Jenkins NA, Garber J, Kane M, Kolodner R: The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. Cell 1993;75:1027-1038.

Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N: Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal 2013;6:pl1.

Geoerger B, Kang HJ, Yalon-Oren M, Marshall LV, Vezina C, Pappo A, Laetsch TW, Petrilli AS, Ebinger M, Toporski J, Glade-Bender J, Nicholls W, Fox E, DuBois SG, Macy ME, Cohn SL, Pathiraja K, Diede SJ, Ebbinghaus S, Pinto N: Pembrolizumab in paediatric patients with advanced melanoma or a PD-L1-positive, advanced, relapsed, or refractory solid tumour or lymphoma (KEYNOTE-051): interim analysis of an open-label, single-arm, phase 1-2 trial. Lancet Oncol 2020;21:121-133.

Guan J, Lu C, Jin Q, Lu H, Chen X, Tian L, Zhang Y, Ortega J, Zhang J, Siteni S, Chen M, Gu L, Shay JW, Davis AJ, Chen ZJ, Fu YX, Li GM: MLH1 Deficiency-Triggered DNA Hyperexcision by Exonuclease 1 Activates the cGAS-STING Pathway. Cancer Cell 2021;39:109-121 e105.

Hengel SR, Spies MA, Spies M: Small-Molecule Inhibitors Targeting DNA Repair and DNA Repair Deficiency in Research and Cancer Therapy. Cell Chem Biol 2017;24:1101-1119.

Incorvaia L, Passiglia F, Rizzo S, Galvano A, Listi A, Barraco N, Maragliano R, Calo V, Natoli C, Ciaccio M, Bazan V, Russo A: "Back to a false normality": new intriguing mechanisms of resistance to PARP inhibitors. Oncotarget 2017;8:23891-23904.

Johannsdottir HK, Jonsson G, Johannesdottir G, Agnarsson BA, Eerola H, Arason A, Heikkila P, Egilsson V, Olsson H, Johannsson OT, Nevanlinna H, Borg A, Barkardottir RB: Chromosome 5 imbalance mapping in breast tumors from BRCA1 and BRCA2 mutation carriers and sporadic breast tumors. Int J Cancer 2006;119:1052-1060.

Kim MC, Borcherding N, Ahmed KK, Voigt AP, Vishwakarma A, Kolb R, Kluz PN, Pandey G, De U, Drashansky T, Helm EY, Zhang X, Gibson-Corley KN, Klesney-Tait J, Zhu Y, Lu J, Lu J, Huang X, Xiang H, Cheng J, Wang D, Wang Z, Tang J, Hu J, Wang Z, Liu H, Li M, Zhuang H, Avram D, Zhou D, Bacher R, Zheng SG, Wu X, Zakharia Y, Zhang W: CD177 modulates the function and homeostasis of tumor-infiltrating regulatory T cells. Nat Commun 2021;12:5764.

Kolb R, De U, Khan S, Luo Y, Kim MC, Yu H, Wu C, Mo J, Zhang X, Zhang P, Zhang X, Borcherding N, Koppel D, Fu YX, Zheng SG, Avram D, Zheng G, Zhou D, Zhang W: Proteolysis-targeting chimera against BCL-X(L) destroys tumor-infiltrating regulatory T cells. Nat Commun 2021;12:1281.

Le DT, Kim TW, Van Cutsem E, Geva R, Jager D, Hara H, Burge M, O'Neil B, Kavan P, Yoshino T, Guimbaud R, Taniguchi H, Elez E, Al-Batran SE, Boland PM, Crocenzi T, Atreya CE, Cui Y, Dai T, Marinello P, Diaz LA, Jr., Andre T: Phase II Open-Label Study of Pembrolizumab in Treatment-Refractory, Microsatellite Instability-High/Mismatch Repair-Deficient Metastatic Colorectal Cancer: KEYNOTE-164. J Clin Oncol 2020;38:11-19.

Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, Skora AD, Luber BS, Azad NS, Laheru D, Biedrzycki B, Donehower RC, Zaheer A, Fisher GA, Crocenzi TS, Lee JJ, Duffy SM, Goldberg RM, de la Chapelle A, Koshiji M, Bhaijee F, Huebner T, Hruban RH, Wood LD, Cuka N, Pardoll DM, Papadopoulos N, Kinzler KW, Zhou S, Cornish

TC, Taube JM, Anders RA, Eshleman JR, Vogelstein B, Diaz LA, Jr.: PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. N Engl J Med 2015;372:2509-2520.
Leach FS, Nicolaides NC, Papadopoulos N, Liu B, Jen J, Parsons R, Peltomaki P, Sistonen P, Aaltonen LA, Nystrom-Lahti M, et al.: Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. Cell 1993;75:1215-1225.
Ledermann JA, Harter P, Gourley C, Friedlander M, Vergote I, Rustin G, Scott C, Meier W, Shapira-Frommer R, Safra T, Matei D, Fielding A, Spencer S, Rowe P, Lowe E, Hodgson D, Sovak MA, Matulonis U: Overall survival in patients with platinum-sensitive recurrent serous ovarian cancer receiving olaparib maintenance monotherapy: an updated analysis from a randomised, placebo-controlled, double-blind, phase 2 trial. Lancet Oncol 2016;17:1579-1589.
Li F, Mao G, Tong D, Huang J, Gu L, Yang W, Li GM: The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutSalpha. Cell 2013a;153:590-600.
Li J, Lu Y, Akbani R, Ju Z, Roebuck PL, Liu W, Yang JY, Broom BM, Verhaak RG, Kane DW, Wakefield C, Weinstein JN, Mills GB, Liang H: TCPA: a resource for cancer functional proteomics data. Nat Methods 2013b;10:1046-1047.
Lord CJ, Ashworth A: PARP inhibitors: Synthetic lethality in the clinic. Science 2017;355:1152-1158.
Lu C, Guan J, Lu S, Jin Q, Rousseau B, Lu T, Stephens D, Zhang H, Zhu J, Yang M, Ren Z, Liang Y, Liu Z, Han C, Liu L, Cao X, Zhang A, Qiao J, Batten K, Chen M, Castrillon DH, Wang T, Li B, Diaz LA, Jr., Li GM, Fu YX: DNA Sensing in Mismatch Repair-Deficient Tumor Cells Is Essential for Anti-tumor Immunity. Cancer Cell 2021;39:96-108 e106.
Lupo B, Trusolino L: Inhibition of poly(ADP-ribosyl)ation in cancer: old and new paradigms revisited. Biochim Biophys Acta 2014;1846:201-215.
Mandal R, Samstein RM, Lee KW, Havel JJ, Wang H, Krishna C, Sabio EY, Makarov V, Kuo F, Blecua P, Ramaswamy AT, Durham JN, Bartlett B, Ma X, Srivastava R, Middha S, Zehir A, Hechtman JF, Morris LG, Weinhold N, Riaz N, Le DT, Diaz LA, Jr., Chan TA: Genetic diversity of tumors with mismatch repair deficiency influences anti-PD-1 immunotherapy response. Science 2019;364:485-491.
Marabelle A, Le DT, Ascierto PA, Di Giacomo AM, De Jesus-Acosta A, Delord JP, Geva R, Gottfried M, Penel N, Hansen AR, Piha-Paul SA, Doi T, Gao B, Chung HC, Lopez-Martin J, Bang YJ, Frommer RS, Shah M, Ghori R, Joe AK, Pruitt SK, Diaz LA, Jr.: Efficacy of Pembrolizumab in Patients With Noncolorectal High Microsatellite Instability/Mismatch Repair-Deficient Cancer: Results From the Phase II KEYNOTE-158 Study. J Clin Oncol 2020;38:1-10.
Modrich P: Mismatch repair, genetic stability, and cancer. Science 1994;266:1959-1960.
Mowat C, Mosley SR, Namdar A, Schiller D, Baker K: Anti-tumor immunity in mismatch repair-deficient colorectal cancers requires type I IFN-driven CCL5 and CXCL10. J Exp Med 2021;218.
Murata S, Zhang C, Finch N, Zhang K, Campo L, Breuer EK: Predictors and Modulators of Synthetic Lethality: An Update on PARP Inhibitors and Personalized Medicine. Biomed Res Int 2016;2016:2346585.
Natrajan R, Lambros MB, Rodriguez-Pinilla SM, Moreno-Bueno G, Tan DS, Marchio C, Vatcheva R, Rayter S, Mahler-Araujo B, Fulford LG, Hungermann D, Mackay A,

Grigoriadis A, Fenwick K, Tamber N, Hardisson D, Tutt A, Palacios J, Lord CJ, Buerger H, Ashworth A, Reis-Filho JS: Tiling path genomic profiling of grade 3 invasive ductal breast cancers. Clin Cancer Res 2009;15:2711-2722.
Parsons R, Li GM, Longley MJ, Fang WH, Papadopoulos N, Jen J, de la Chapelle A, Kinzler KW, Vogelstein B, Modrich P: Hypermutability and mismatch repair deficiency in RER+ tumor cells. Cell 1993;75:1227-1236.
Rakha EA, Reis-Filho JS, Ellis IO: Basal-like breast cancer: a critical review. J Clin Oncol 2008;26:2568-2581.
Ray Chaudhuri A, Nussenzweig A: The multifaceted roles of PARP1 in DNA repair and chromatin remodelling. Nat Rev Mol Cell Biol 2017;18:610-621.
Sagiv-Barfi I, Kohrt HE, Czerwinski DK, Ng PP, Chang BY, Levy R: Therapeutic antitumor immunity by checkpoint blockade is enhanced by ibrutinib, an inhibitor of both BTK and ITK. Proc Natl Acad Sci U S A 2015;112:E966-972.
Toft DJ, Cryns VL: Minireview: Basal-like breast cancer: from molecular profiles to targeted therapies. Mol Endocrinol 2011;25:199-211.
Turner N, Lambros MB, Horlings HM, Pearson A, Sharpe R, Natrajan R, Geyer FC, van Kouwenhove M, Kreike B, Mackay A, Ashworth A, van de Vijver MJ, Reis-Filho JS: Integrative molecular profiling of triple negative breast cancers identifies amplicon drivers and potential therapeutic targets. Oncogene 2010;29:2013-2023.
Zimmer AS, Gillard M, Lipkowitz S, Lee JM: Update on PARP Inhibitors in Breast Cancer. Curr Treat Options Oncol 2018;19:21.

Figure Legends

Figure 1. DNA mismatch repair MSH2 protein is elevated in BLBC and predicts poor survival. **A.** Cox regression hazard ratio and p-values across all 224 proteins on the RPPA for 119 BLBC samples. DNA repair proteins are highlighted in red. **B.** Hazard ratio (HR) summary for the 6 RPPA proteins with a *P* value less than 0.05. HR above one (dotted line) indicates poor survival for the higher protein level. **C.** Proportional cut points for the six significant RPPA proteins based on log-rank optimal *P* value. **D.** Survival curve for MSH2 in 119 BLBC samples comparing high (red, n=60) to low (black, n=59), $P = 0.0025$ using the log-rank test. **D.** Survival curve for MSH6 in 119 BLBC samples comparing high (red, n=26) to low (black, n=93), $P = 0.013$ using the log-rank test. **E.** MSH2-MSH6 protein correlations for breast cancer samples with molecular subtype designations (n=670, $P < 1e-16$). **F.** MSH2 RPPA protein quantification across molecular subtypes, BLBC (n=119), HER2+ (N=61), Luminal A (n=328), and Luminal B (n=162). $P < 0.000001$ for MSH2 in BLBC based on one-way ANOVA with Tukey HSD adjustment for

multiple comparisons. **G.** MSH6 RPPA protein quantification across molecular subtypes. $P < 0.000001$ based on one-way ANOVA with Tukey HSD adjustment for multiple comparisons.

Figure 2. Opposing functions of MSH2 and MLH1 in breast cancer metastasis. **A-B.** Primary tumor growth curve (A) or lung metastasis (B) of control (Con), MSH2-KO or MLH1-KO cells of Py8119 orthotopically injected into #4 fatpad of in 8-week old female C57BL/6J (n = 7 per group). **C-D.** Primary tumor growth curve (C) or lung metastasis (D) of control (Con), MSH2-KO or MLH1-KO cells of 4T1 orthotopically injected into #4 fatpad of 8-week old female Balb/C mice (n = 7 per group). **E-F.** Lung (E) or liver (F) metastasis of Con, MSH2-KO or MLH1-KO cells of Py8119 *i.v.* injected into 8-week old female C57BL/6J (n = 7 per group). **G.** Lung metastasis of Con, MSH2-KO or MLH1-KO cells of Py8119 *i.v.* injected into 8-week old female NSG mice (n = 7 per group).

Figure 3. MSH2 KO leads to an overall increase in tumor-infiltrating leukocytes. Tumors and spleens from Fig. 2A were analyzed for immune profiles including (A-B) CD45+ leukocytes; (C-D) CD3+ T cells; (C) NK1.1+ NK cells, NK1.1+GZMB+ active NK cells, or CD3+NK1.1+ NKT cells; (E-F) CD3+CD4+Foxp3- conventional T cells (Tconv) or CD3+CD4+Foxp3+ Treg cells; (G-H) CD3+CD8+ T cells or CD3+CD8+GZMB+ active CD8 T cells; (I-J) CD11c+MCHII+ dendritic cells or CD86+ DCs (act-DC); (K-L) CD11b+Ly6G+ neutrophils; and (M-N) two distinct tumor associated macrophages (CD11c+F4/80+ and CD11c-F4/80+) or monocytes (CD11b+Ly6C+). N = 4-5 per groups.

Figure 4. MSH2 KO induces immune active tumor microenvironment via chemokine induction. **A-D**. Tumors formed from parental (Con), MSH2-KO (*Msh2*) or MLH1-KO (Mlh1) cells of Py8119 (Fig. 2A) or 4T1 (Fig. 2C) were subject to RNA sequencing, following with GSEA analysis using the C5: ontology gene sets. MSH2-KO tumors enrich in pathways related to

chemokine activity in (A) Py8119 or (B) 4T1 BLBC model related to Con and MLH1-KO tumors; (C) Venn diagram showing the 12 common ontology genesets in the MSH2-KO tumors of both Py8119 and 4T1 models; and (D) MSH2-KO tumors enrich in pathways related to antigen presentation. **E.** MSH2-KO increases immune modulatory chemokine/cytokine proteins in conditional medium using Proteome Profiler Mouse Cytokine Array Kit, Panel A (Biotechne, R&D). **F-I.** Correlation of MSH2 protein with immune modulatory molecules within the TCGA Breast Cancer BLBC dataset, including (F) chemokines; (G) MHC II genes; (H) MHC I genes; (I) immune cell lineage markers. **J-K.** Immune infiltration score based on the ESTIMATE algorithm by MSH2 mean protein value, comparing MSH2-high (n=60) to MSH2-low (n=59) BLBC specimens from the TCGA breast cancer dataset. **(J)** total ESTIMATE score for all leukocytes (P value is determined using Welch's t-test); (**K)** Condensed CIBERSORT estimates of immune cell lineages based on mRNA values available for the BLBC samples.

Figure 5. Long-term MSH2 KO induces an exhaustion phenotype of T cells and sensitizes tumors to anti-PD-1 therapy. **A.** MSH2 KO increases the expression of immune checkpoints in MDA-MB-231 BLBC cells in late passages (P30). Lentiviral particles encoding Cas9 and guide RNA for MSH2 were used to infect MDA-MB-231 cells, following 5-day selection process with puromycin to induce MSH2 KO. The MSH2 KO cells were passaged to P5 or P30, following with RNAseq and differential expression genes (DEGs). Control cells were infected lentiviral particles encoding Cas9 and scramble control. n = 3 per group. **B-D**. Long-term MSH2 KO in 4T1 model leads to T cell exhaustion and sensitizes tumors to anti-PD-1 therapy. (B) Experimental diagram, including whole 4T1 cell vaccine 7 days before tumor cell injection. Anti-PD-1 treatment starts one day after tumor cell injection, every 4 days upto 2 weeks. (C) survival curves and (B) tumor growth curves were shown (n = 7 per group). **E-F**. MSH2 KO induces T cell exhaustion and sensitizes tumors to anti-PD-1 treatment. (E) Example of flow cytometry

using tumors in C, using PD-1+TIM-3+ as exhaustion T cell population; and (F) Histogram showing the statistics of PD-1+TIM-3+ T cell population (n = 4-5 per group).

Figure 6. MSH2 is involved in epigenetic regulation of immune checkpoints via the inhibition of DNMT1 activity. A. Reverse correlation between CpG methylation and gene expression of PVR and CD274 (PD-L1) within the TCGA breast cancer cohorts. **B-C.** MSH2 KO releases the expression of PVR and CD274 in both (B) epigenetic and (C) transcriptional pathways. (B) Control, P5 or P30 of MSH2 KO MDA-MB-231 cells were treated with 0.5 μM of 5-azacytidine for 3 days. (C) Control, P5 or P30 of MSH2 KO MDA-MB-231 cells were treated with 10 ng/ml of IFNγ for the indicated hrs. B-D. RNA extraction, reverse transcription and real-time PCR were used to quantitate gene expression (n = 3). **E-I**. MutSα contributes to DNMA1-dependent DNA methylation. (E) Diagram showing the principle of DNMT1-mediated DNA methylation. A G/T mismatch hemi-methylated substrate – labeled with Cy5, is used to recruit MutSα for determining its impact on DNMT1 enzymatic activity. Fully methylated substrate is digested by GlaI endonuclease; the partial product with Cy5 fluorescence can be separated by SDS-PAGE and visualized. **F-G**. MutSα contributes to DNMT1-dependent DNA methylation. (F) Representative image showing that MutSα further enhances the DNA methyltransferase activity of DNMT1 and (G) summary of DNMT1 activity with or without the presence of MutSα (n = 3). **H-I.** MutSα-enhanced DNA methylation is specifically mediated by the DNMT1 enzymatic activity. Similar DNA methylation assay was performed with or without the presence of RFTS – a DNMT1-specific inhibitor, at different doses. (H) representative image and (I) summary of summary of DNMT1 activity with or without the presence of MutSα or RFTS (n = 3).

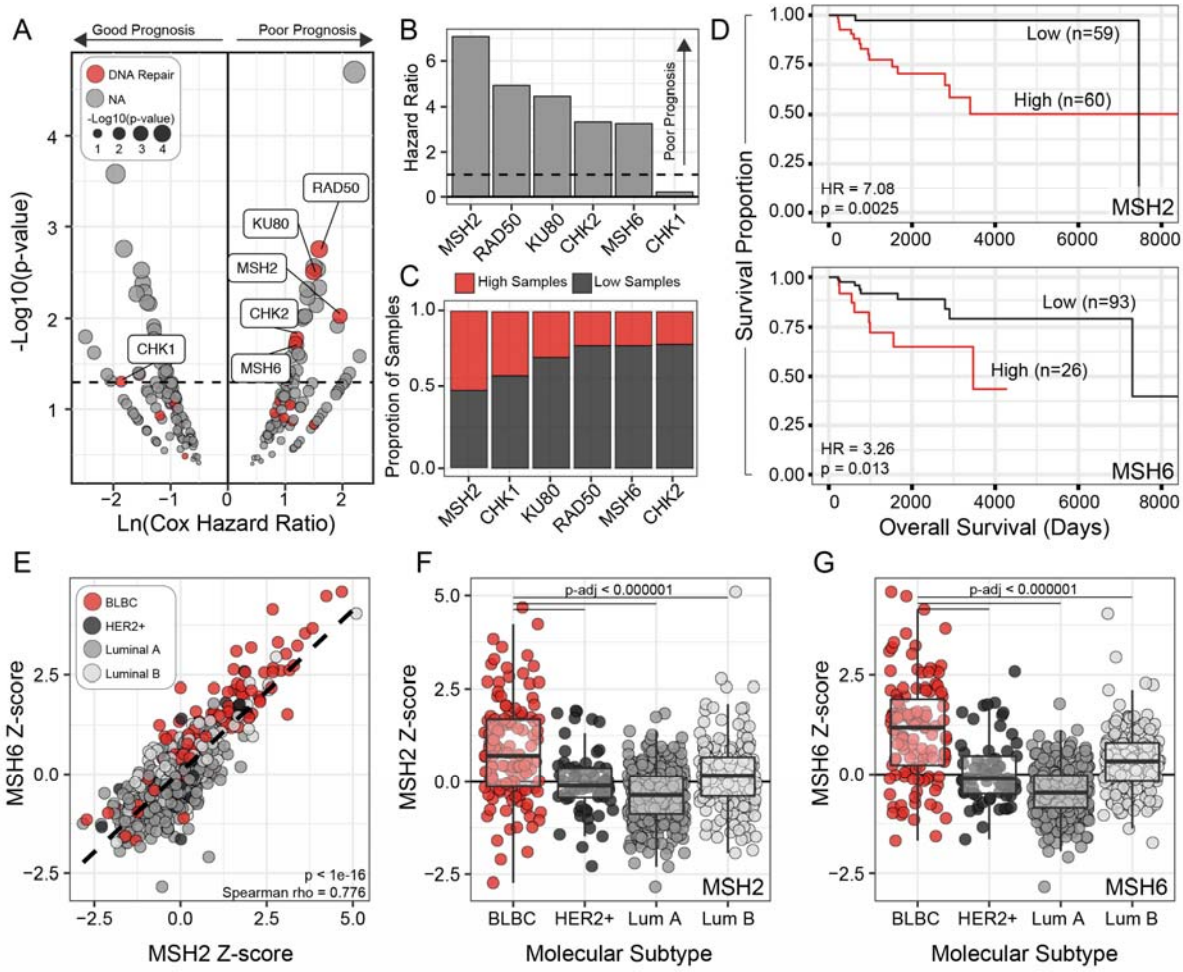

A
Good Prognosis
Poor Prognosis
DNA Repair
NA
-Log10(p-value)
1 2 3 4
RAD50
KU80
MSH2
CHK2
MSH6
CHK1
-Log10(p-value)
Ln(Cox Hazard Ratio)
B
Hazard Ratio
Poor Prognosis
MSH2
RAD50
KU80
CHK2
MSH6
CHK1
C
High Samples
Low Samples
Proprotion of Samples
MSH2
CHK1
KU80
RAD50
MSH6
CHK2
D
Low (n=59)
High (n=60)
HR = 7.08
p = 0.0025
MSH2
Low (n=93)
High (n=26)
HR = 3.26
p = 0.013
MSH6
Survival Proportion
Overall Survival (Days)
E
BLBC
HER2+
Luminal A
Luminal B
MSH6 Z-score
MSH2 Z-score
p < 1e-16
Spearman rho = 0.776
F
p-adj < 0.000001
MSH2 Z-score
MSH2
BLBC
HER2+
Lum A
Lum B
Molecular Subtype
G
p-adj < 0.000001
MSH6 Z-score
MSH6
BLBC
HER2+
Lum A
Lum B
Molecular Subtype

Figure 1

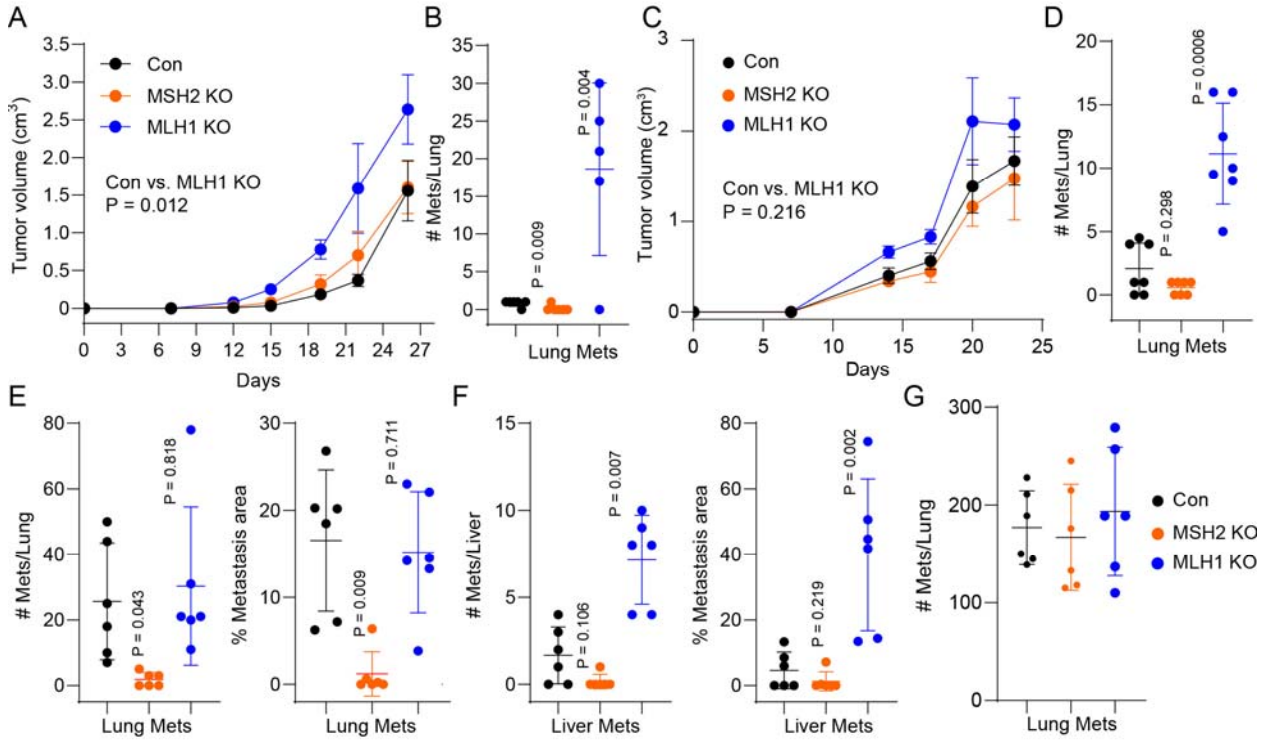

A
Con
MSH2 KO
MLH1 KO
Con vs. MLH1 KO
P = 0.012
Tumor volume (cm³)
Days
B
# Mets/Lung
P = 0.009
P = 0.004
Lung Mets
C
Con
MSH2 KO
MLH1 KO
Con vs. MLH1 KO
P = 0.216
Tumor volume (cm³)
Days
D
# Mets/Lung
P = 0.298
P = 0.0006
Lung Mets
E
# Mets/Lung
P = 0.043
P = 0.818
Lung Mets
% Metastasis area
P = 0.009
P = 0.711
Lung Mets
F
# Mets/Liver
P = 0.106
P = 0.007
Liver Mets
% Metastasis area
P = 0.219
P = 0.002
Liver Mets
G
Mets/Lung
Con
MSH2 KO
MLH1 KO
Lung Mets

Figure 2

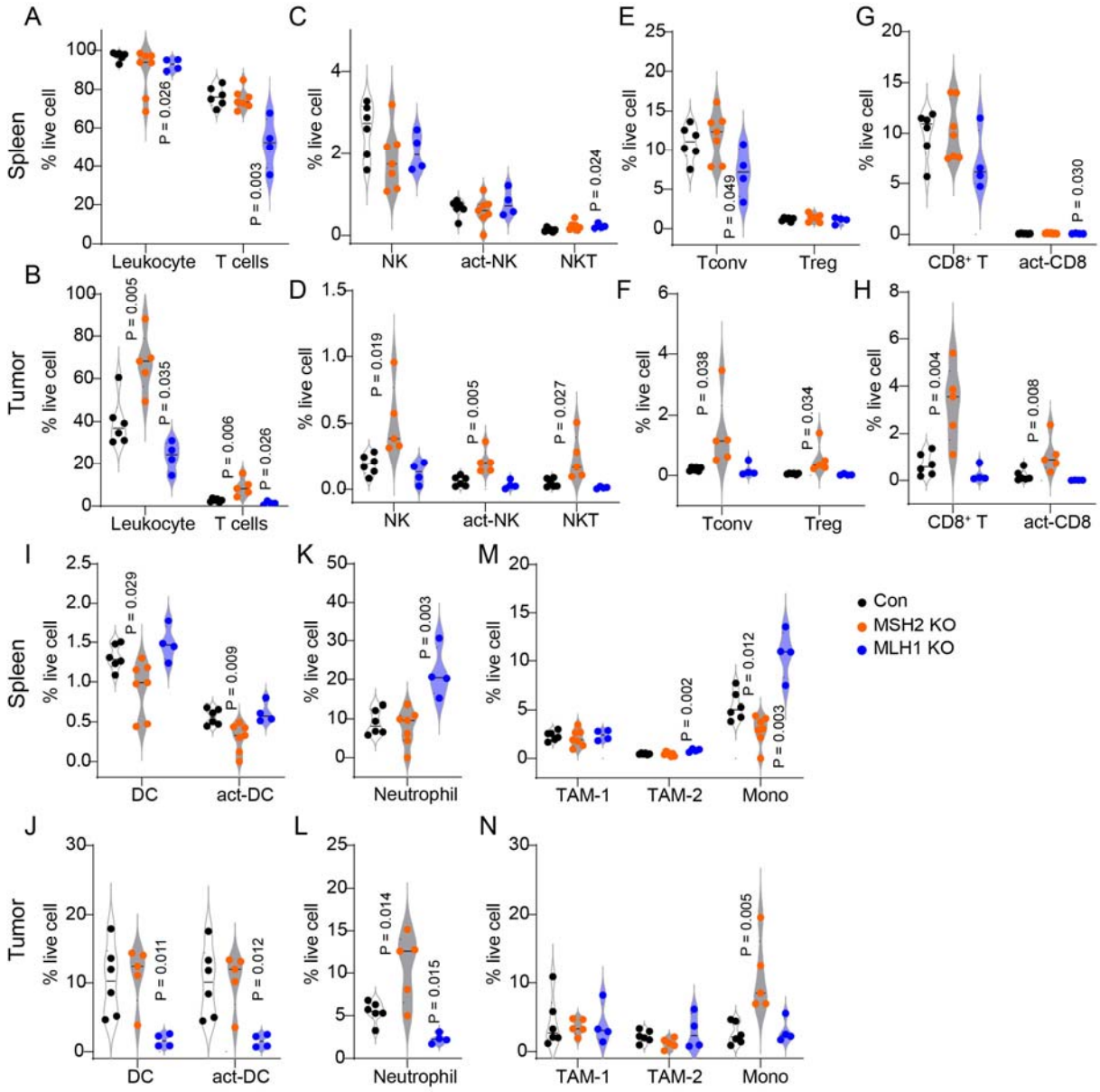

A
B
C
D
E
F
G
H
I
J
K
L
M
N
Spleen
Tumor
% live cell
Leukocyte
T cells
NK
act-NK
NKT
Tconv
Treg
CD8⁺ T
act-CD8
DC
act-DC
Neutrophil
TAM-1
TAM-2
Mono
P = 0.026
P = 0.003
P = 0.024
P = 0.049
P = 0.030
P = 0.005
P = 0.035
P = 0.006
P = 0.026
P = 0.019
P = 0.005
P = 0.027
P = 0.038
P = 0.034
P = 0.004
P = 0.008
P = 0.029
P = 0.009
P = 0.003
P = 0.002
P = 0.012
P = 0.003
P = 0.011
P = 0.012
P = 0.014
P = 0.015
P = 0.005
Con
MSH2 KO
MLH1 KO

Figure 3

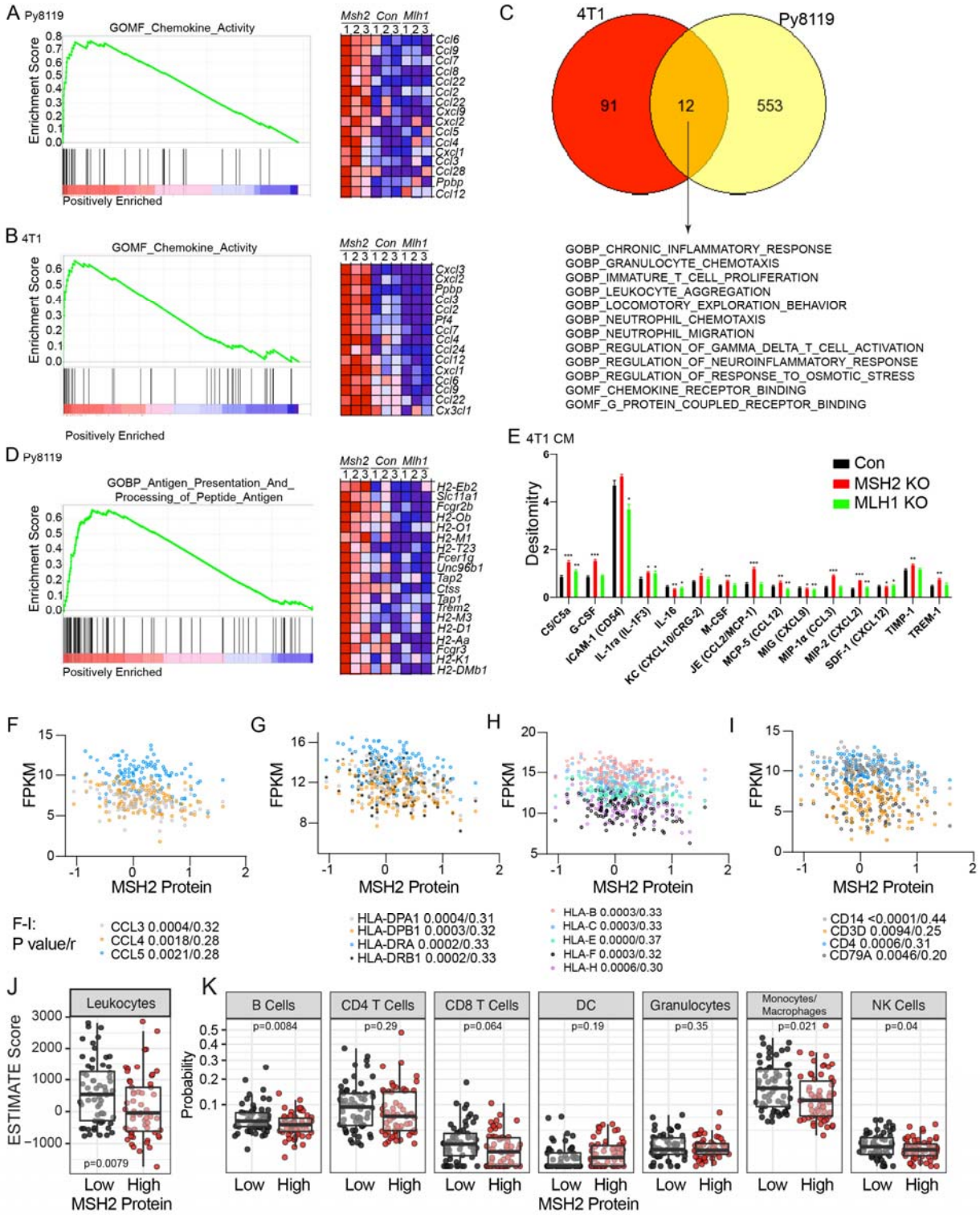

A Py8119
GOMF_Chemokine_Activity
Enrichment Score
Positively Enriched
Msh2 Con Mlh1
Ccl6
Ccl9
Ccl7
Ccl8
Ccl22
Ccl2
Ccl22
Cxcl9
Cxcl2
Ccl5
Ccl4
Cxcl1
Ccl3
Ccl28
Ppbp
Ccl12
B 4T1
GOMF_Chemokine_Activity
Enrichment Score
Positively Enriched
Msh2 Con Mlh1
Cxcl3
Cxcl2
Ppbp
Ccl3
Ccl2
Pf4
Ccl7
Ccl4
Ccl24
Ccl12
Cxcl1
Ccl6
Ccl9
Ccl22
Cx3cl1
C
4T1
Py8119
91
12
553
GOBP_CHRONIC_INFLAMMATORY_RESPONSE
GOBP_GRANULOCYTE_CHEMOTAXIS
GOBP_IMMATURE_T_CELL_PROLIFERATION
GOBP_LEUKOCYTE_AGGREGATION
GOBP_LOCOMOTORY_EXPLORATION_BEHAVIOR
GOBP_NEUTROPHIL_CHEMOTAXIS
GOBP_NEUTROPHIL_MIGRATION
GOBP_REGULATION_OF_GAMMA_DELTA_T_CELL_ACTIVATION
GOBP_REGULATION_OF_NEUROINFLAMMATORY_RESPONSE
GOBP_REGULATION_OF_RESPONSE_TO_OSMOTIC_STRESS
GOMF_CHEMOKINE_RECEPTOR_BINDING
GOMF_G_PROTEIN_COUPLED_RECEPTOR_BINDING
D Py8119
GOBP_Antigen_Presentation_And_Processing_of_Peptide_Antigen
Enrichment Score
Positively Enriched
Msh2 Con Mlh1
H2-Eb2
Slc11a1
Fcgr2b
H2-Ob
H2-O1
H2-M1
H2-T23
Fcer1g
Unc96b1
Tap2
Ctss
Tap1
Trem2
H2-M3
H2-D1
H2-Aa
Fcgr3
H2-K1
H2-DMb1
E 4T1 CM
Desitomitry
Con
MSH2 KO
MLH1 KO
C5/C5a
G-CSF
ICAM-1 (CD54)
IL-1ra (IL-1F3)
IL-16
M-CSF
KC (CXCL10/CRG-2)
JE (CCL2/MCP-1)
MCP-5 (CCL12)
MIG (CXCL9)
MIP-1α (CCL3)
MIP-2 (CXCL2)
SDF-1 (CXCL12)
TIMP-1
TREM-1
F
FPKM
MSH2 Protein
G
FPKM
MSH2 Protein
H
FPKM
MSH2 Protein
I
FPKM
MSH2 Protein
F-I:
P value/r
CCL3 0.0004/0.32
CCL4 0.0018/0.28
CCL5 0.0021/0.28
HLA-DPA1 0.0004/0.31
HLA-DPB1 0.0003/0.32
HLA-DRA 0.0002/0.33
HLA-DRB1 0.0002/0.33
HLA-B 0.0003/0.33
HLA-C 0.0003/0.33
HLA-E 0.0000/0.37
HLA-F 0.0003/0.32
HLA-H 0.0006/0.30
CD14 <0.0001/0.44
CD3D 0.0094/0.25
CD4 0.0006/0.31
CD79A 0.0046/0.20
J
Leukocytes
ESTIMATE Score
p=0.0079
Low High
MSH2 Protein
K
Probability
B Cells
p=0.0084
CD4 T Cells
p=0.29
CD8 T Cells
p=0.064
DC
p=0.19
Granulocytes
p=0.35
Monocytes/ Macrophages
p=0.021
NK Cells
p=0.04
Low High
MSH2 Protein

Figure 4

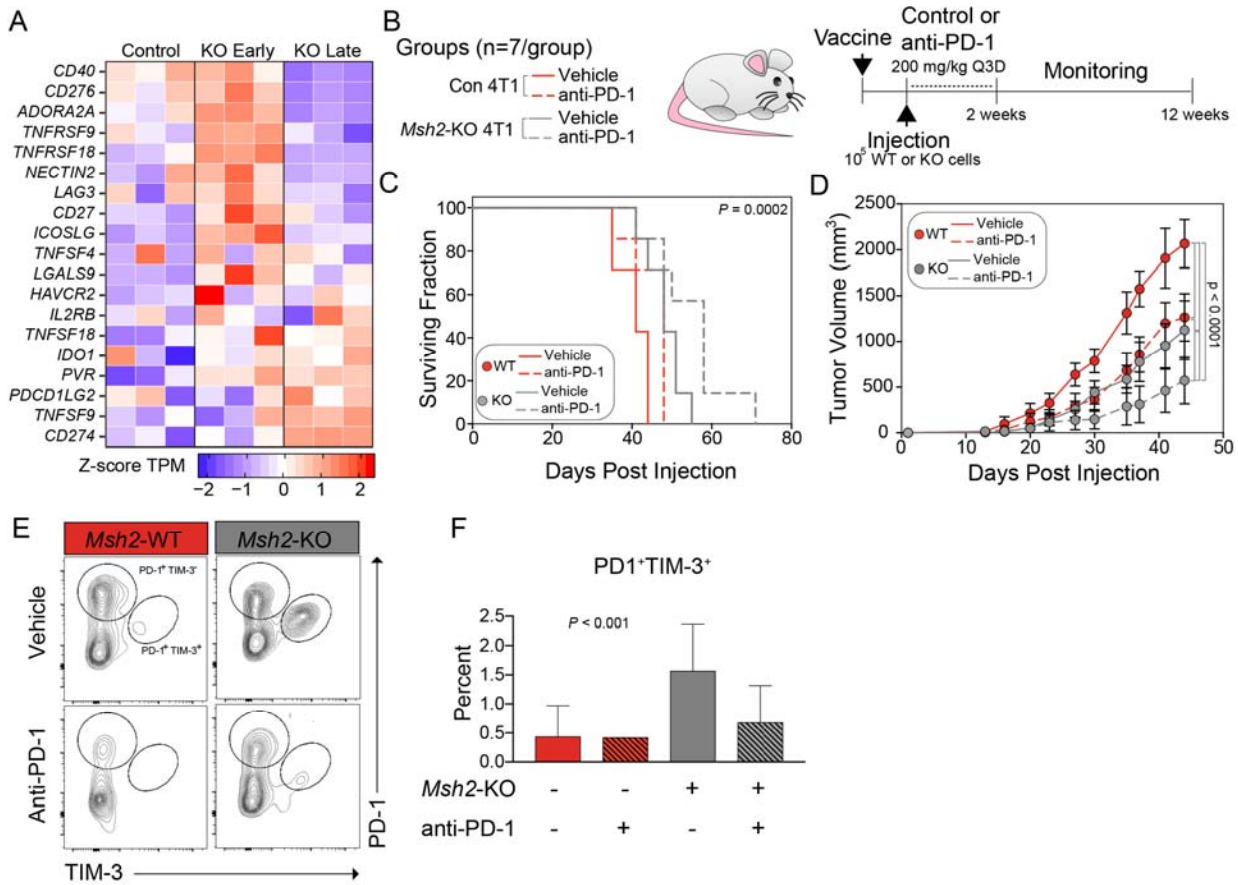

A
Control
KO Early
KO Late
CD40
CD276
ADORA2A
TNFRSF9
TNFRSF18
NECTIN2
LAG3
CD27
ICOSLG
TNFSF4
LGALS9
HAVCR2
IL2RB
TNFSF18
IDO1
PVR
PDCD1LG2
TNFSF9
CD274
Z-score TPM
−2 −1 0 1 2
B
Groups (n=7/group)
Con 4T1
Vehicle
anti-PD-1
Msh2-KO 4T1
Vehicle
anti-PD-1
Vaccine
Control or anti-PD-1
200 mg/kg Q3D
Monitoring
2 weeks
12 weeks
Injection
10^5 WT or KO cells
C
P = 0.0002
Surviving Fraction
Days Post Injection
WT
KO
Vehicle
anti-PD-1
D
Tumor Volume (mm^3)
Days Post Injection
p < 0.0001
E
Msh2-WT
Msh2-KO
Vehicle
Anti-PD-1
PD-1
TIM-3
F
PD1+TIM-3+
P < 0.001
Percent
Msh2-KO
anti-PD-1

Figure 5

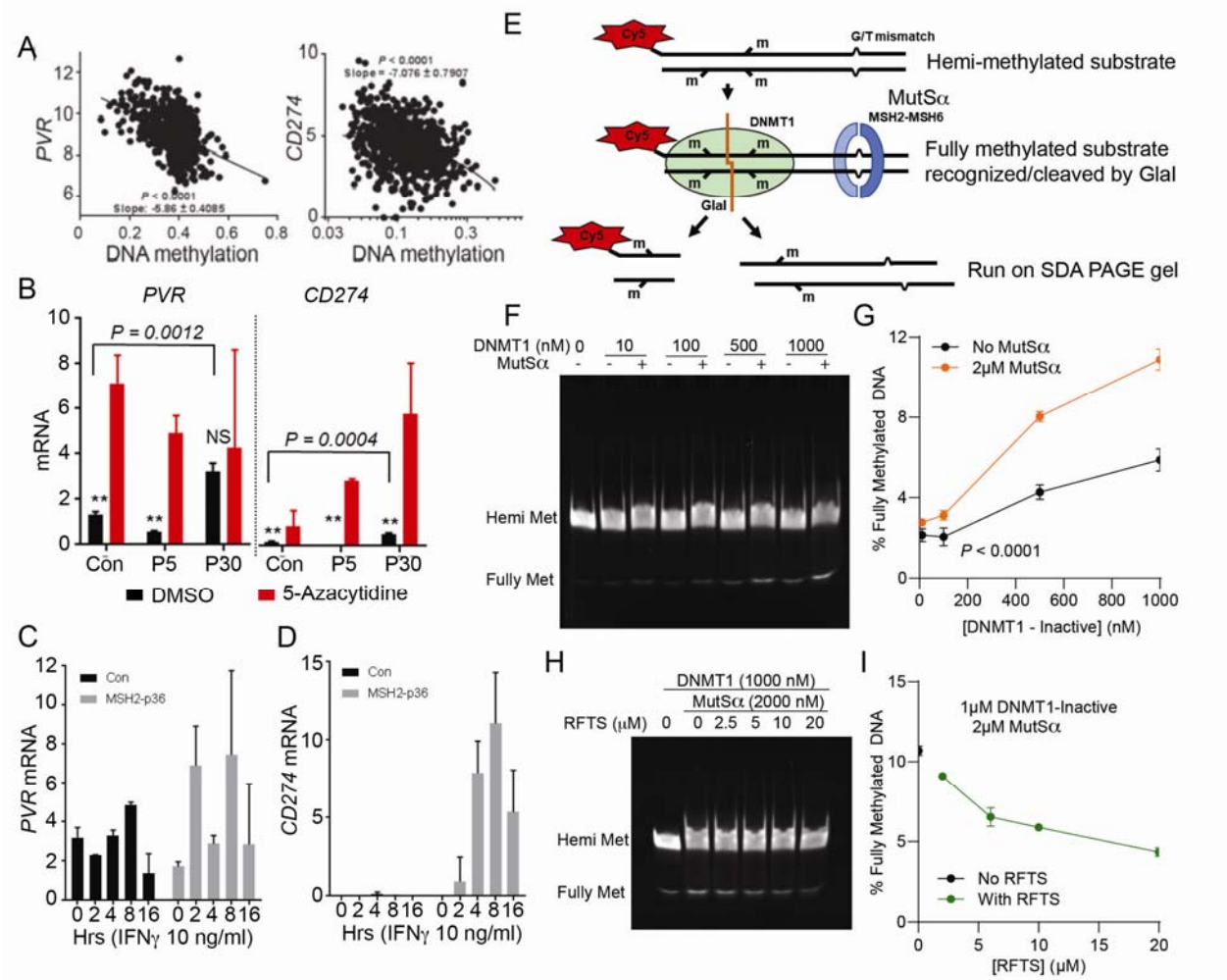

A
PVR
DNA methylation
CD274
DNA methylation
B
PVR
CD274
P = 0.0012
P = 0.0004
NS
mRNA
Con
P5
P30
DMSO
5-Azacytidine
C
Con
MSH2-p36
PVR mRNA
Hrs (IFNγ 10 ng/ml)
D
Con
MSH2-p36
CD274 mRNA
Hrs (IFNγ 10 ng/ml)
E
Cy5
G/T mismatch
Hemi-methylated substrate
MutSα
MSH2-MSH6
DNMT1
Fully methylated substrate
recognized/cleaved by GlaI
GlaI
Run on SDA PAGE gel
F
DNMT1 (nM) 0 10 100 500 1000
MutSα
Hemi Met
Fully Met
G
No MutSα
2μM MutSα
% Fully Methylated DNA
P < 0.0001
[DNMT1 - Inactive] (nM)
H
DNMT1 (1000 nM)
MutSα (2000 nM)
RFTS (μM) 0 0 2.5 5 10 20
Hemi Met
Fully Met
I
1μM DNMT1-Inactive
2μM MutSα
% Fully Methylated DNA
No RFTS
With RFTS
[RFTS] (μM)

Figure 6

Supplementary information.

Supplementary Table 1. The correlation between patient prognosis and DNA repair genes/proteins in the TCGA BLBC dataset.

Hazard Ratios and –log10(p-value) Variables of Major DNA Repair Genes or Proteins (MSH2 and MSH6) from the TCGA and TPGA Datasets. Greater than 1.3 of -log10(p-value) is considered significant.

Supplementary Table 2. MSH2-low human BLBC specimens have significantly enriched signaling pathways involving in immune regulations.

GSEA analysis using the C5: ontology gene sets comparing MSH2 protein-low versus -high human TCGA BLBC specimens (lowest 1/3 versus highest 1/3), showing positively enriched C5: ontology gene sets in the MSH2-low specimens.

Supplementary Table 3. MSH2-high BLBC specimens have elevated pathways related to cancer cell proliferation and progression.

GSEA analysis using the C5: ontology gene sets comparing MSH2 protein-low versus -high human TCGA BLBC specimens (lowest 1/3 versus highest 1/3), showing negatively enriched C5: ontology gene sets in the MSH2-low specimens.

Supplementary Table 4. *MLH1*-low human BLBC specimens have significantly enriched signaling pathways related to active and aggressive cancer types.

GSEA analysis using the C5: ontology gene sets comparing *MLH1* mRNA-low versus -high human TCGA BLBC specimens (lowest 1/3 versus highest 1/3), showing positively enriched C5: ontology gene sets in the *MLH1*-low specimens.

Supplementary Table 5. *MLH1*-high human BLBC specimens have no significantly enriched signaling pathways related to cancer progression.

GSEA analysis using the C5: ontology gene sets comparing *MLH1* mRNA-low versus -high human TCGA BLBC specimens (lowest 1/3 versus highest 1/3), showing negatively enriched C5: ontology gene sets in the *MLH1*-low specimens.

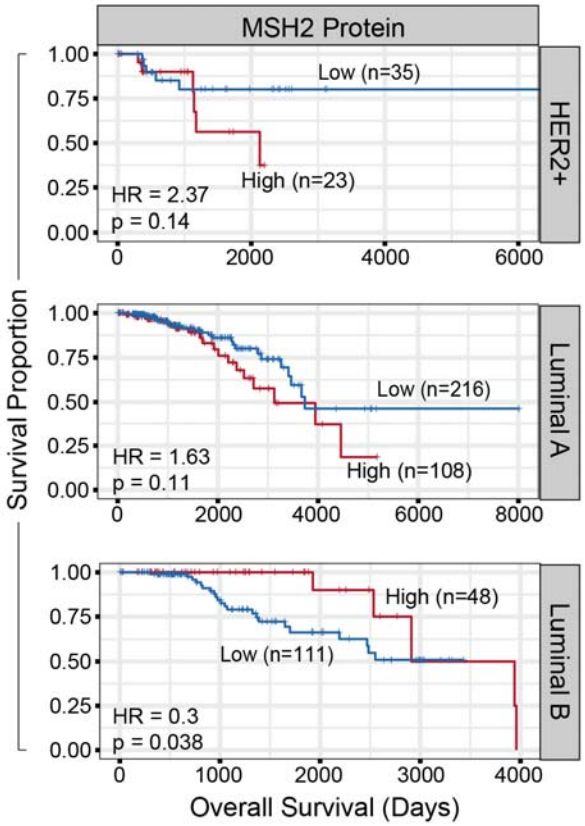


Figure S1. Survival curve based on MSH2 protein levels in the TCGA HER2+, Luminal A, or Luminal B samples, comparing high (red) to low (black). HR and P values are indicated.

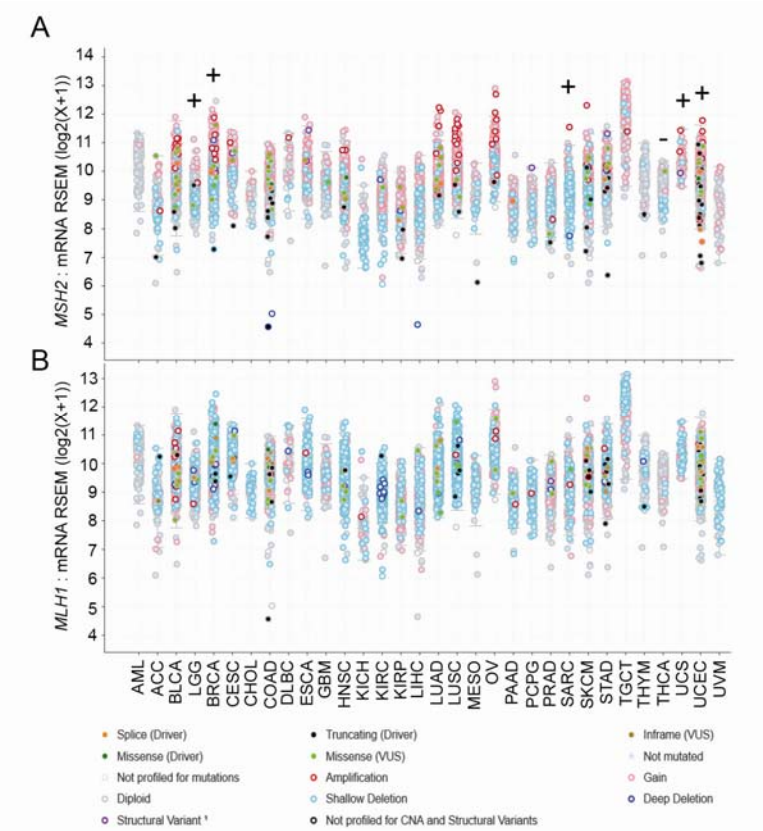
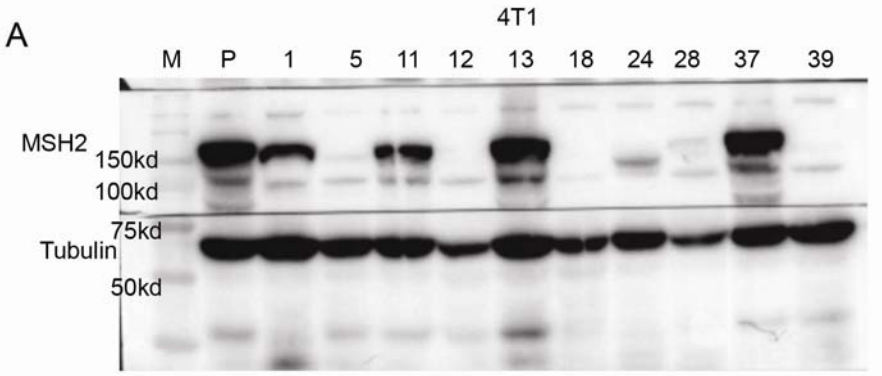
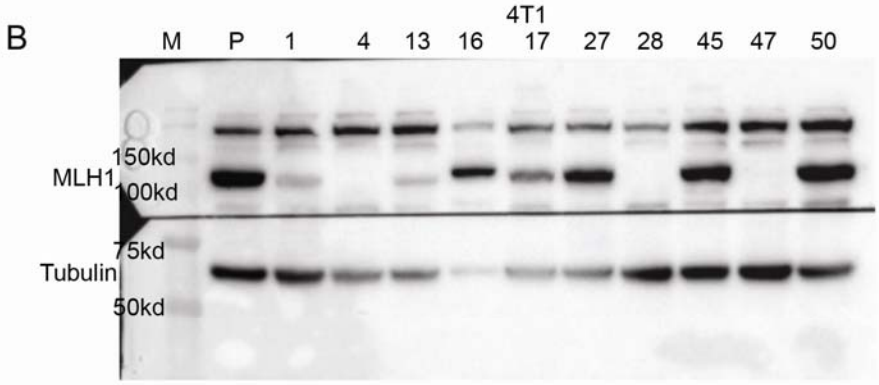
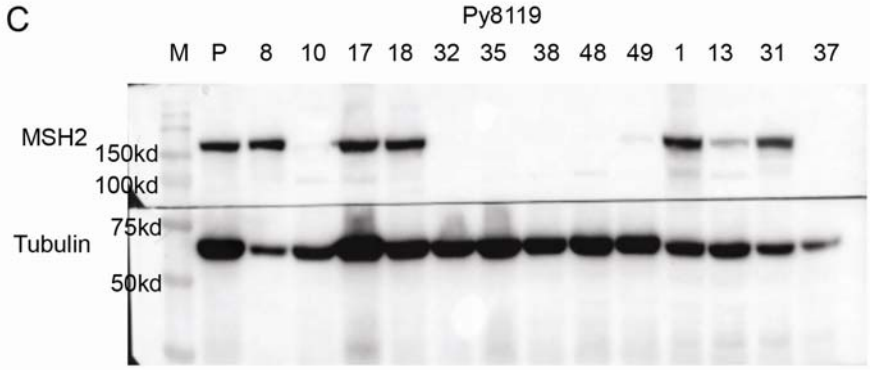


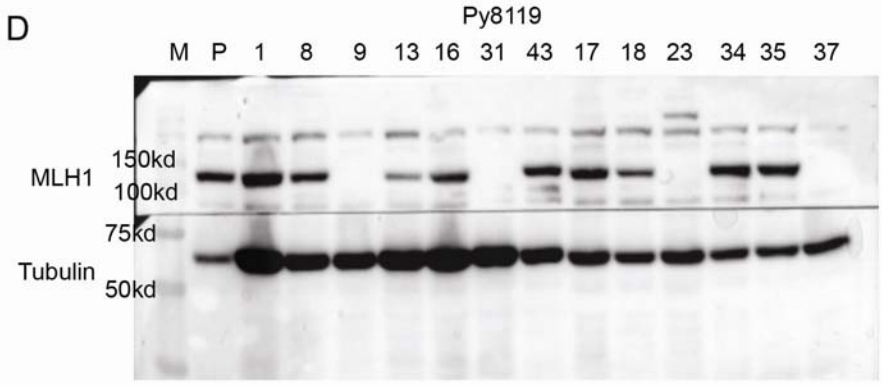
Figure S2. MSH2 – but not MLH1 – exhibits frequent amplification or gain of expression in human cancers. cBioPortal for Cancer Genomics was used to extract **(A)** *MSH2* or **(B)** *MLH1* mRNA expression and genomic variations from the TCGA PanCancer Atlas Studies (n = 10967). Cancer types are using the standard TCGA abbreviations.





Con: A1, A37, A13, A11, B27, B45, B50 combined
MSH2 KO: A5, A12, A18, A28, A39 combined
MLH1 KO: B4, B28, B47 combined





Con: C8, C17, C18, D1, D8 combined
MSH2 KO: C10, C32, C35, C38, C48, C49 combined
MLH1 KO: D9, D31, D23, D37 combined

z

Figure S3. The Generation of MSH2-KO or MLH1-KO cells of Py8119 or 4T1 mouse breast cancer cells using CRISPR/Cas9-based genomic editing method. A-D. Cas9/Guide RNA complex were transiently transfected into Py8119 or 4T1 cells, following with single clone selection, validation using immunoblotting. Several clones were combined for experiments in Figure 2.

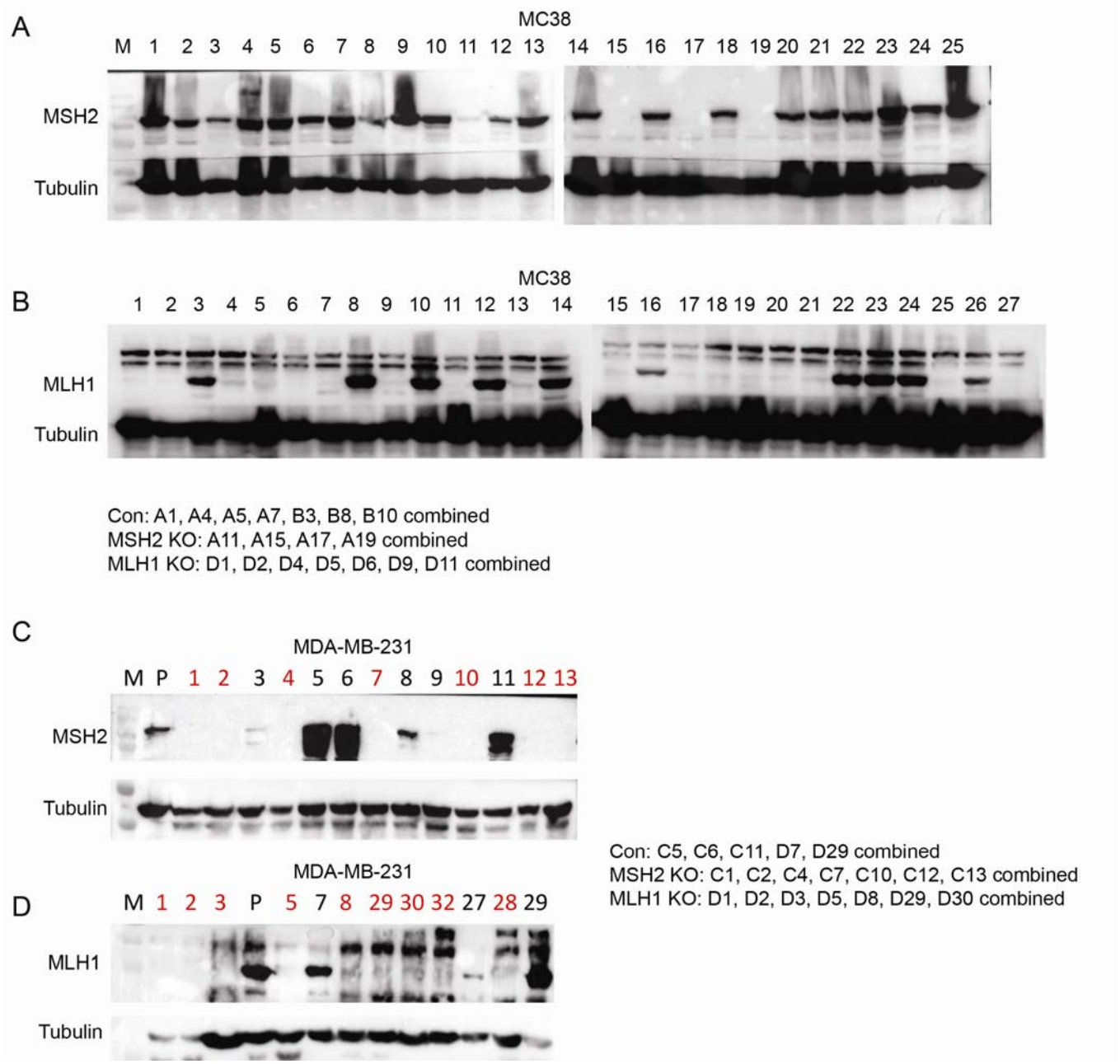


Figure S4. The Generation of MSH2-KO or MLH1-KO cells of MC38 mouse colon cancer cells or MDA-MB-231 human breast cancer cells using CRISPR/Cas9-based genomic editing method. A-D. Cas9/Guide RNA complex were transiently transfected into MC38 or MBA-MD-231 cells, following with single clone selection, validation using immunoblotting. Several clones were combined for experiments in Figure S5.

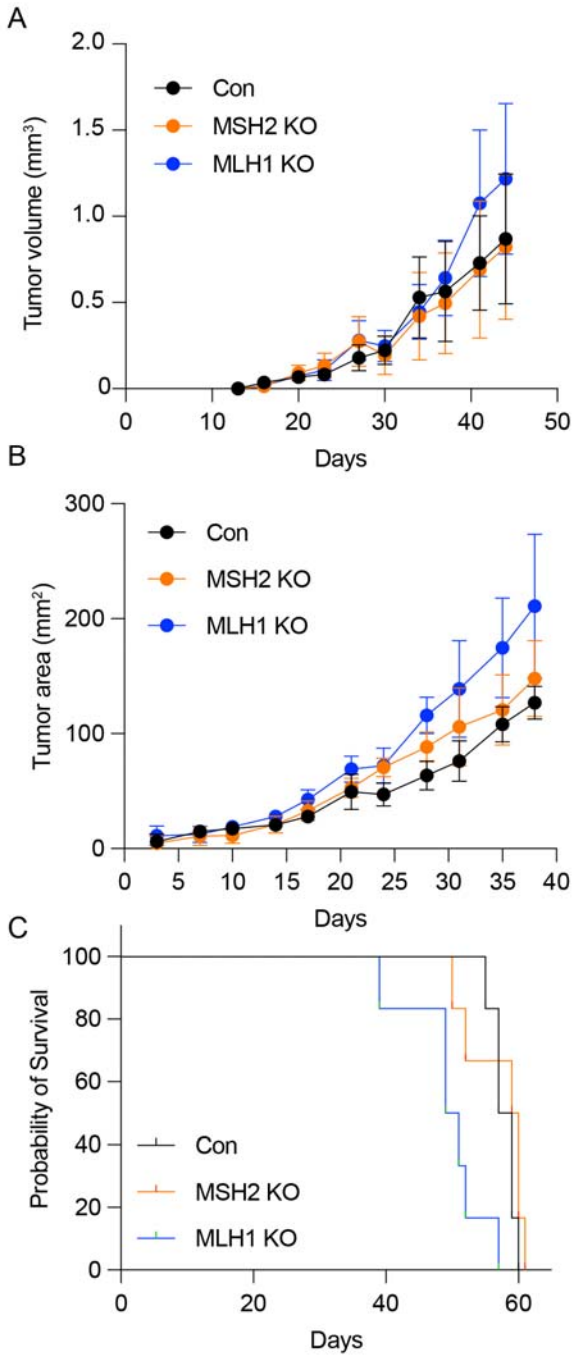


Figure S5. Opposing functions of MSH2 and MLH1 in breast cancer metastasis in immune competent mouse. **A-B.** Primary tumor growth curve (A-B) or survival (C) of control (Con), MSH2-KO or MLH1-KO cells of MC38 cells i.v. injected into 8-week old female C57BL/6J (n = 7 per group) (A) or into #4 mammary fatpad of 8-week old NSG mice (n = 7 per group) (B-C). Tumor growth and survival are monitored.

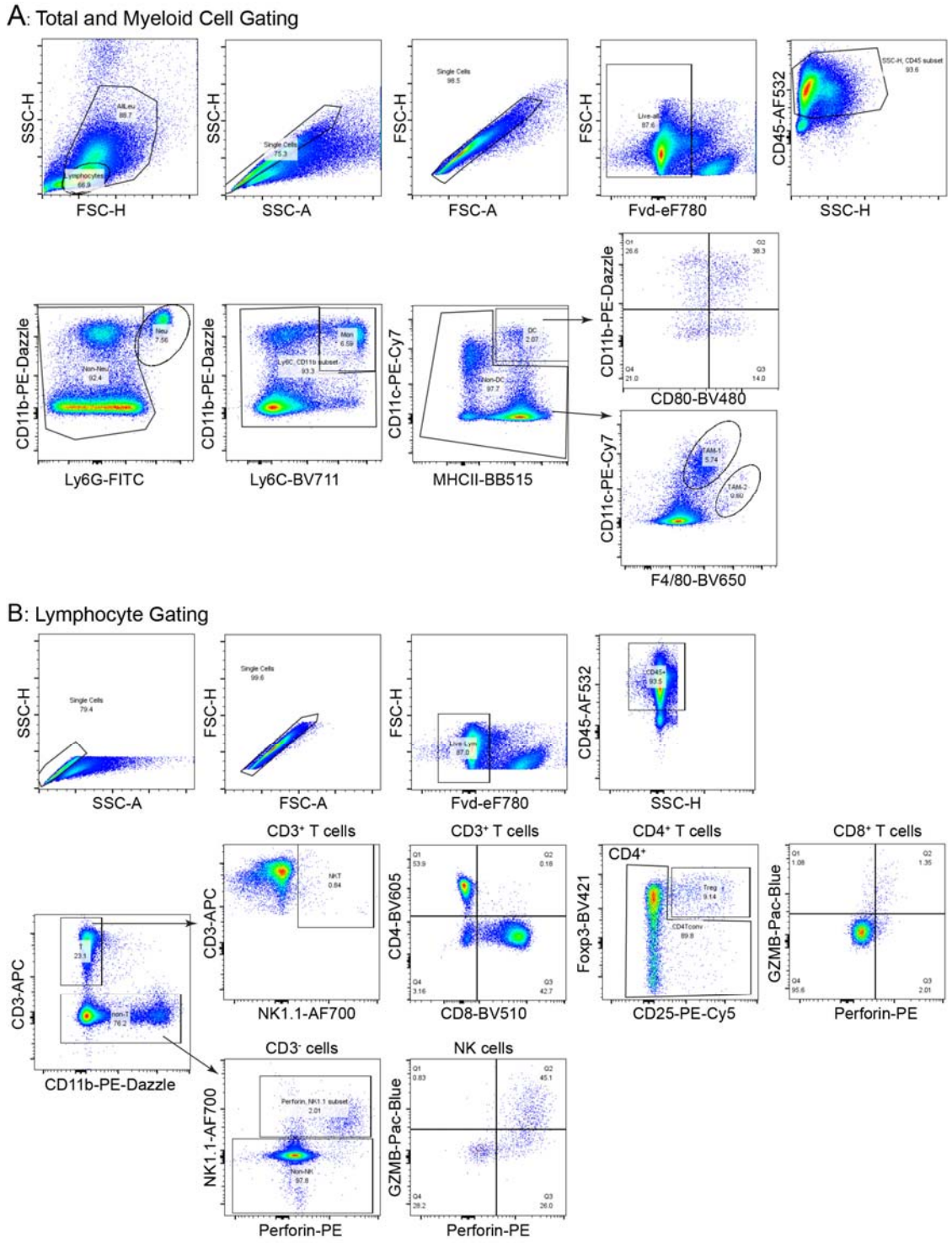


Figure S6. Gating scheme and supplementary information for Figure 3. A. Scheme showing the total live cell and all leukocyte gating. **B.** Scheme showing gating on lymphocytes.

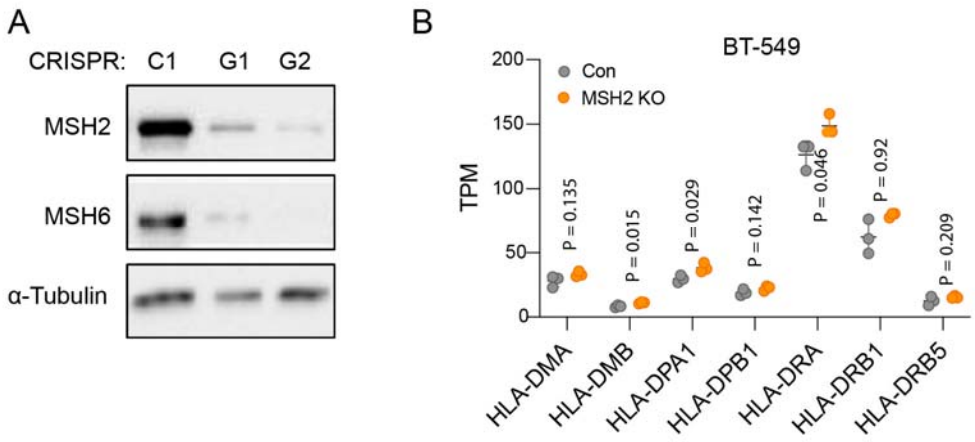


Figure S7. MSH2 KO increases the expression of MHC genes in BT-549 BLBC cells. A. Lentiviral particles encoding Cas9 and guide RNA for MSH2 were used to infect BT-549 cells, following 5-day selection process with puromycin to induce MSH2 KO. The MSH2 KO cells were passaged to P30 and KO efficiency was confirmed using immunoblotting, with concomitant loss of MSH6. **B.** RNAseq and differential expression genes (DEGs) were analyzed from WT and MSH2 KO cells, showing elevation of MHC genes in the MSH2 KO cells.

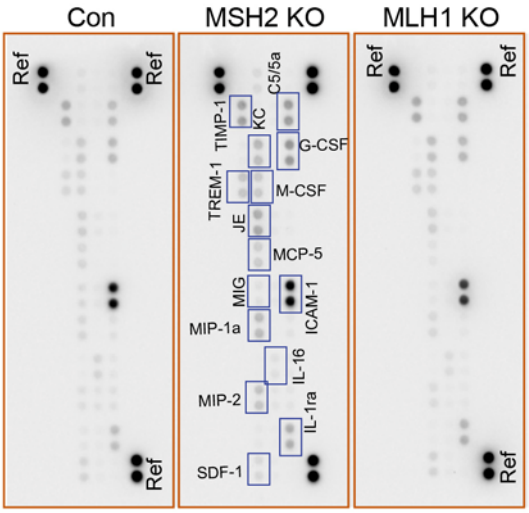


Figure S8. Original images used in and Supplementary information for Fig. 4E (Proteome Profiler Mouse Cytokine Array Kit, Panel A, Biotechne, R&D).

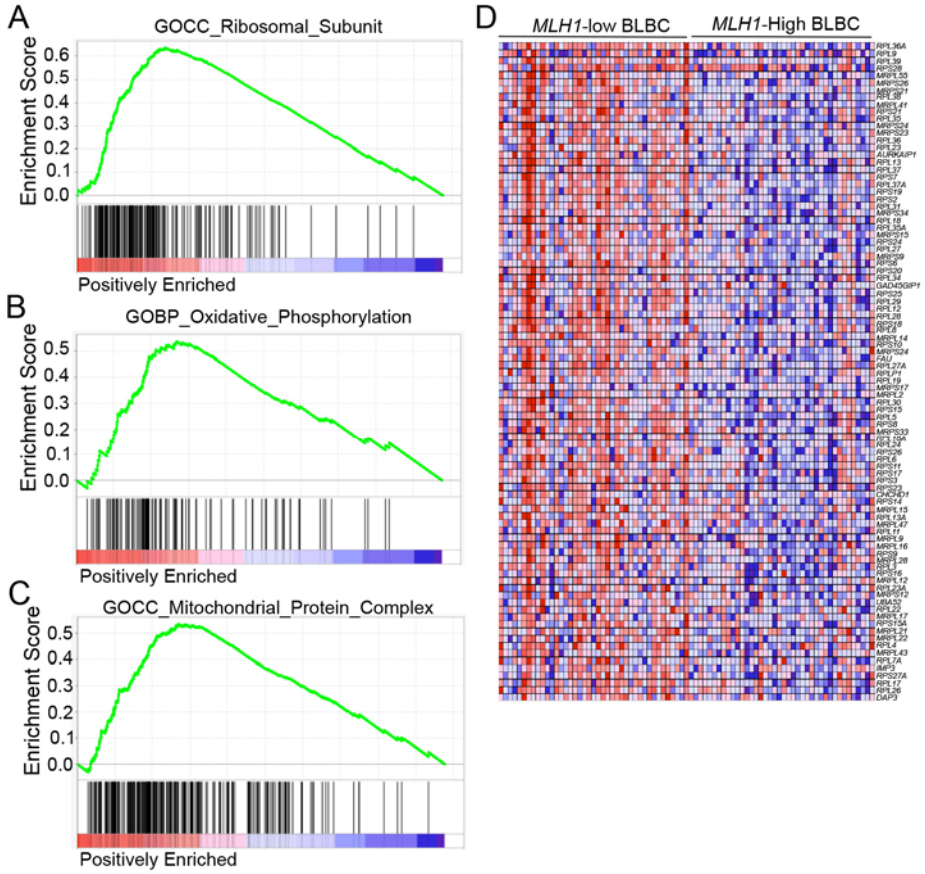


Figure S9. MLH1-low human BLBC specimens have significantly enriched signaling pathways related to metabolism and protein synthesis. Also refer to Supplementary Table 4.

GSEA analysis using the C5: ontology gene sets comparing MLH1 mRNA-low versus -high human TCGA BLBC specimens (lowest 1/3 versus highest 1/3), showing positively enriched C5: ontology gene sets in the MLH1-low specimens, including (A) Enriched Ribosomal Subunit; (B) Increased Oxidative Phosphorylation; (C) Mitochondrial Protein Complex; and (D) individual Ribosomal Subunit listed from (A).

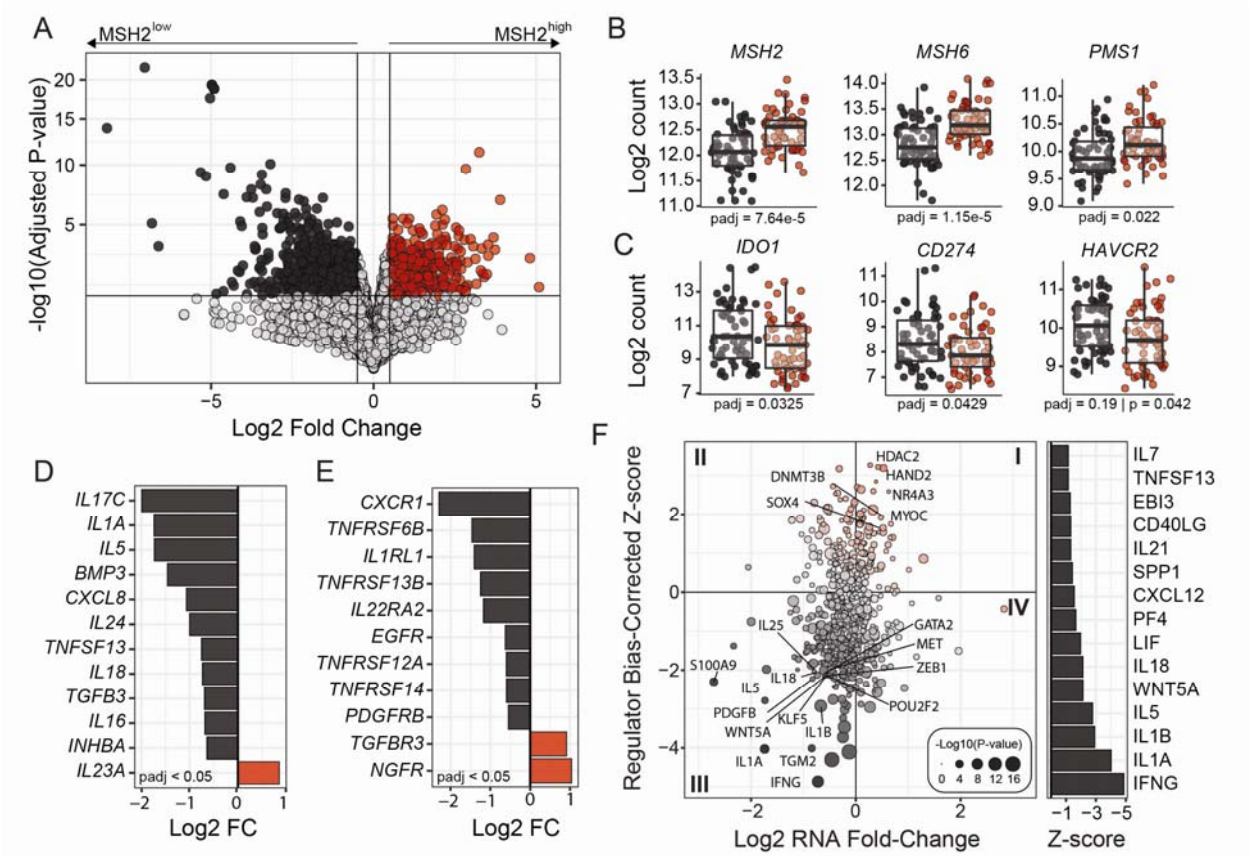


Figure S10. Differential gene expression results based on MSH2 protein in the TCGA BLBC patient specimens. A. Differential gene expression using HTseq count data comparing MSH2-high (n=60) to MSH2-low (n=59) found 650 genes significantly increased and 1,444 significantly decreased. Significance was defined as an adjusted P-value < 0.05 and log2 fold-change ≥ |0.5|. **B.** Regularized log2 transformed count expression of DNA mismatch repair pathway constituents that demonstrated significantly differential between MSH2-high versus low. **C.** Regularized log2 transformed count expression of immune checkpoint targets. **D.** Log2-fold change of significant chemokines and cytokines. **E.** Log2-fold change of significant chemokine and cytokine receptors. **F.** IPA upstream regulators results based on differential gene expression. Quadrant I correspond to regulators with increased expression and predicted increase activity in MSH2-high samples, while Quadrant III corresponds to increased expression

and predicted activity in MSH2-low samples. Bar graph summarizes the significant predicted activation of cytokines.

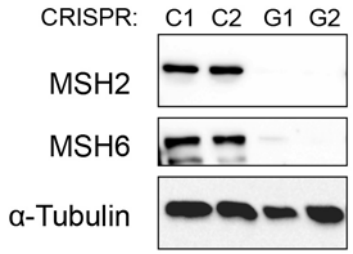


Figure S12. Generation of MSH2 KO MDA-MB-231 BLBC cells. Lentiviral particles encoding Cas9 and guide RNA for MSH2 were used to infect MDA-MB-231 cells, following 5-day selection process with puromycin to induce MSH2 KO. The MSH2 KO cells were passaged to P30 and KO efficiency was confirmed using immunoblotting, with concomitant loss of MSH6.