Helicobacter pylori infection causes characteristic DNA damage patterns in human cells

Max Koeppel, Fernando Garcia-Alcalde, Frithjof Glowinski, Philipp Schlaermann and Thomas F Meyer[#]

Supplemental Information

Supplemental Figure S1

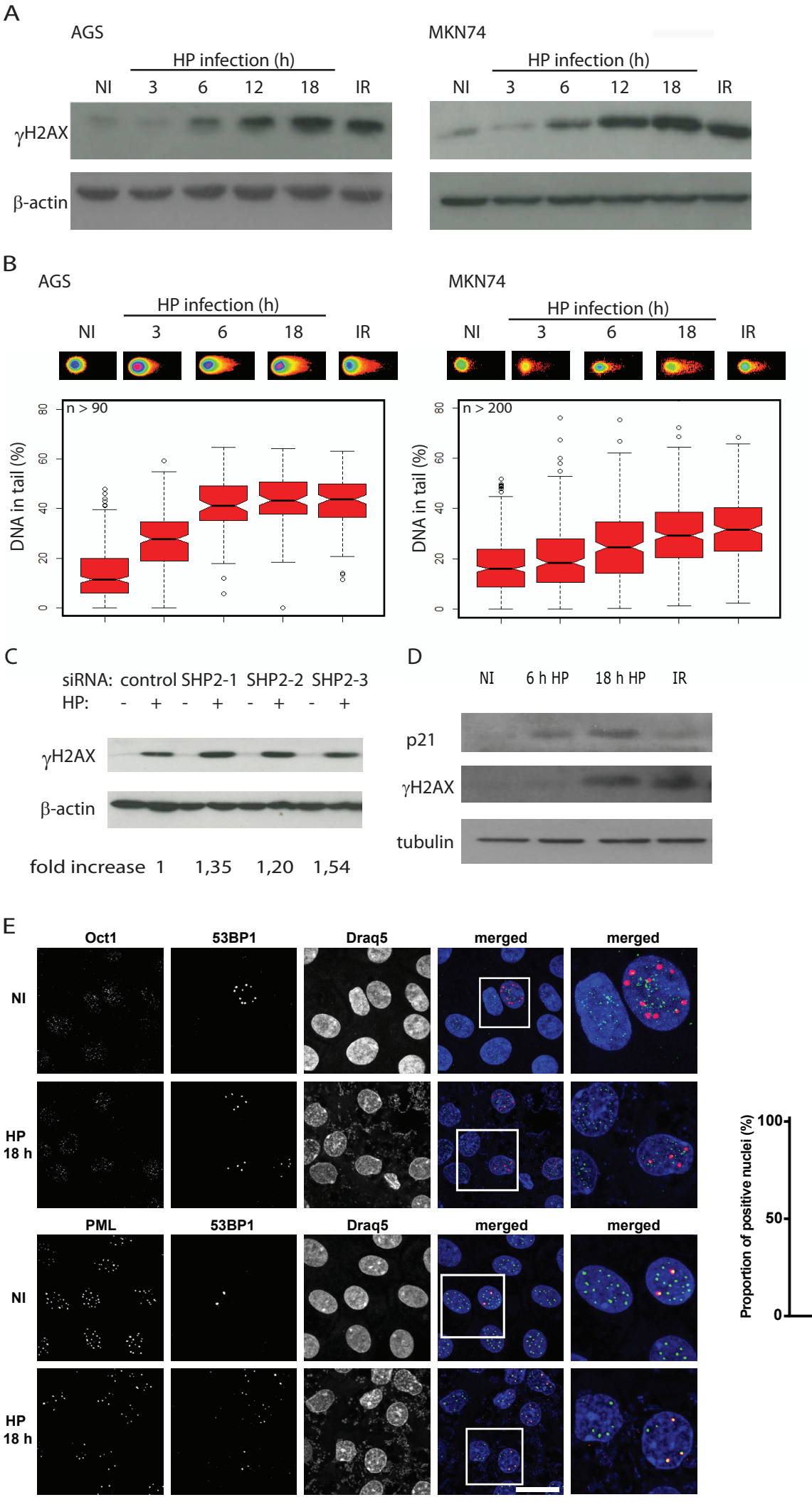


Figure S1, Related to Fig. 1

A) Western blot analysis shows progressive induction of the DNA damage marker γH2AX over time after infection of AGS (left) or MKN74 (right) cells. Cells were infected with *H. pylori* strain P12 at MOI 50, or subjected to 10 Gy IR for comparison. B) Comet assay of infected AGS (left) or MKN74 cells (right) shows the extent of DSBs in individual cells. Infections and irradiation control were performed as in (A). C) AGS cells were transfected with siRNAs against SHP2 or control siRNA and infected 60 h later with *H. plyori* P12 at MOI 50 for 20 h. Representative western blot for the indicated proteins; the fold increase indicates the average ratio of γH2AX-accumulation of SHP2 over control-knockdown after normalization to actin of three biological replicates. D) AGS cells were infected with *H. pylori* strain P12 at MOI 50, or subjected to 10 Gy IR prior to western blot against the indicated proteins showing an accumulation of p21 after infection. E) Confocal immunofluorescence was performed in human primary gastric epithelial cells infected for 18 h with *H. pylori* strain P12 at MOI 25 to show the co-localization of DNA damage foci with proteins common to the indicated subnuclear domains; Oct1 or PML (green) and 53BP1 (red). Nuclei were visualized with Draq5; images are collapsed confocal stacks. Scale bars: 20µm. Proportion of PML and Oct1 positive cells is shown on the right.

Supplemental Fig S2

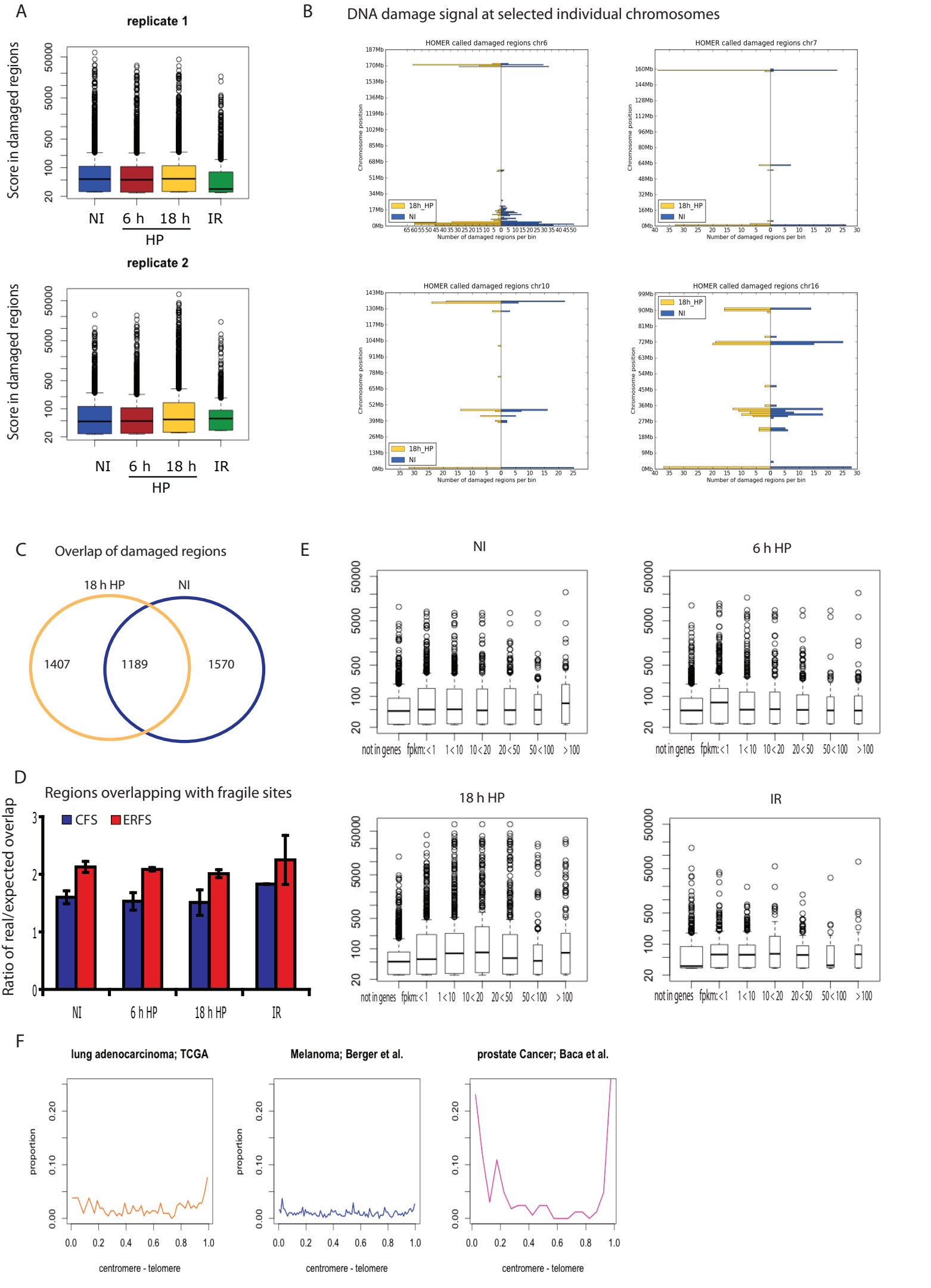
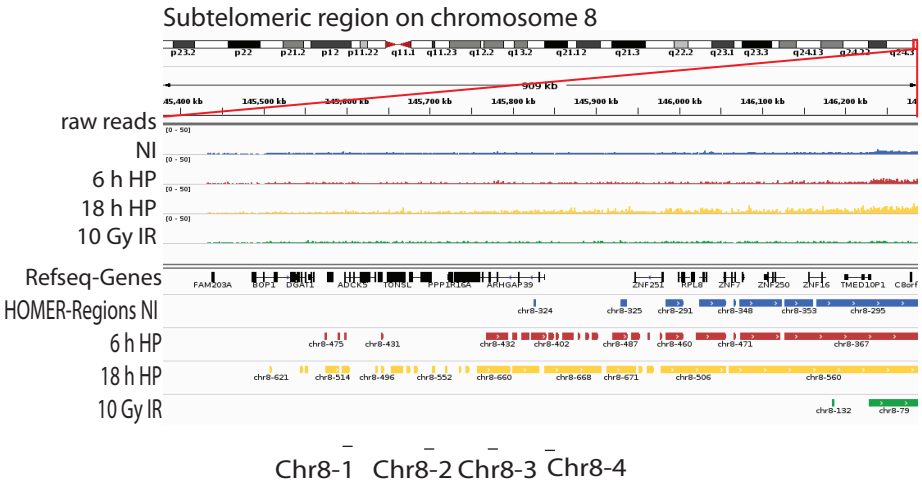


Figure S2, Related to Fig. 3

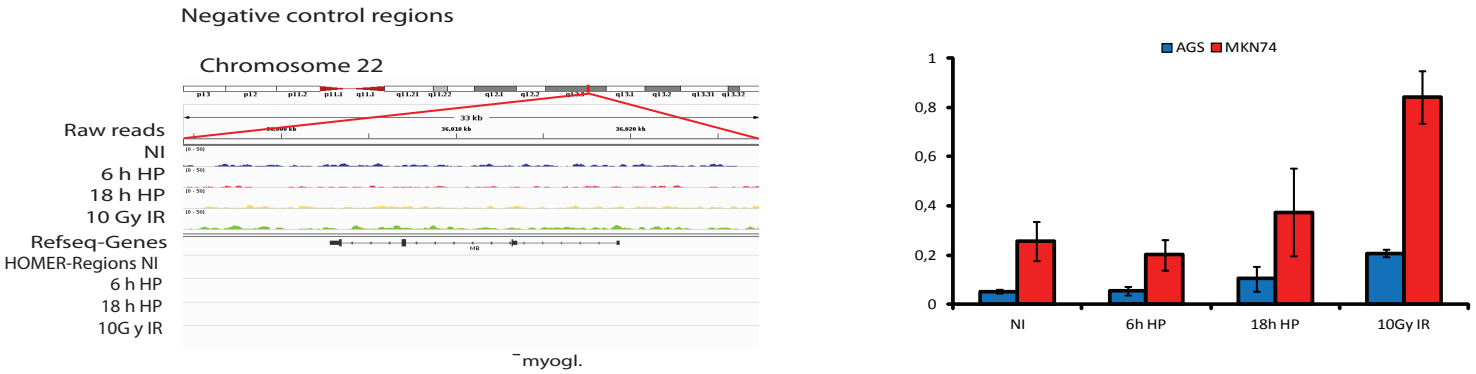
A) Boxplot of the regions identified by HOMER to be enriched under the respective conditions. The HOMER score corresponds to the extent of each region and to the overall signal accumulated within. The black line in each box represents the median. Shown are the scores of the individual replicates. B) AGS cells were subjected to *H. pylori* infection followed by ChIP-seq for H2AXpS139. Damaged regions were identified by HOMER, each chromosome was divided into bins and the damage signal in each bin was calculated. Shown are the accumulated signals from the biological replicates for non-infected cells (blue) and 18 h-infected cells (yellow). C) Damaged regions of non-infected cells and 18 h p.i. were separated into unique and overlapping damaged sites showing slightly more condition-specific regions. D) Hypergeometric distribution testing indicates the ratio between real vs. expected random overlap between damaged enriched regions and CFS or ERFS, respectively (in all cases p-value< 10^{-11}) E) Damaged regions from non-infected, infected and irradiated cells were overlapped with non-genic regions and those of different transcription levels, as defined by FPKM. Boxplots depict the extent of damage in the respective categories; width of the boxes corresponds to the number of regions. F) Distribution of structural variation in different cancers across chromosomal arms. Whole genome sequencing data of multiple donors from the indicated studies was used to plot the location of structural variations in aggregated bins of chromosomal arms.

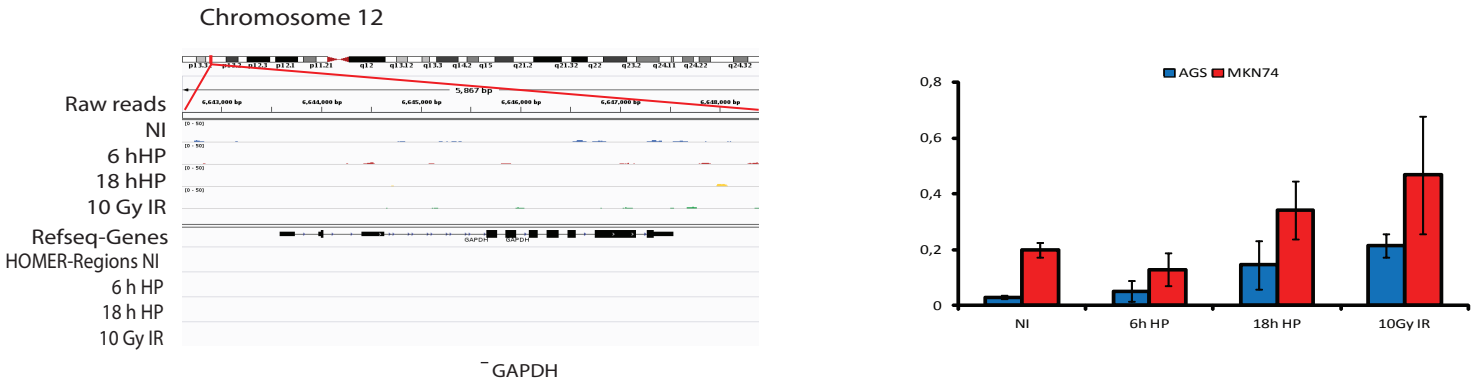
Supplemental Fig S3

A

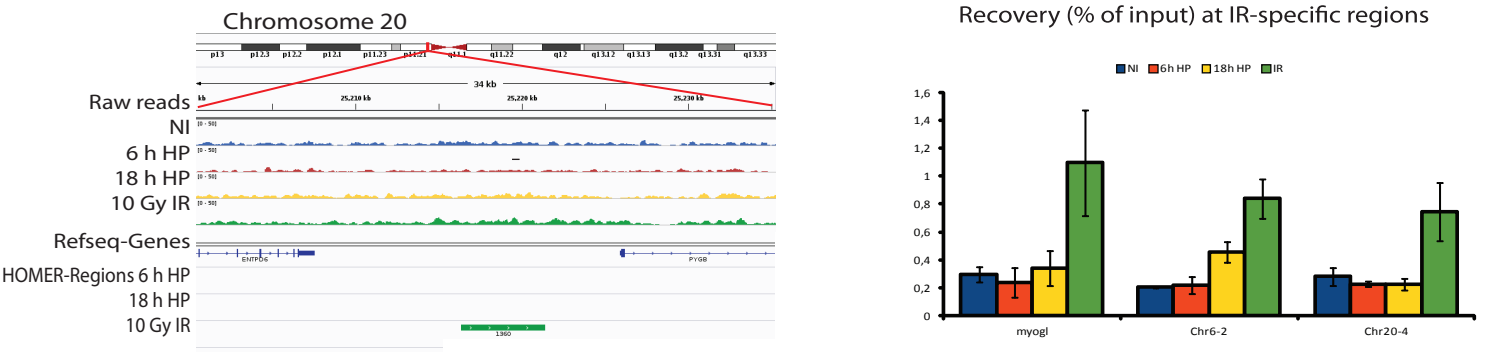


B





C



D

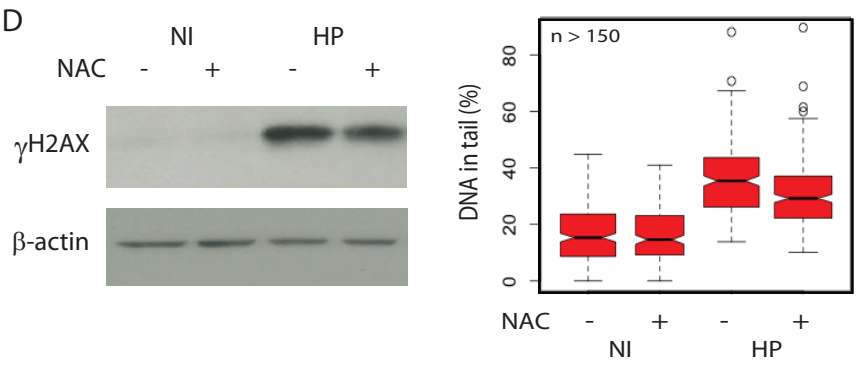


Figure S3, Related to Fig. 4

A) IGV-browser indicating the observed damage at a telomere proximal site on chr 8 for the individual conditions. Below the graph the location of primers used for ChIP-qPCR validations are indicated. B) IGV-browser (left panels) shows two control regions without significant damage used as negative control regions in ChIP-qPCR experiments. ChIP-qPCR (right panel) shows the individual damage that occurred at the two control regions during infection or after irradiation of AGS (blue) and MKN74 cells (red); error bars indicate S.D. from three independent experiments. C) IGV-browser shows an example of an IR-specific region (left panel). The validation of two IR specific regions on chr6 and chr20 by ChIP-qPCR is shown in the right panel; Error bars represent S.D. derived from three independent experiments. D) Limited influence of ROS-quenching on the occurrence of *H. pylori* induced DNA-damage. AGS cells were either infected or non-infected for 18 h with strain P12 at MOI 50 and treated with 5 mM NAC where indicated for the duration of infection. Western blot shows only a minor reduction of γH2AX upon ROS-quenching (left). Comet assay shows a similar slight reduction in the amount of broken DNA after infection and treatment with NAC (right).

Uniquely mapped reads		
Condition	**Replicate 1**	**Replicate 2**
Non-infected	26736362	36036351
6 h *H. pylor*i	30439962	32680690
18 h *H. pylori*	26028966	26597785
10 Gy IR	31250375	30835875
Total number of HOMER called regions		
Condition	**Replicate 1**	**Replicate 2**
Non-infected	2583	2944
6 h *H. pylori*	2873	2675
18 h *H. pylori*	2882	2596
10 Gy IR	1261	1457
Reproducibility according to ENCODE		
Condition	**% top40 in Repl. 2**	**% top40 in Repl. 1**
Non-infected	63.8086	73.7787
6 h *H. pylori*	68.0451	78.3835
18 h *H. pylori*	80.8285	62.8131
10 Gy IR	51.7361	56.5972
Combined enriched peaks		
Condition		
6 h *H. pylori*	431	
18 h *H. pylori*	1580	
10 Gy IR	536	

Table S1, Related to Fig. 3

Reproducibility between biological replicates

Control regions	
Primer name	**Sequence**
myoE2 Fw	CTCATGATGCCCCTTCTTCT
myoE2 Rv	GAAGGCGTCTGAGGACTTAAA
GAPDH Fw	ATGACCCCTTCATTGACCTC
GAPDH Rv	GGGGAATACGTGAGGGTATG
chr6-2_F	GGGACTCCTCAAGTTCTTCG
chr6-2_R	GGGAGTTAGATTTGGCATGG
chr20-4_F	TCGATTGCTGGTGTAATGGT
chr20-4_R	CCAAAGCCCTCTCAGTAAGG
Specific regions	
Primer name	**Sequence**
chr6-1F	TGGCACTAAGAACCAAGCAG
chr6-1R	ATTTGGGTAGGGTGGAAATG
chr8-1F	CCTCAGAGTCACACCTCCAG
chr8-1R	GACTACCTGGCCACCTCCTA
chr8-2F	CAGCCAGACTGCTCTCTCAG
chr8-2R	AGCCCTGTGCAACGATTACT
chr8-3F	GTGCTGCAGTCAGAGAGGAG
chr8-3R	AGGAGACGGTACCTCACAGG
chr8-4F	TTCCAGGGAAAATGCACTAA
chr8-4R	GGCTCTTTCTGTCCTTGACC
chr10-1F	CAGCGTAGCTCAAGCTGTTT
chr10-1R	GAGGGTTCAGTGAGGAAGGA
chr15-1F	TTTGCCTGAGGTCTTGTCAG
chr15-1R	CTGGGTCCTTATCTGAGGATG
chr16-1F	AGTGGCTTTAGCCGTCTGTT
chr16-1R	AGGTGTGCGTGTGTTTTAGC
FANCA_1Fw	GGAGCGTAGAACTGTGGTGA
FANCA_1Rv	CAGCCGACAGACAGGAAAAC
KDM5A_1Fw	GGAACCAGTCTTTGCAGAGC
KDM5A_1Rv	GCCATGCATTCTCTCAGAGC
AXIN1_1Fw	CAGCGTCGTGTGTACAACTT
AXIN1_1Rv	GCCCCGTCATGAGAGAAGAT
TFDP1_1Fw	CAGTGAGGGAGAAAGGAGGG
TFDP1_1Rv	GGAGGAGGTCTTACACAGCG
POLE_1Fw	ACACTGTAACGGCCTCCAAA
POLE_1Rv	TCATTGAGGGCCACAGAGAG
ZMYND8_1Fw	CCCCAAGTGCCCCTACTATT
ZMYND8_1Rv	TTGCTGTGCGATGATCCTCA
TRIM7_1Fw	GACTGAAACACAAGCTCCGG
TRIM7_1Rv	CAGTCCCCAGTCTCTCCTTG

Table S2, Related to Fig. 4

Primers used for ChIP

Supplemental Experimental Procedures

Cell culture, bacterial strains, infections and treatments

AGS and MKN74 cells were grown in RPMI supplemented with 10% heat inactivated FCS. The *H. pylori* strains P12 (strain collection number no 243) and its derivatives P12ΔPAI (no P387), P12ΔVacA (no P216) and P12ΔCagA (no 378) were generated and grown on GC agar plates as described by Koch et al (2012). The cells were washed once with RPMI containing 0.1% heat inactivated FCS, two h prior to infection. Infections were done under the same serum conditions for the indicated times using the same multiplicity of infection (MOI). Irradiation was performed with a Gammacell 40 Exactor (Cs-137 source; Nordion) with a dose of 10 Gy, followed by 90 min recovery, while treatment with 250 µM H_2O_2, 10 mM Hydroxy-Urea or 10 µM etoposide dissolved in DMSO was performed for 18-20 h, respectively, under identical conditions as infections. ROS-quenching was done with 5 mM N-acetyl-cystein (NAC), added to the infection medium 2 h prior to infection. For knockdown experiments cells were either transfected with different siRNAs against SHP2 or control siRNA and cultured further for 60 h prior to infection. 18-20 h p.i. proteins were harvested for western blotting.

Primary cell culture

Human primary gastric epithelial cells were isolated and cultured as described elsewhere (Schlaermann et al., 2014). Briefly, gastric glands were isolated from tissue removed during routine procedures at Berlin Charité University Hospital. After removal of fat and connective tissue, samples were minced, washed and glands were obtained through gentle squeezing of material between glass slides. Isolated gastric glands were grown as spheroids in Matrigel (growth factor reduced, phenol red free; BD Biosciences) overlaid with specific primary cell medium and passaged every 10-21 days at a ratio of 1:8. For 2D culture and infection, sheared spheroids were seeded into collagen-coated wells (type 1; 10 µg/cm^2; Sigma) plates for infection with *H. pylori* or irradiation.

Western blot

Samples for western blotting were scraped in RIPA (50 mM Tris-HCl pH 7.4, 300 mM NaCl; 5 mM EDTA; 1% Triton-X-100) containing protease inhibitor cocktail (Complete, Roche). One volume of 4 x SDS sample buffer (250 mM Tris-HCl pH 6.8, 8% SDS; 40% glycerol; 0.3 M DTT) was added and samples were boiled prior to separation of proteins on 12% SDS-PAGE. After transfer of proteins, the PVDF-membranes were blocked in 3% non-fat dry milk in TBS/0.05% Tween-20. The following antibodies were used: rabbit anti-γH2AX, 1:1,000 (2577; Cell Signaling), rabbit anti-phospho-MRE11, 1:1,000 (4859; Cell Signaling), rabbit anti-NBS1, 1:2,000 (ab32074; Abcam), rabbit anti-phospho-ATR, 1:1,000 (2853; Cell Signaling), rabbit anti-phospho-ATRIP, 1:1,000 (5161; Cell Signaling), mouse anti-β-actin, 1:20,000 (A5411: Sigma), HRP-linked sheep anti-mouse, 1:3,000 (NA931V; Amersham); HRP-linked donkey anti-rabbit, 1:3,000 (NA934V; Amersham). ECL Western blotting substrate was detected by Hyperfilm ECL (Amersham).

Comet Assay

Assessing the percentage of broken DNA was done by neutral comet assay using the CometAssay-Kit (4250-050-K) from Trevigen. In brief, cells were washed once and collected in PBS, prior to immobilization in low melting agarose. After lysis, fragmented DNA was subjected to electrophoresis for 1 h at 15 V. Staining of DNA was achieved with SYBR-green solution, prior to detection of comet tails by epifluorescence microscopy. Analysis of DNA in each tail was performed using the CometScore software from TriTek.

Immunofluorescence

Cells were seeded on autoclaved coverslips prior to infection or treatment. After washing in PBS, cells were fixed with 4% paraformaldehyde for 15 min at room temperature followed by three washes in PBS/0.1% Tween-20. Cells were permeabilized by incubation with PBS/0.5% Triton X-100 for 1 h at RT and afterwards blocked with 3% BSA in PBS-T for 30 min at RT. Incubation with antibodies was carried out in 3% PBS-T using mouse anti-γH2AX-ab 1:100 (05-636; Millipore); rabbit anti-53BP1-ab 1:100 (PA1-16565; ThermoScientific); mouse anti-PML 1:100 (sc-966; Santa Cruz); mouse anti-Oct1 1:25 (ab51363; Abcam); secondary goat anti-mouse Cy2 (115-096-003; Dianova) and secondary donkey anti-rabbit Cy3 (711-105-152; Dianova).

ChIP

The chromatin immunoprecipitation (ChIP) was done as described previously (Denissov et al., 2007). For precipitations 2 µg of mouse anti-γH2AX-ab (05-636; Millipore) were used per reaction. Real-time qPCR was performed using the SYBR Green mix (Applied Biosciences) with the Abi-Prism machine (Applied Biosciences). Primers are listed in Supplemental Table S2.

ChIP-seq

Sequencing libraries from 20 ng of precipitated DNA were prepared with the Nugen Ovation Ultralow-Kit (Nugen Part no 0330, 0331), according to manufacturer's instructions. All samples were sequenced on an Illumina HiSeq2500 at the Max Planck Genomic Centre Cologne, Germany, with 100 bp single-end reads. Prior to mapping, raw data was quality-controlled with Qualimap (Garcia-Alcalde et al., 2012).

Data-analysis: Region calling by HOMER

Identification of damaged regions was performed on each replicate with HOMER (Heinz et al., 2010) setting a peak-size of 2,500 bp and a minimal distance between regions to 5,000 bp. Differentially enriched regions were identified using HOMER's getDifferentialPeaks-tool and selecting regions presenting a fold-change of > 1.5 and a p-value of < 0.001 and aggregating replicates of the same condition afterwards.

Identification of cancer related genes in the expressed genes showing *H. pylori*-induced damage

To test whether host genomic regions in which gene expression co-occurs with DNA damage might have a functional relation to known genes implicated in human cancer, we combined available lists of known cancer genes from Futreal et al. (2004), Kandoth et al. (2013) and Tamborero et al. (2013) to obtain a list of 729 non-redundant genomic locations harboring such genes. To test for statistical significance we performed hypergeometric distribution testing with multiple Monte Carlo simulations.

RNA-isolation and RNA-seq

For analysis of the transcriptional response of primary human gastric cells, total RNA was extracted from infected or uninfected cells using the Qiagen AllPrep-Kit according to protocol. After DNAse-treatment (Invitrogen) and Bioanalyzer quality control (Agilent Technologies), mRNA was enriched using oligo-dT beads, followed by strand-specific library-preparation (Parkhomchuk et al., 2009). A size-selected library (300bp) was sequenced from both ends on a HiSeq2000-platform (Illumina) to at least 38 mio reads for each sample. Mapping of the derived reads was performed using TopHat2 and Cufflinks for analysis of expressed genes (Trapnell et al., 2012). Generation and analysis of RNA-seq data in AGS cells will be described extensively elsewhere (F. Glowinski and T.F. Meyer, in preparation).

Overlap with additional genomic data-sets

To overlay region-specific signals with active transcription we defined different subgroups of expressed genes via fragments per kilobase of transcript length per million sequenced reads (FPKM), as calculated by Cufflinks (Trapnell et al., 2012). The resulting genes were binned according to their expression level described in Fig. 3D. Hypergeometric distribution testing was performed using the 'Statistical analysis of tracks'-function of the Galaxy Hyperbrowser (Sandve et al., 2010) to analyze whether the damaged regions under each condition were located inside of all expressed genes taken together at a higher proportion than expected by random overlap. The genomic loci of genes in the different expression groups were then intersected with the damage-enriched regions identified. Genomic locations of common fragile sites (CFS) and early replicating fragile sites (ERFS) were obtained from Crosetto et al. (2013) and Barlow et al., (2013), respectively. Mouse genomic coordinates from ERFS were transferred to human coordinates using the UCSC-browser.

Supplemental References

Barlow, J.H., Faryabi, R.B., Callen, E., Wong, N., Malhowski, A., Chen, H.T., Gutierrez-Cruz, G., Sun, H.W., McKinnon, P., Wright, G.*, et al.* (2013). Identification of early replicating fragile sites that contribute to genome instability. Cell *152*, 620-632.

Crosetto, N., Mitra, A., Silva, M.J., Bienko, M., Dojer, N., Wang, Q., Karaca, E., Chiarle, R., Skrzypczak, M., Ginalski, K.*, et al.* (2013). Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. Nat Methods *10*, 361-365.

Denissov, S., van Driel, M., Voit, R., Hekkelman, M., Hulsen, T., Hernandez, N., Grummt, I., Wehrens, R., and Stunnenberg, H. (2007). Identification of novel functional TBP-binding sites and general factor repertoires. Embo j *26*, 944-954.

Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., and Stratton, M.R. (2004). A census of human cancer genes. Nat Rev Cancer *4*, 177-183.

Garcia-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L.M., Gotz, S., Tarazona, S., Dopazo, J., Meyer, T.F., and Conesa, A. (2012). Qualimap: evaluating next-generation sequencing alignment data. Bioinformatics *28*, 2678-2679.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576-589.

Kandoth, C., McLellan, M.D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J.F., Wyczalkowski, M.A.*, et al.* (2013). Mutational landscape and significance across 12 major cancer types. Nature *502*, 333-339.

Koch, M., Mollenkopf, H.J., Klemm, U., and Meyer, T.F. (2012). Induction of microRNA-155 is TLR- and type IV secretion system-dependent in macrophages and inhibits DNA-damage induced apoptosis. Proc Natl Acad Sci U S A *109*, E1153-1162.

Parkhomchuk, D., Borodina, T., Amstislavskiy, V., Banaru, M., Hallen, L., Krobitsch, S., Lehrach, H., and Soldatov, A. (2009). Transcriptome analysis by strand-specific sequencing of complementary DNA. Nucleic Acids Res *37*, e123.

Sandve, G.K., Gundersen, S., Rydbeck, H., Glad, I.K., Holden, L., Holden, M., Liestol, K., Clancy, T., Ferkingstad, E., Johansen, M.*, et al.* (2010). The Genomic HyperBrowser: inferential genomics at the sequence level. Genome Biol *11*, R121.

Schlaermann, P., Toelle, B., Berger, H., Schmidt, S.C., Glanemann, M., Ordemann, J., Bartfeld, S., Mollenkopf, H.J., and Meyer, T.F. (2014). A novel human gastric primary cell culture system for modelling Helicobacter pylori infection in vitro.

Tamborero, D., Gonzalez-Perez, A., Perez-Llamas, C., Deu-Pons, J., Kandoth, C., Reimand, J., Lawrence, M.S., Getz, G., Bader, G.D., Ding, L.*, et al.* (2013). Comprehensive identification of mutational cancer driver genes across 12 tumor types. Sci Rep *3*, 2650.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc *7*, 562-578.