Supplementary Information for:

Modeling genomic diversity and tumor dependency in malignant melanoma

Updated information pertaining to microarray data processing may be found at the following url: http://www.broad.mit.edu/melanoma

Supplementary Methods

Cultured melanoma cells

The melanoma short-term cultures (STCs) employed in this study were all derived from metastatic disease. In most cases, their melanocytic origins were confirmed by histological or molecular studies.

High-density SNP array hybridization and data processing

Genomic DNA from all samples was tracked with ABgene 2D barcode rack and single tube readers (ABgene). The well designation for samples was randomized, and in each plate there were 6 control wells that contained normal diploid DNA. After the genomic DNA was cut by *StyI* restriction enzyme (New England Biolabs), the DNA was ligated to an adaptor with a T4 ligase (New England Biolabs). PCR amplification was then done using an Applied Biosystems 9700 Thermal Cycler I and Titanium Taq (Clontech) to obtain 200-110 bp fragments. After the amplified DNA was pooled and concentrated, it was fragmented using *DNaseI* (Affymetrix, Inc.). The DNA was then end-labeled, denatured and hybridized to the probe sets in each well of the plate and washed. DNA was then hybridized to the microarrays according to the manufacturer's directions. 50K *XbaI* SNP hybridization was performed as described previously.

Using the SNPFileCreator GenePattern module, a .snp file was created by normalizing intensities and modeling using PM-MM difference modeling method (Li, 2001). Each .snp file contained normalized raw intensities for samples along with SNP marker information and chromosomal position (based upon Affymetrix annotations and hg17 build of the human genome sequence from the University of California, Santa Cruz: http://genome.ucsc.edu/).

Next, a series of manipulations were implemented to increase the signal–to-noise ratio. First, melanoma samples and normal samples from each of the .snp files were selected, sorted by chromosomal position, and merged into a single matrix after median adjusted batch-correction. In order to convert raw intensity data into copy-number values, normal samples with the highest similarity to a given tumor sample were chosen by median $\log_2$-transformed Euclidean distance across the corresponding markers. To remove normal samples with sub-optimal data quality, a Gaussian histogram analysis was generated so that copy number distribution could be examined. Here, high-quality normal samples typically show a single narrow peak centered at two copies, whereas hyperploid tumor samples typically have multiple peaks with a broader signal range (this method removes

samples with significant stromal admixture but may also exclude pure tumors that lack copy number abnormalities). The copy number data for each tumor sample was then re-computed, referencing the high-quality normal samples. Heatmap visualization of SNP array data uses a beta version of the GenePattern SNP visualizer (A full-version is expected soon, http://www.broad.mit.edu/cancer/software/genepattern/index.html).

Next, the GLAD segmentation algorithm (1) was employed to smooth the copy number data with default parameters. Noisy samples reflecting technical artifacts were removed at this step using a distribution metric of the segmented data. SNP markers in regions of documented copy-number variants (http://projects.tcag.ca/variation/, October 2006) were removed from subsequent analysis, thereby reducing the density of markers. To avoid bias from using multiple clones from a single patient, such samples (where present) were identified through a genotype comparison. Next, samples with the least noise from among the clones were chosen based on noise between neighboring probes. In combining 250K and 50K data sets, both data sets were aligned by genomic position, and missing cells in the matrix were filled in with the mean of neighboring values. All source code was written in MATLAB R2006b. Further details of processing can be found in a forthcoming paper by Beroukhim *et al.* (manuscript submitted).

RNA preparation and gene expression profiling

RNA was extracted from melanoma short term cultures grown to 80% confluency with 1.0 mL of Trizol Reagent with 200 µL of chloroform. After shaking and incubation at 30°C for 10 minutes, the samples were centrifuged at 13,000 RPM for 10 minutes at 2°C. The supernatant was subsequently removed and the RNA pellet was washed with 1.0 mL of 75% ethanol. After another centrifugation, the ethanol supernatant was removed and the pellet was dissolved in DEPC-treated water. Following a final incubation for 5 minutes at 55°C, the amount of RNA was quantified with a spectrophotometer. Use of the Affymetrix HT-HGU133A chips allowed all RNAs to be profiled within a single experiment, thereby minimizing batch effects and other technical concerns that might adversely affect data quality.

Tiling array CGH data processing

The tiling array CGH was downloaded from the GEO database (GSE6779). Log_2-ratios were calculated as follows: first, Cy3 and Cy5 channel values were determined. In the event that Cy3 (signal-background) fell below the median Cy3 background, the Cy3 (signal-background) was given the median Cy3 background value. This was similarly done for the Cy5 channel. Log_2-ratios were obtained by dividing the Cy3 (signal - background) by the Cy5 (signal - background) divided by the (median Cy3 background divided by the Cy5 background). Missing values were replaced with the mean of the closest two neighboring markers. Finally, the log_2-ratios were median-column normalized. Samples were then smoothed using the GLAD segmentation algorithm with default parameters. Markers in regions of known copy number polymorphisms were removed (http://projects.tcag.ca/variation/, October 2006) as in the other data sets.

SNP array and CGH data set comparisons

Q-value distributions determined by GISTIC were used as a basis for quantitative comparison by Pearson's-correlation coefficient. Since the various data sets employed had different numbers of markers, a composite marker set spanning the genome was first created from the combined data sets. Next, each q-value distribution was mapped to its original genomic location. Regions of the q-value distribution that were not reflected in the original data set were then filled in with neighboring original values. For the data set

comparisons, this process was performed on both q-value distributions, thus creating two distributions of equal size mapped by genomic position (see Supplementary Table 2). In our intra-data set comparison, a random number generator was used to assign individual samples to one of two subgroups. This randomization was repeated 100 times, GISTIC analysis was performed on each subgroup and the median Pearson's correlation coefficient was taken. The Pearson's correlation test was chosen to focus on the genomic landscape rather than the absolute q-value differences, thereby allowing fairer comparisons between data sets with different sample sizes and markers numbers. Similar trends were observed when the –log(q-value) or G score were used as the metric, as well as a distribution of medians and means of all samples at a particular marker. Source code was written in MATLAB R2006b.

Inferring loss-of-heterozygosity

LOH calls were inferred on 86 normal samples (the same normal samples chosen as most similar to compute copy number) using dChipSNP software. This data was then used to infer LOH regions for tumor samples with a 0.5 inferred LOH call rate and removing LOH regions consistent with more than 5% of reference samples. The resulting data matrix was extracted from dChip and loaded into MATLAB, where the GISTIC algorithm was performed on the inferred LOH matrix.

Western blotting and biochemical studies

Cells were lysed in TNN buffer containing a mini complete protease inhibitor tablet (Roche), 1 mM NaF, and 1 mM $NaVO_3$. Lysate concentration was determined via Bradford assay; protein was then prepared in SDS gel loading buffer and boiled for 8 minutes at 95°C. Lysates of 60 µg of protein were resolved by denaturing gel electrophoresis using the BioRad Minigel system with 10% Tris-HCl gels (BioRad, Catalog Number: 161-1158). Protein was transferred to nitrocellulose membranes (80V for 80 minutes at 4°C) and probed using primary antibodies against p-Erk (anti-phospho-p44/42), total ERK (anti-p44/42), p-MEK (anti-phospho-MEK1/2 (Ser217/221)), total MEK (anti-MEK1/2) (Cell Signaling Technology), Cyclin D1 (sc-20044, Santa Cruz Biotechnology), and alpha-tubulin (anti-alpha-tubulin, Cell Signaling Technology). All primary antibodies were used at a 1:1000 dilution. The secondary antibody was anti-rabbit IgG, HRP-linked (Cell Signaling Technology) at a 1:1000 dilution. Biochemical MEK inhibition by CI-1040 was examined by treating sub-confluent melanoma cells with varying dilutions of CI-1040 for 24 hours. After harvest and protein preparation, Western blotting was performed with the p-ERK and total ERK antibodies described above.

Supplementary Figure Legends

Figure S1: GISTIC comparative plots of 250K SNP vs 50K SNP and tiling array data
Top: statistically significant amplifications **(A)** (orange, red) and deletions **(B)** (light blue, dark blue) by GISTIC analysis, comparing the 50K *XbaI* (n=31) and 250K *StyI* (n=70) arrays, respectively. Bottom: amplifications **(C)** and deletions **(D)** for the Jonsson *et al.* tiling array CGH cell line data set (n=45) were compared to the 250K *StyI* array data set. Left axes show genomic position; top axes show FDR q-values and bottom axes show corresponding *G-scores*. FDR threshold = 0.25.

Figure S2: Genomic distinctions and similarities between melanoma clinical subtypes
Significant amplifications (red) and deletions (blue) by GISTIC are shown for the Curtin et al. melanoma CGH data set(2), including 30 chronic sun induced samples **(A),** 40 non-chronic sun induced samples **(B)**, 36 acral melanoma samples **(C)**, and 20 mucosal melanoma samples **(D)**. Left axes show genomic position; top axes show FDR q-values and bottom axes show corresponding *G-scores*. FDR threshold = 0.25. Known oncogene or tumor suppressor genes are labeled to right of the corresponding peak.

Figure S3: Chromosomal copy number changes at the *BRAF* and *PTEN* loci
Heatmap views of the *PTEN* and *BRAF* locus with samples (columns) sorted by level of deletion (blue) or amplification (red) are shown for the Jonsson 33K tiling array CGH data set (n=45) (**(A)** and **(D)**, respectively) and the 2K Curtin *et al.* BAC array CGH data set (n=70) (**(B)** and **(E)**, respectively). The *BRAF* locus in the 250K and 50K data sets is also shown **(C)** (n=101). Genes at the respective loci are indicated (Y axis).

Figure S4: Phospho-ERK inhibition by the MEK inhibitor CI-1040
Immunoblots show p-ERK, total ERK, and cyclin D levels in selected melanoma short term cultures following incubation with increasing CI-1040 concentrations. The MALME-3M melanoma cell line (MALME) is included as a control. Mutation status of BRAF and NRAS are denoted, along with the CI-1040 GI_{50}.

Figure S5: Chromosomal correlates of MAP kinase dependency
A whole genome copy number heatmap is shown (red=copy gains; blue=copy losses) for selected short term cultures sorted by CI-1040 GI_{50} values. BRAF and NRAS mutation status is indicated.

Figure S6: Matched expression and copy number for chromosome 10 outliers
Gene expression values are plotted for CUL2 **(A)** and KLF6 **(B),** sorted by chromosomal hemizygous deletion or retention. These genes were identified by whole genome SAM analysis (Fig. 3D of main text).

Supplementary Table 1: BRAF, NRAS, and PTEN mutation status

Sample Name	Sample Type	SNP array platform	HTA Gene Expression data	BRAF	NRAS	PTEN*	Other
501MEL	Cell line	50K *XbaI*	yes	WT	G12D	no data	
A375	Cell line	50K *XbaI*	yes	V600E	WT	no data	
HS944	Cell line	250K *StyI*	no	no data	Q61K**	Del exon 2; frameshift/premature stop**	
KO13	Short term culture	50K *XbaI*	no	no data	no data	no data	
KO16	Short term culture	50K *XbaI*	no	V600E	WT	WT	
KO18	Short term culture	50K *XbaI*	no	WT	WT	WT	
KO25	Short term culture	50K *XbaI*	no	no data	no data	no data	
KO27	Short term culture	50K *XbaI*	no	V600E	WT	no data	
KO28	Short term culture	50K *XbaI*	no	V600E	WT	WT	PIK3CA E542K
KO29AX	Short term culture	50K *XbaI*	no	V600E	WT	WT	
KO33b	Short term culture	50K *XbaI*	no	no data	no data	no data	
KOO6	Short term culture	50K *XbaI*	no	WT	WT	no data	
KOO8	Short term culture	50K *XbaI*	no	V600E	WT	WT	
LOX_IMVi	Cell line	50K *XbaI*	no	V600E	WT	no data	
M000216	Short term culture	250K *StyI*	yes	V600E	WT	WT	
M000301	Short term culture	250K *StyI*	yes	WT	Q61K	WT	
M000907	Short term culture	-	yes	WT	G13R	WT	
M000921	Short term culture	250K *StyI*	yes	V600E	WT	T319fs*1	
M010124	Short term culture	250K *StyI*	yes	WT	Q61K	WT	
M010308	Short term culture	250K *StyI*	yes	WT	Q61K	WT	
M010322	Short term culture	-	yes	V600E	WT	WT	
M010403	Short term culture	250K *StyI*	yes	no data	no data	WT	
M010606	Short term culture	250K *StyI*	yes	WT	WT	WT	
M010718	Short term culture	-	yes	WT	Q61K	WT	
M010817	Short term culture	250K *StyI*	yes	WT	Q61R	P213L	
M010920	Short term culture	250K *StyI*	yes	WT	WT	WT	
M13	Short term culture	50K *XbaI*	no	WT	Q61K	WT	
M14	Cell line	50K *XbaI*	no	V600E	no data	no data	
M16	Short term culture	50K *XbaI*	no	WT	WT	WT	
M19	Short term culture	50K *XbaI*	no	no data	no data	no data	
M23	Short term culture	50K *XbaI*	no	WT	Q61K	no data	
M25	Short term culture	50K *XbaI*	no	V600E	no data	no data	
M26	Short term culture	50K *XbaI*	no	WT	WT	WT	

M34	Short term culture	50K *XbaI*	no	V600E	WT	K164fs*16	
M970109	Short term culture	-	yes	V600E	WT	WT	
M970131	Short term culture	-	yes	WT	WT	WT	
M970805	Short term culture	-	yes	WT	WT	WT	
M980409	Short term culture	-	yes	WT	Q61K	WT	
M980513	Short term culture	-	yes	no data	no data	WT	
M980928	Short term culture	250K *StyI*	yes	WT	WT	WT	
M981201	Short term culture	-	yes	WT	Q61K	WT	
M990114	Short term culture	250K *StyI*	no	WT	WT	WT	
M990514	Short term culture	-	yes	no data	no data	WT	
M990719	Short term culture	250K *StyI*	yes	WT	G13R	no data	
M990802	Short term culture	250K *StyI*	yes	V600E	WT	WT	
M991121	Short term culture	-	yes	WT	WT	WT	
MALME3M	Cell line	50K *XbaI*	yes	V600E	WT	no data	
MCF7	Breast cancer cell line	-	yes	no data	no data	no data	
MDAMB435	Cell line	50K *XbaI*	no	V600E	no data	no data	
MEL35	Short term culture	50K *XbaI*	no	WT	Q61R	no data	
MEL9	Short term culture	50K *XbaI*	no	no data	no data	no data	
MelanocyteW2	Normal melanocyte	-	yes	no data	no data	no data	
MelanocyteW3	Normal melanocyte	-	yes	no data	no data	no data	
MelanocyteW4	Normal melanocyte	-	yes	no data	no data	no data	
MelanocyteW6	Normal melanocyte	-	yes	no data	no data	no data	
MelanocyteW7	Normal melanocyte	-	yes	no data	no data	no data	
MeWo	Cell line	250K *StyI*	no	WT*	no data	no data	
ML310	Short term culture	-	yes	no data	no data	no data	
SK-MEL-119	Cell line	250K *StyI*	no	no data	Q61R**	WT*	
SK-MEL-19	Cell line	250K *StyI*	no	V600E**	no data	no data	
SK-MEL-2	Cell line	50K *XbaI*	yes	WT	Q61R	no data	
SKMEL28	Cell line	50K *XbaI*	yes	V600E	WT	T167A*	CDK4 R24C
SK-MEL-30	Cell line	250K *StyI*	no	no data	Q61K**	WT*	
SKMEL5	Cell line	50K *XbaI*	no	V600E	WT	no data	
SK-MEL-63	Cell line	250K *StyI*	no	no data	Q61K**	WT*	
UACC257	Cell line	50K *XbaI*	no	V600E	WT	no data	
UACC62	Cell line	50K *XbaI*	no	V600E	WT	no data	
WM1026	Short term culture	250K *StyI*	yes	WT	WT	WT	
WM1325	Short term culture	250K *StyI*	yes	no data	no data	no data	
WM1346	Short term culture	250K *StyI*	yes	no data	no data	no data	
WM1385	Short term culture	250K *StyI*	yes	no data	no data	no data	

WM1419	Short term culture	250K *StyI*	yes	no data	no data	no data
WM1433	Short term culture	250K *StyI*	yes	no data	no data	no data
WM1480	Short term culture	250K *StyI*	yes	no data	no data	no data
WM1575	Short term culture	-	yes	no data	no data	no data
WM1716	Short term culture	250K *StyI*	yes	V600E	WT	no data
WM1719	Short term culture	-	yes	no data	no data	no data
WM1720	Short term culture	250K *StyI*	yes	V600E	WT	WT
WM1745	Short term culture	250K *StyI*	yes	V600E	WT	no data
WM1852	Short term culture	250K *StyI*	yes	V600E	WT	WT
WM1862	Short term culture	250K *StyI*	yes	V600E	WT	no data
WM1930	Short term culture	250K *StyI*	yes	V600E	WT	WT
WM1931	Short term culture	250K *StyI*	no	V600E	WT	WT
WM1942	Short term culture	-	yes	V600E	WT	WT
WM1960	Short term culture	250K *StyI*	yes	WT	Q61K	WT
WM1963	Short term culture	-	yes	WT	WT	WT
WM1968	Short term culture	250K *StyI*	yes	no data	no data	WT
WM1976	Short term culture	250K *StyI*	yes	V600E	WT	WT
WM2029	Short term culture	250K *StyI*	yes	WT	WT	WT
WM262	Short term culture	250K *StyI*	yes	no data	no data	no data
WM3061	Short term culture	250K *StyI*	yes	WT	G13R	WT
WM3066	Short term culture	-	yes	no data	no data	no data
WM3077	Short term culture	250K *StyI*	yes	no data	no data	no data
WM3130	Short term culture	250K *StyI*	yes	K601E	WT	WT
WM3163	Short term culture	250K *StyI*	yes	V600E	WT	WT
WM3215	Short term culture	250K *StyI*	yes	V600E	WT	T319fs*1
WM3218	Short term culture	250K *StyI*	yes	V600E	WT	WT
WM3228	Short term culture	250K *StyI*	yes	V600E	WT	WT
WM3243NCI-A5	Short term culture	250K *StyI*	yes	V600E	WT	WT
WM3244NCI-B	Short term culture	-	yes	no data	no data	WT
WM3246	Short term culture	250K *StyI*	yes	WT	WT	K80_R84del
WM3259	Short term culture	250K *StyI*	yes	V600E	WT	WT (homozygous deletion)
WM3282	Short term culture	250K *StyI*	yes	V600K	WT	WT
WM3297	Short term culture	250K *StyI*	yes	V600E	WT	WT
WM3311	Short term culture	250K *StyI*	yes	WT	WT	WT
WM3314	Short term culture	-	yes	no data	no data	no data

WM3381A	Short term culture	250K *StyI*	yes	V600E	WT	no data	
WM3450	Short term culture	250K *StyI*	no	WT	Q61K	WT	
WM3451	Short term culture	-	yes	no data	no data	WT	
WM3456	Short term culture	250K *StyI*	yes	WT	Q61K	WT	
WM3457	Short term culture	250K *StyI*	yes	V600E	WT	splice+2 insG, following aa267	
WM3482	Short term culture	250K *StyI*	yes	V600E	WT	WT	
WM3506	Short term culture	250K *StyI*	yes	WT	Q61R	WT	
WM3526	Short term culture	250K *StyI*	yes	V600E	WT	WT	
WM3619	Short term culture	250K *StyI*	yes	WT	Q61R	WT	
WM3623	Short term culture	250K *StyI*	yes	WT	Q61K	WT	
WM3627	Short term culture	250K *StyI*	yes	V600E	WT	WT	
WM3629	Short term culture	250K *StyI*	yes	D594G	G12D	WT	
WM3670	Short term culture	250K *StyI*	yes	G469E	G12D	WT	
WM3682	Short term culture	-	yes	WT	WT	WT	FGFR1 S125L
WM3702	Short term culture	250K *StyI*	yes	WT	WT	no data	FGFR1 S125L
WM3727	Short term culture	250K *StyI*	yes	V600E	WT	WT	
WM451Lu	Short term culture	250K *StyI*	yes	V600E	WT	WT	
WM46	Short term culture	-	yes	no data	no data	no data	
WM806	Short term culture	250K *StyI*	no	no data	no data	no data	
WM853-2	Short term culture	250K *StyI*	yes	V600E	WT	WT	
WM858	Short term culture	250K *StyI*	yes	no data	no data	no data	
WM984	Short term culture	250K *StyI*	yes	V600E	WT	WT	
WW94	Cell line	50K *XbaI*	yes	no data	no data	no data	

*PTEN mutation nomenclature follows COSMIC with the exception of sample WM3457
**previously reported (18)

Supplementary Table 2: Genomic diversity of melanoma short term cultures

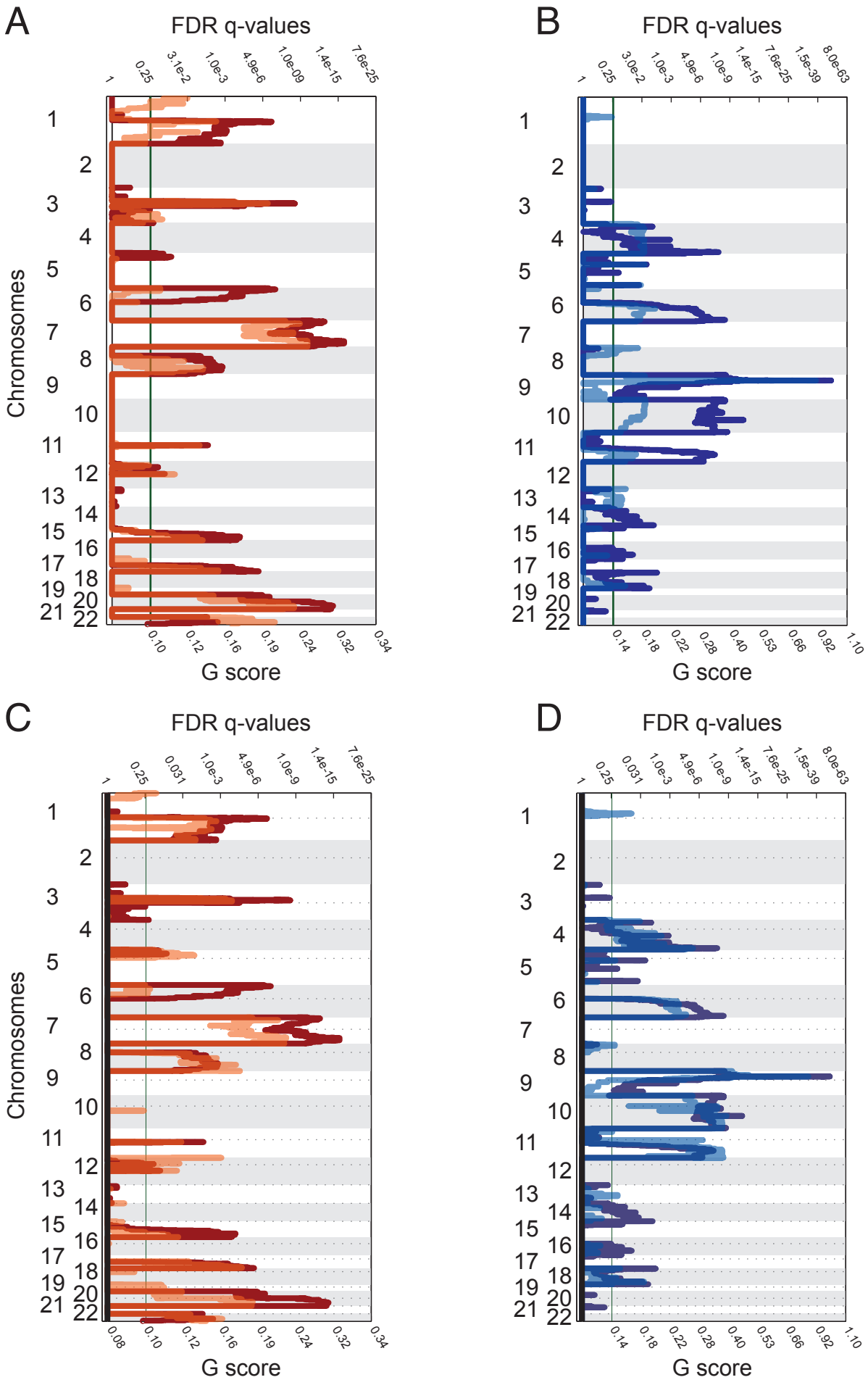
Data set 1	Data set 2	Pearson Correlation Coefficient
cultured (a)	cultured (b)	0.85
noncsd + csd	cultured	0.71
noncsd	cultured	0.66
mucosal	cultured	0.55
csd	cultured	0.52
acral	cultured	0.48
lung	noncsd + csd	0.42
lung	cultured	0.35
normal	cultured	0.016
normal	lung	0.02
normal	noncsd + csd	0.01
normal	jonsson	0.01
jonsson	cultured	0.80
jonsson	noncsd + csd	0.74
jonsson	acral	0.49
jonsson	mucosal	0.56
jonsson	csd	0.47
jonsson	noncsd	0.69

cultured	short term cultures and cell lines (n=89, 250K *StyI* and 50K Xba) (a) and (b) were randomly selected mutually exclusive subsets from 89 (see Supplementary Methods)
csd	primary chronic sun-induced damage melanoma (n=30, 2K BAC array CGH, Curtin *et al*.)
noncsd	primary non chronic sun-induced damage melanomas (n=40, 2K BAC array CGH, Curtin *et al*.)
csd + noncsd	cutaneous primary melanomas (csd and noncsd, n=70, 2K BAC array CGH, Curtin *et al*.)
acral	primary acral melanoma (n=36, 2K BAC array CGH, Curtin *et al*.)
mucosal	primary mucosal melanoma (n=20, 2K BAC array CGH, Curtin *et al*.)
lung	primary lung adenocarcinomas (n=462, 250K Sty I, Weir et al., submitted)
normal	normal tissue (n=156, 250K *StyI*)
jonsson	melanoma cell lines (n=45, 30K tiling array CGH, Jonsson et al.)

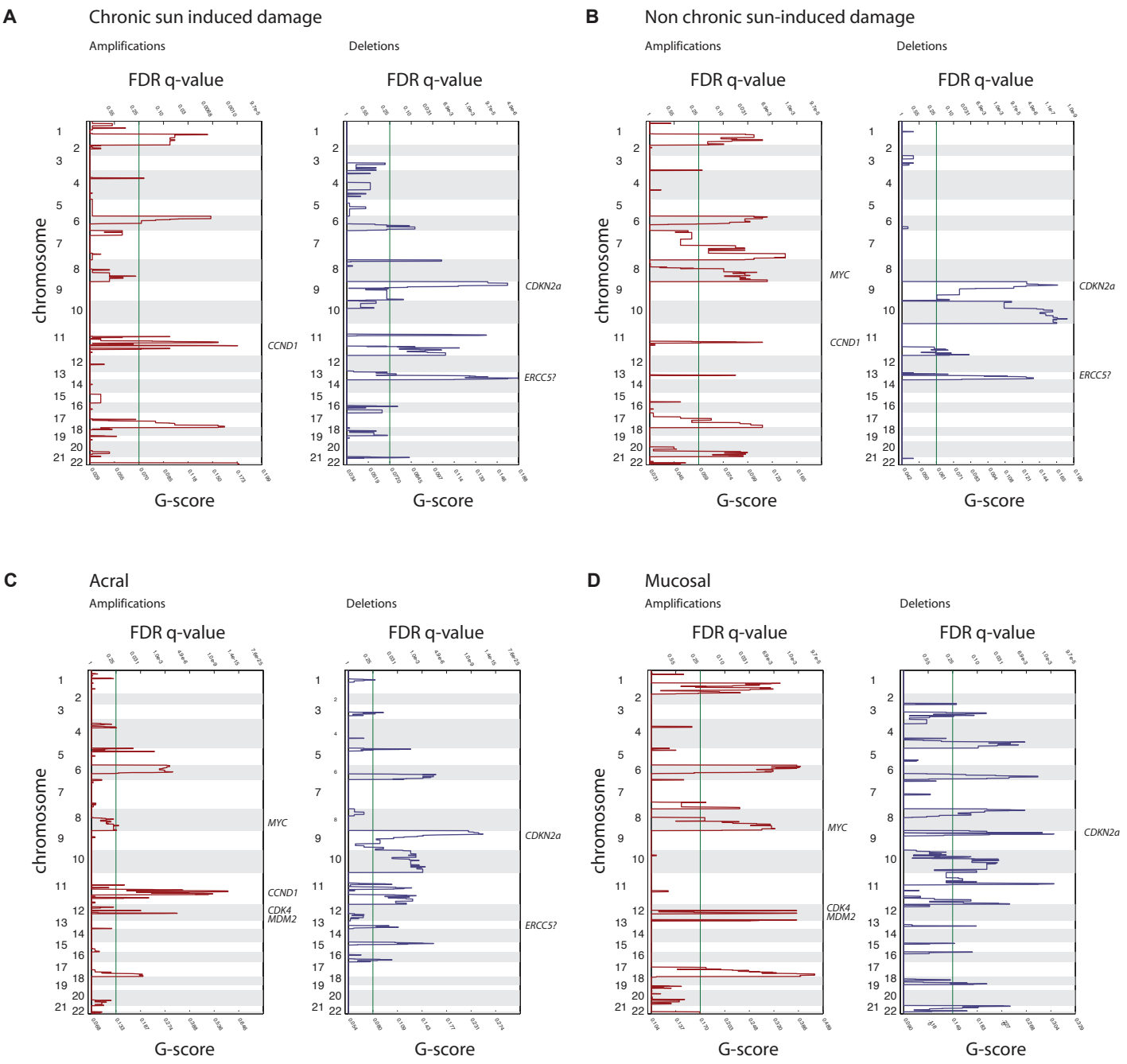
References

1. Hupe P, Stransky N, Thiery JP, Radvanyi F, Barillot E. Analysis of array CGH data: from signal ratio to gain and loss of DNA regions. Bioinformatics 2004;20:3413-3422.
2. Curtin JA, Fridlyand J, Kageshita T, et al. Distinct sets of genetic alterations in melanoma. N Engl J Med 2005;353:2135-2147.

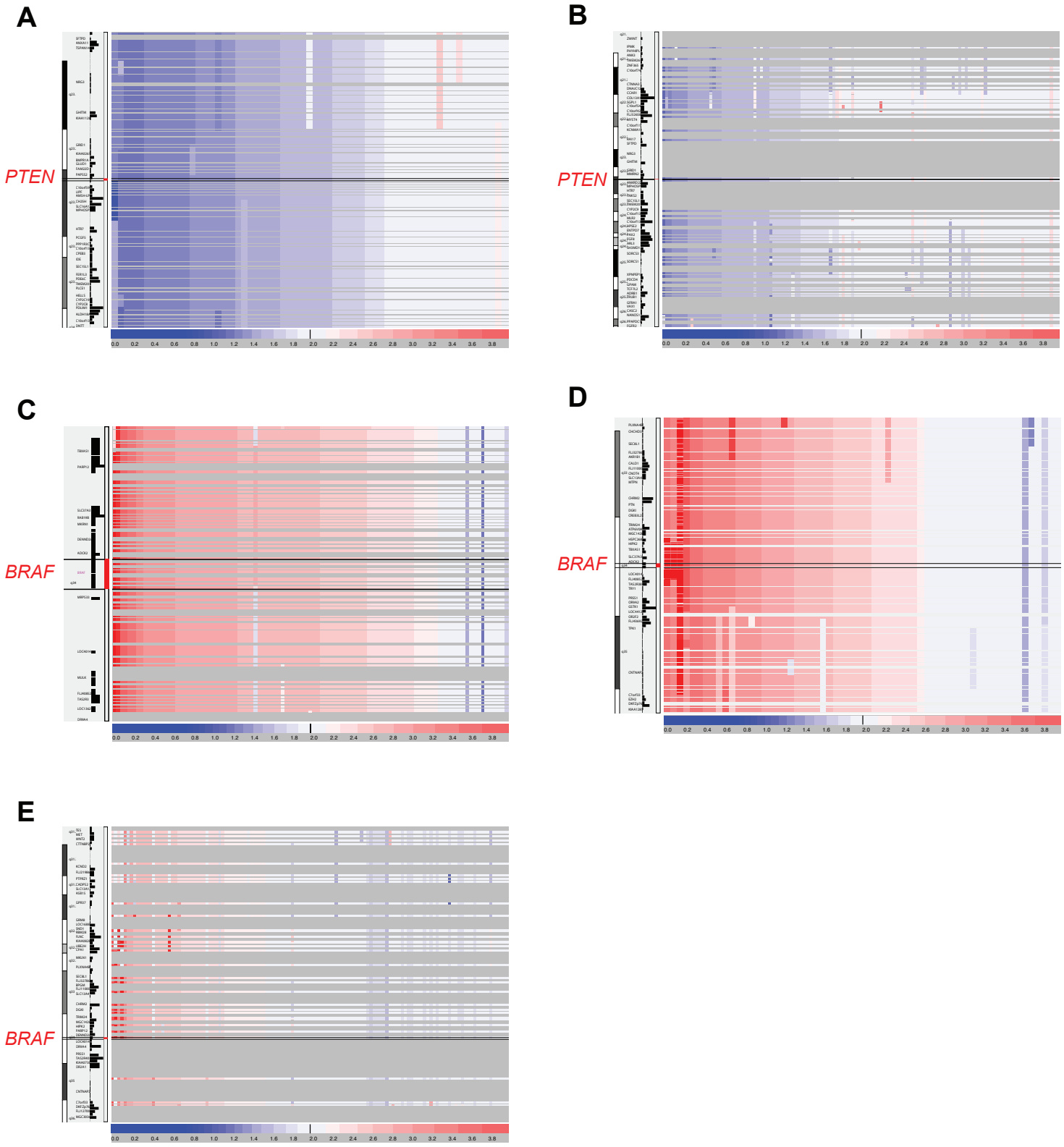
Supplementary Figure 1 - Lin *et al.*



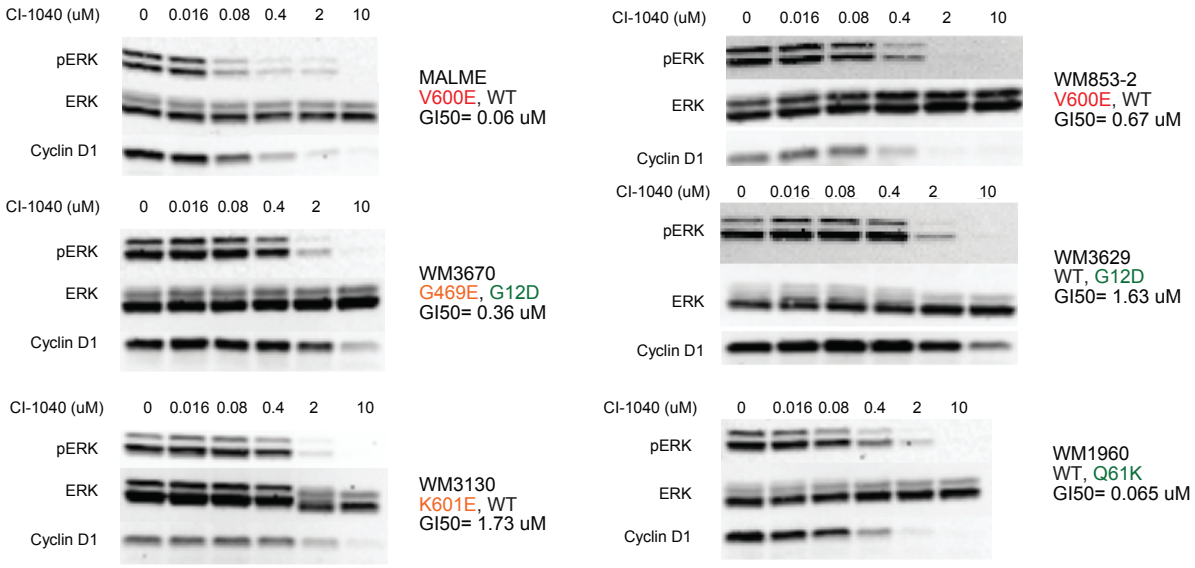
Supplementary Figure 2 - Lin *et al.*



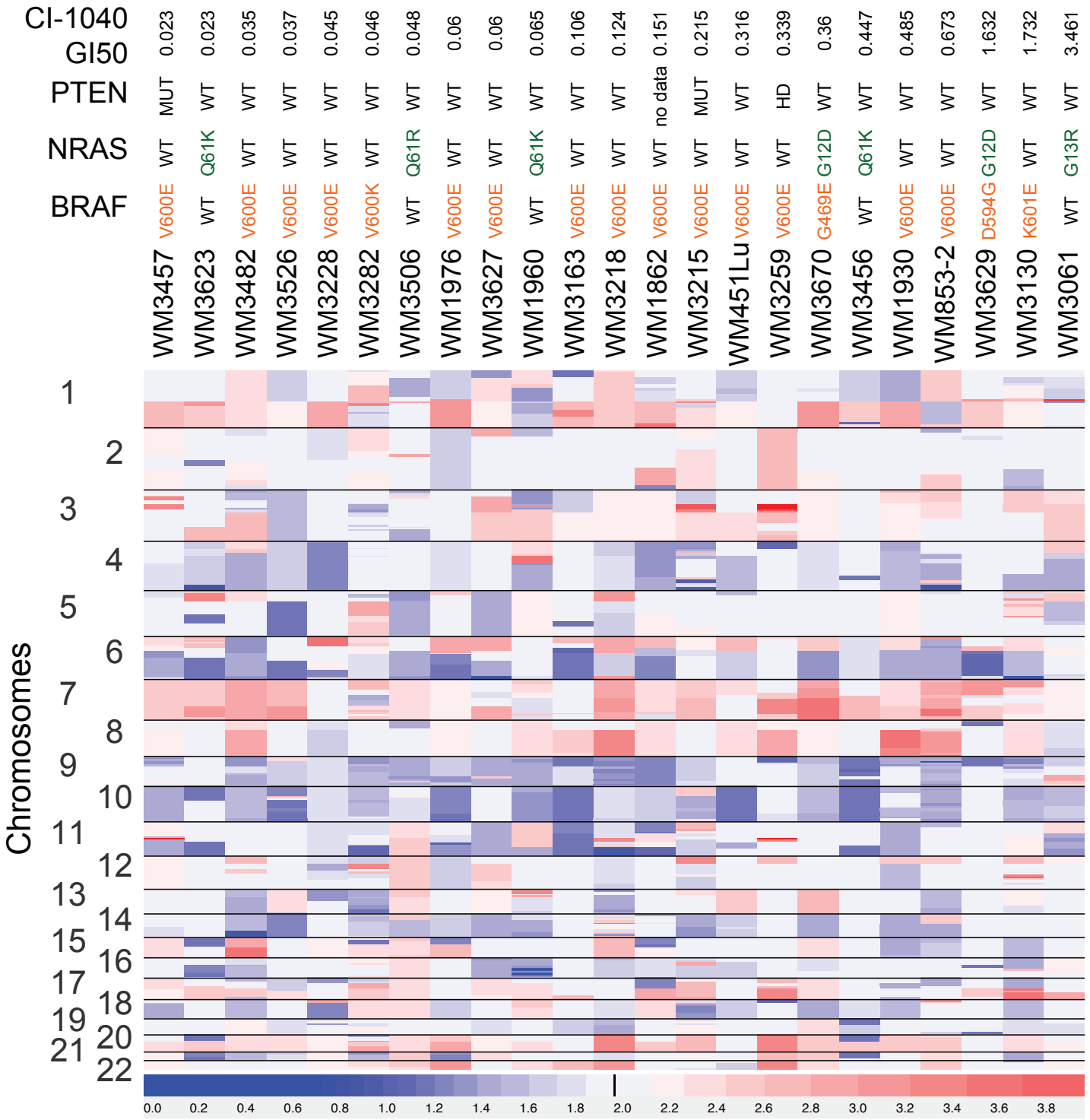
Supplementary Figure 3 - Lin *et al.*



Supplementary Figure 4 - Lin *et al.*



Supplementary Figure 5 - Lin *et al.*



Supplementary Figure 6 - Lin *et al.*