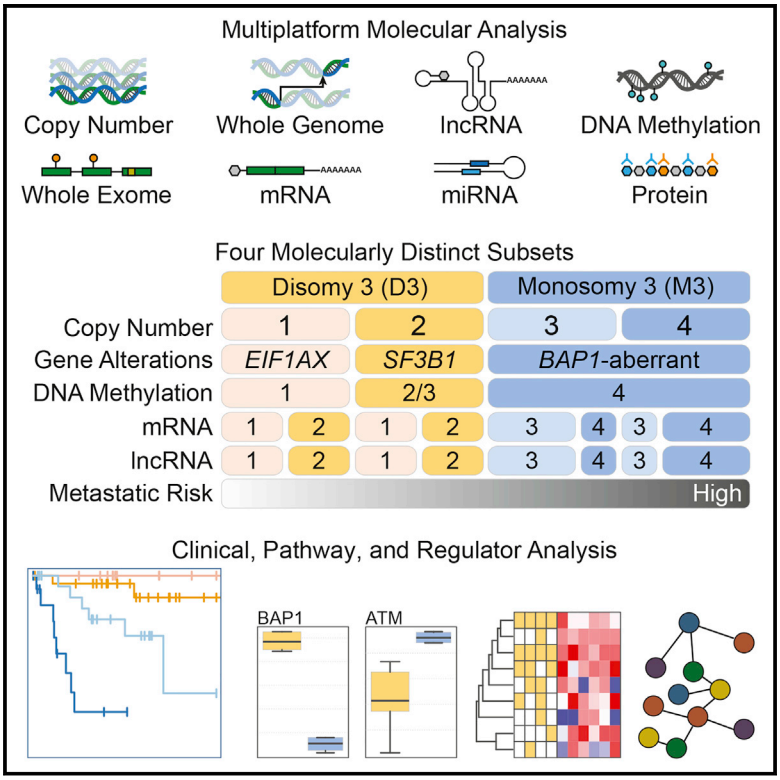
Article

Cancer Cell

Integrative Analysis Identifies Four Molecular and Clinical Subsets in Uveal Melanoma

Graphical Abstract



Highlights

- Both D3 and M3-UM divide into molecularly distinct subsets with different outcomes
- Poor-prognosis M3-UM are characterized by a global DNA methylation pattern
- Poor-prognosis M3-UM subsets have distinct genomic, signaling, and immune profiles
- *EIF1AX* and *SRSF2/SF3B1* mutant D3-UM have different genomic/DNA methylation profiles

Authors

A. Gordon Robertson, Juliann Shih, Christina Yau, ..., Bita Esmaeli, Cyriac Kandoth, Scott E. Woodman

Correspondence

besmaeli@mdanderson.org (B.E.), kandothc@mskcc.org (C.K.), swoodman@mdanderson.org (S.E.W.)

In Brief

Robertson et al. analyze 80 uveal melanomas (UM) and divide poor-prognosis monosomy 3 UM into subsets with divergent genomic aberrations, transcriptional features, and clinical outcomes. Somatic copy number changes and DNA methylation profiles separate better-prognosis disomy 3 UM into low or intermediate risk.

Robertson et al., 2017, Cancer Cell *32*, 204–220
August 14, 2017 © 2017 The Authors. Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.ccell.2017.07.003

CellPress

CellPress

Cancer Cell
Article

Integrative Analysis Identifies Four Molecular and Clinical Subsets in Uveal Melanoma

A. Gordon Robertson,[1,35] Juliann Shih,[2,3,35] Christina Yau,[4,35] Ewan A. Gibb,[1] Junna Oba,[5] Karen L. Mungall,[1] Julian M. Hess,[2] Vladislav Uzunangelov,[6] Vonn Walter,[7,8] Ludmila Danilova,[9] Tara M. Lichtenberg,[10] Melanie Kucherlapati,[11,12] Patrick K. Kimes,[7] Ming Tang,[13] Alexander Penson,[14,15] Ozgun Babur,[16] Rehan Akbani,[17] Christopher A. Bristow,[18] Katherine A. Hoadley,[7,19] Lisa Iype,[20] Matthew T. Chang,[14,21] TCGA Research Network, Andrew D. Cherniack,[2,3] Christopher Benz,[4] Gordon B. Mills,[22] Roel G.W. Verhaak,[13,17] Klaus G. Griewank,[23]

(Author list continued on next page)

[1]Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, BC V5Z 4S6, Canada
[2]The Eli and Edythe L. Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, MA 02142, USA
[3]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA
[4]Buck Institute for Research on Aging, Novato, CA 94945, USA
[5]Department of Melanoma Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
[6]Department of Biomolecular Engineering, Center for Biomolecular Sciences and Engineering, University of California, Santa Cruz, CA 95064, USA
[7]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
[8]Department of Public Health Sciences, Penn State College of Medicine, 500 University Drive, Hershey, PA 17033, USA
[9]The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins University, Baltimore, MD 21287, USA
[10]The Research Institute at Nationwide Children's Hospital, Columbus, OH 43205, USA
[11]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA
[12]Division of Genetics, Brigham and Women's Hospital, Boston, MA 02115, USA
[13]Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
[14]Human Oncology and Pathogenesis Program, Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY 10021, USA
[15]Marie-Josée and Henry R. Kravis Center for Molecular Oncology, Memorial Sloan Kettering Cancer Center, New York, NY 10021, USA
[16]Molecular and Medical Genetics, Computational Biology, Oregon Health and Science University, Portland, OR 97239, USA
[17]Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
[18]Institute for Applied Cancer Science, Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
[19]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
[20]Institute for Systems Biology, Seattle, WA 98109, USA
[21]Departments of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA 94122, USA

(Affiliations continued on next page)

SUMMARY

Comprehensive multiplatform analysis of 80 uveal melanomas (UM) identifies four molecularly distinct, clinically relevant subtypes: two associated with poor-prognosis monosomy 3 (M3) and two with better-prognosis disomy 3 (D3). We show that *BAP1* loss follows M3 occurrence and correlates with a global DNA methylation state that is distinct from D3-UM. Poor-prognosis M3-UM divide into subsets with divergent genomic aberrations, transcriptional features, and clinical outcomes. We report change-of-function *SRSF2* mutations. Within D3-UM, *EIF1AX*- and *SRSF2*/*SF3B1*-mutant tumors have distinct somatic copy number alterations and DNA methylation profiles, providing insight into the biology of these low- versus intermediate-risk clinical mutation subtypes.

Significance

Using sequence assembly approaches, we identified complex alterations in *BAP1* in multiple UM that were not revealed by applying standard SNP/indel algorithms to next-generation sequencing data, suggesting that many *BAP1* alterations are undetected using current techniques. We show that poor-prognosis UM initially develop monosomy 3 (M3), followed by *BAP1* alterations that are associated with a unique global DNA methylation profile. Despite this shared methylation state, poor-prognosis M3-UM separated into two subsets by copy number alterations, RNA (mRNA/lncRNA/miRNA) expression, and cellular pathway activity profiles. Our integrated analysis reveals that the somatic copy number and associated gene expression subtypes correlate with differential clinical outcomes. Our findings reveal four distinct molecular and clinical UM profiles, emphasizing the need for stratified UM patient management.

 © 2017 The Authors. Published by Elsevier Inc.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Ina Felau,[24] Jean C. Zenklusen,[24] Jeffrey E. Gershenwald,[25] Lynn Schoenfield,[26] Alexander J. Lazar,[27] Mohamed H. Abdel-Rahman,[28] Sergio Roman-Roman,[29] Marc-Henri Stern,[29] Colleen M. Cebulla,[30] Michelle D. Williams,[27] Martine J. Jager,[31] Sarah E. Coupland,[32,34] Bita Esmaeli,[33,36,*] Cyriac Kandoth,[15,36,*] and Scott E. Woodman[5,22,36,37,*]
[22]Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
[23]Department of Dermatology, University Hospital Essen, 45157 Essen, Germany
[24]Center for Cancer Genomics, National Cancer Institute, Bethesda, MD 20892, USA
[25]Department of Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
[26]Department of Pathology, The Ohio State University, Wexner Medical Center, Columbus, OH 43210, USA
[27]Department of Pathology, Dermatology and Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
[28]Departments of Ophthalmology and Internal Medicine, Division of Human Genetics, The Ohio State University, Columbus, OH 43210, USA
[29]Department of Translational Research, Institut Curie, PSL Research University, Paris 75248, France
[30]Havener Eye Institute, The Ohio State University Wexner Medical Center, Columbus, OH 43212, USA
[31]Department of Ophthalmology, Leiden University Medical Center, Leiden, the Netherlands
[32]Department of Molecular & Clinical Cancer Medicine, Institute of Translational Medicine, University of Liverpool, Liverpool L7 8TX, UK
[33]Orbital Oncology & Ophthalmic Plastic Surgery, Department of Plastic Surgery, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
[34]Department of Cellular Pathology, Royal Liverpool University Hospital, Liverpool, L69 3GA, UK
[35]These authors contributed equally
[36]Senior authors
[37]Lead Contact
*Correspondence: besmaeli@mdanderson.org (B.E.), kandothc@mskcc.org (C.K.), swoodman@mdanderson.org (S.E.W.)
http://dx.doi.org/10.1016/j.ccell.2017.07.003

INTRODUCTION

Uveal melanoma (UM), which arises from melanocytes resident in the uveal tract, is the second most common melanoma subtype after cutaneous melanoma (CM) (Singh et al., 2011; Virgili et al., 2007). Although both UM and CM tend to occur in people with light iris color and fair skin (Weis et al., 2006), their clinical and molecular characteristics are very different (Coupland et al., 2013; Woodman, 2012). While primary UM is treated with either surgery or radiation and has a low local recurrence rate, up to 50% of UM patients develop distant metastatic disease, often to the liver, after treatment of the primary tumor. At present there are no effective therapies for metastatic UM, and most patients survive less than 12 months after diagnosis of metastases (Blum et al., 2016; Chattopadhyay et al., 2016).

UM displays chromosome aberrations and gene mutations that correlate strongly with clinical outcome and are not present in CM. Loss of one copy of chromosome 3 (monosomy 3, M3) in UM is associated with an increased risk of metastasis and a poor prognosis (Damato et al., 2010; Shields et al., 2010). Loss-of-function mutations in *BAP1*, which is located on 3p21, have been identified in M3-UM (Harbour et al., 2010), and decreased BAP1 mRNA and protein expression, indicating *BAP1* aberrancy, are highly correlated with the development of UM metastases (Kalirai et al., 2014; Koopmans et al., 2014). Currently either disomy 3 (D3) versus M3 status or a 12-gene microarray-based gene expression panel is used to determine whether a patient is in a low- or a high-risk prognostic group (Harbour, 2014; Tschentscher et al., 2003). Recent analysis of a large D3-UM cohort showed *SF3B1* mutation to be associated with an intermediate risk of developing later-onset metastatic UM (Yavuzyigitoglu et al., 2016).

Despite prognosis being clearly correlated with the expression of a small panel of marker genes, with M3, and with *BAP1* aberrancy or *SF3B1* mutation, the molecular pathways involved in the development of metastatic disease have not been elucidated. In this Rare Tumor Project of The Cancer Genome Atlas (TCGA), we performed a global and integrated molecular characterization of 80 primary UM, seeking to generate insights into biological processes that underlie UM tumors that have distinctly different prognoses.

RESULTS

Sample and Data Collection

Eighty primary UM tumors were available for multiplatform analysis (Table S1). Cancer cell contents were high based on ABSOLUTE (median purity = 0.95, Figure S1A), DNA methylation-derived leukocyte fraction, and histopathological assessment. All cases were $\geq$T2 (seventh edition of the AJCC TNM-staging system). As in Diener-West et al. (2005), ~10% of patients developed another primary malignancy.

Chromosome Copy Number Aberrations

In primary UM, recurrent chromosome aberrations include losses in 1p, 6q, 8p, and 16q; gains in 6p and 8q; and M3 (Coupland et al., 2013). We used the ABSOLUTE and FACETS algorithms to estimate clonal and subclonal somatic copy number alterations (SCNA) from SNP microarray and whole-exome sequencing (WES) data. Unsupervised SCNA clustering defined four subtypes that had diverse aneuploid events and divided D3-UM and M3-UM into two subgroups each (Figure 1A). In D3-UM, cluster 1 showed the least aneuploidy and was enriched for partial or total 6p gain, with no other significant chromosome aberrations; cluster 2 showed 6p gain and partial 8q arm gains. In M3-UM, both clusters 3 and 4 showed 8q whole-arm gain in nearly all samples, with median 8q copy numbers 3 versus 5 (i.e., 1 versus 3 extra copies) respectively. Evidence for 8q isochromosome (i.e., chromosome 8 with two q arms) was seen in all 20 samples in cluster 4, but in only 4 of 22 samples in cluster 3 (Table S1). Thus, while both M3 and 8q gain co-occurred in clusters 3 and 4, the 8q copy number burden and

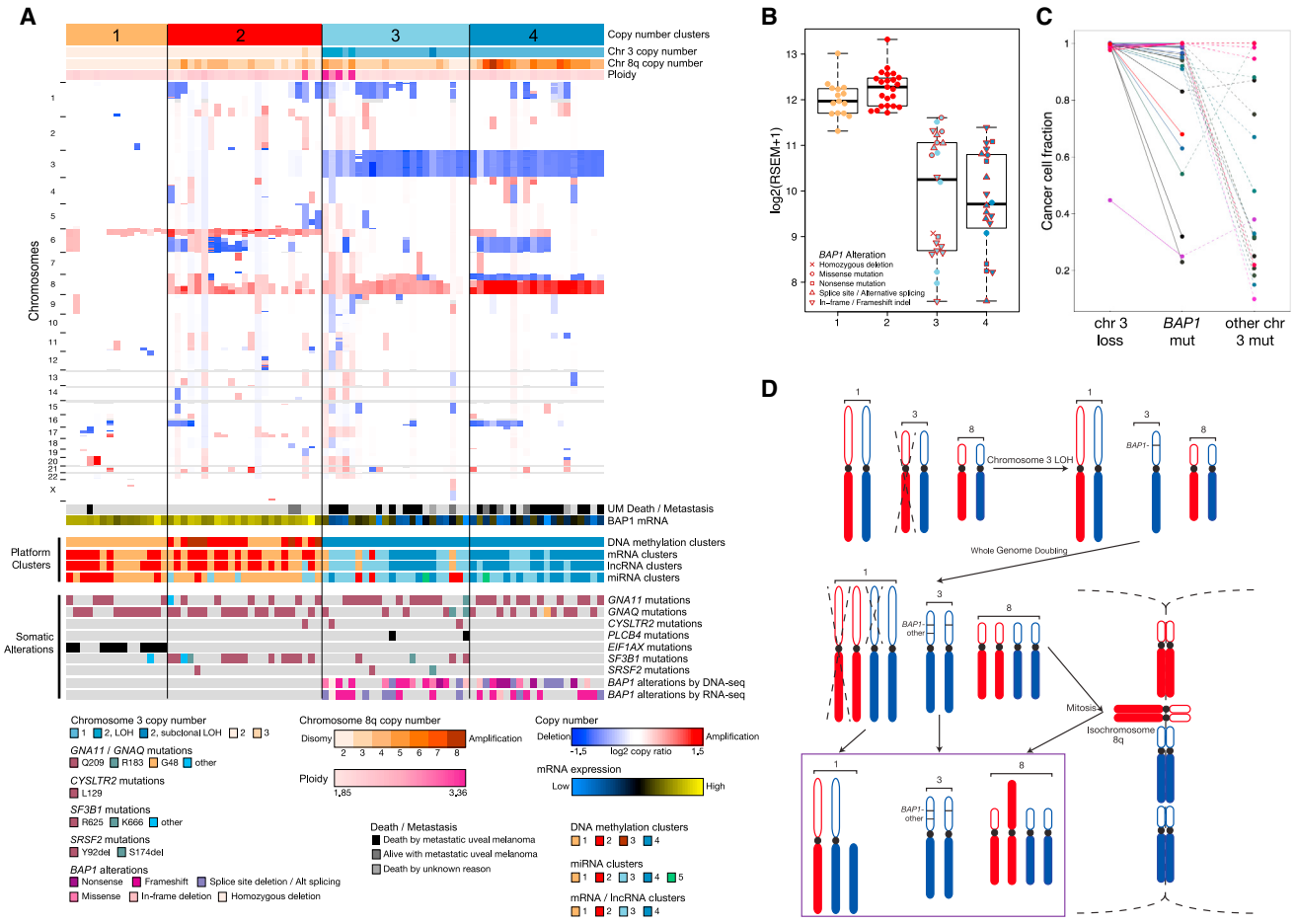


Figure 1. Genomic Landscape of Primary UM

(A) Unsupervised clustering of somatic copy number alterations (SCNAs) separated 80 primary UM into four clusters: 1 (n = 15), 2 (n = 23), 3 (n = 22), and 4 (n = 20), ordered by increasing chromosomal instability. The upper covariate tracks show SCNA clusters (1–4), chromosome 3 and 8q copy number, and ploidy level. The heatmap shows somatic copy number ratio (diploid = 0, white). Lower covariate tracks show (i) clinical outcome; (ii) BAP1 mRNA expression; (iii) unsupervised clusters for DNA methylation, mRNA, lncRNA, and miRNA; (iv) mutations in G-protein-signaling genes, splicing factors, and *EIF1AX*; (v) *BAP1* alterations that include alternate splicing and rearrangements detected by assembly of DNA-seq and RNA-seq data.

(B) BAP1 mRNA expression, grouped by SCNA clusters, with *BAP1* alteration status determined by at least one method in (A). Dots show all data values. Box plots show median values, and the 25th to 75th percentile range in the data, i.e., the interquartile range (IQR). Whiskers extend 1.5 times the IQR.

(C) Cancer cell fractions for chromosome 3 loss, *BAP1* alterations, and other somatic mutations on chromosome 3, for tumors with *BAP1* alterations detected either by standard SNP-indel algorithms or by local reassembly of WES data. Lines connect events that occurred in the same tumor.

(D) Schematic depicting a probable sequence of somatic events resulting in those detected in the cluster 3 case V4-A9EO (M3, *BAP1* mutation, WGD, then isochromosome 8q).

See also Figure S1 and Table S1.

type varied between the two clusters. Finally, one tumor in cluster 2 and four in cluster 3 showed higher ploidy values, and were predicted to have undergone whole-genome doubling (WGD).

Gene Mutations Identified by Standard Algorithms

In WES data for matched tumor-blood pairs, the median somatic mutation density of 1.1 per Mb was markedly lower than in CM (Cancer Genome Atlas Research Network, 2015), other melanoma subtypes, or other common solid tumors (Tetzlaff et al., 2015). As in (Johansson et al., 2016), we observed no evidence of the UV radiation mutational signature seen in ~80% of CM (Cancer Genome Atlas Research Network, 2015); rather, there were varying proportions of three non-UV-associated signatures (Figure S1B).

Nine significantly mutated genes (SMGs) were detected using MutSig2CV or CoMet: *GNAQ*, *GNA11*, *SF3B1*, *EIF1AX*, *BAP1*, *CYSLTR2*, *SRFF2*, *MAPKAPK5*, and *PLCB4* (Figures S1B and S1C). None of these have been identified as SMGs in CM (Johnson et al., 2016). We found mutually exclusive somatic mutations in the G-protein pathway-associated *GNAQ* and/or *GNA11* (92.5%), *CYSLTR2* (4%), and *PLCB4* (2.5%) genes, consistent with previous findings (Johansson et al., 2016; Moore et al., 2016; Van Raamsdonk et al., 2009, 2010) (Figure S1C and Table S1).

EIF1AX and *SF3B1* mutations in 27 of the 80 UM (34%) were nearly mutually exclusive, consistent with Martin et al. (2013). Nine of ten *EIF1AX*-mutant cases had their mutations in the protein N-terminal region (G6–G15), as in papillary thyroid

carcinomas (Cancer Genome Atlas Research Network, 2014c) (Table S1). *EIF1AX* mutations were present only in UM with neither M3 nor 8q gain, and were exclusively in SCNA cluster 1 (Figure 1A). *SF3B1* mutations resulted in R625C/H amino acid alterations in 14 of 18 samples, while in four UM, mutations resulted in H662R (n = 2), K666T, or T663P, which are frequently-altered sites in other malignancies (Alsafadi et al., 2016). Only one UM harbored both an *EIF1AX* and a *SF3B1* mutation; the latter was an atypical T663P. As was the case for *EIF1AX* mutations, the majority (78%) of UM with *SF3B1* mutations were present in D3-UM, consistent with Johnson et al. (2017). However, unlike *EIF1AX* mutations, *SF3B1* mutations in D3-UM were associated with SCNA cluster 2, most with partial 8q gains. Thus, *EIF1AX*- and *SF3B1*-mutant D3-UM were each associated with nearly mutually exclusive SCNA profiles.

We identified *SRSF2* as an SMG that harbored in-frame Y92 deletions (Y92del) in two UM and an S174del in a third. Tumors with *SRSF2* mutations had neither *SF3B1* nor *EIF1AX* mutations, and were found in both D3-UM and M3-UM with 8q gains, suggesting functional similarities between *SRSF2*- and *SF3B1*-mutant UM.

Mutant Gene-Specific Splicing Events

Missense mutations at K666 and R625 in splicing factor *SF3B1* are associated with alternative branchpoint usage (Alsafadi et al., 2016), and missense mutations at P95 in splicing factor *SRSF2* are associated with exon exclusion in myelodysplastic syndrome/acute myeloid leukemia (Kim et al., 2015; Zhang et al., 2015). Using rMATS to compare RNA sequencing (RNA-seq) data for UM with mutations in either gene versus UM with wild-type *SF3B1* and *SRSF2* suggested that such mutations may alter translation initiation in a large subset of UM. For example, when *SF3B1* has a K666/R625 mutation, the initiation factor *EIF4A2* used a neo-acceptor that resulted in a frameshift in the open reading frame (Figure S1D), and when *SRSF2* had a Y92del, *EIF4A2* had a skipped exon. In *SRSF2* Y92del UM, Src kinase *FYN* had a skipped exon and a larger ratio of FYN-T versus FYN-B isoforms (Figures S1E and S1F). Finally, an exon in the C-terminal domain of *EIF2S3* had among the largest fold changes in expression in all *SF3B1*-mutant UM, but was absent in all UM with wild-type *SF3B1*/*SRSF2*.

BAP1 Alterations Identified by DNA-Seq and RNA-Seq Assembly

Both germline and somatic *BAP1* alterations have been described in UM (Abdel-Rahman et al., 2011; Harbour et al., 2010). While Sanger sequencing initially identified truncating and non-truncating *BAP1* mutations in 81.5% of M3-UM (Harbour et al., 2010), in our cohort standard SNP/indel analysis of WES data identified only 40.5% (17/42) of M3-UM as having *BAP1* mutations. To recover alterations that were inaccessible to our SNP/indel-calling methods, we applied MuTect2 local reassembly to exome capture DNA sequencing (DNA-seq) data, and Trans-ABySS global *de novo* assembly to RNA-seq data. Combining results from both methods and data types identified an additional 18 UM with *BAP1* alterations, often long or complex, raising the percentage of samples with *BAP1* alterations to 83.3% (Figure S1G). The additional *BAP1* genetic alterations were present only in M3-UM that displayed low levels of BAP1 mRNA expression, consistent with *BAP1* loss of heterozygosity.

BAP1 mRNA expression was significantly ($p = 5.3 \times 10^{-16}$) higher in SCNA clusters 1 and 2 (D3) than in SCNA clusters 3 and 4 (M3). However, we found no significant difference in BAP1 mRNA expression in M3-UM with versus without *BAP1* aberrancy, indicating that our approach may not have detected some *BAP1* alterations, or that BAP1 regulation may involve additional epigenetic mechanisms (Figure 1B).

We used ABSOLUTE to determine the relative timing of chromosome 3 loss and of *BAP1* alterations (Figure 1C). Most *BAP1* alterations were predicted to be either subclonal or clonally homozygous. Three of the four UM with WGD in SCNA cluster 4 had homozygous *BAP1* alterations with multiplicity 2, indicating that both M3 and *BAP1* alterations occurred before WGD. Additionally, with one exception in which M3 was clearly subclonal, the cancer cell fractions of M3 were close to 1 (mean = 0.97), suggesting that M3 was an early event that propagated through nearly all clones within each tumor. Cancer cell fractions of *BAP1* alterations were lower (mean = 0.88) and fractions of other putative passenger mutations on chromosome 3 were even lower (mean = 0.60). From these results, we infer that M3 occurs prior to *BAP1* alterations, and that both events occur prior to other mutations on the remaining chromosome 3, followed by WGD in some cases (Figure 1D).

BAP1-Aberrant UM Correlates with a Global DNA Methylation Profile

Unsupervised consensus clustering on the most variable 1% of CpG probes yielded a four-cluster solution (Figure 2). *EIF1AX*-mutant tumors were only present in DNA methylation cluster 1, while UM in DNA methylation clusters 2 and 3 were highly enriched (12 of 16 tumors) with *SF3B1*/*SRFR2* mutations. Thus, D3-UM with *EIF1AX* versus *SF3B1*/*SRFR2* mutations possessed distinct DNA methylation patterns. M3/*BAP1*-aberrant UM tumors showed a single global DNA methylation profile.

Four Transcription-Based UM Subsets

We used RNA-seq data to profile the expression of 20,531 mRNAs and of 8,167 long non-coding RNAs (lncRNAs) and processed transcripts, and identified four-cluster consensus solutions for both mRNA and lncRNA (Figure 3). D3-UM divided into transcription-based clusters 1 and 2, M3-UM into clusters 3 and 4, and the 12-gene panel's two prognostic groups were each further separated into two groups. Specific mRNAs and lncRNAs were differentially and highly expressed in each subset (Figure S2). We noted that lncRNAs LINC00152 (CYTOR) and BANCR, well-established cancer-associated lncRNAs, had higher abundance in poor-prognosis clusters 3 and 4 compared with good-prognosis clusters 1 and 2 (Figure S2A). Other functionally characterized lncRNAs such as NEAT1 and MALAT1 were differentially expressed between poor-prognosis clusters 3 and 4. We identified mRNAs and lncRNAs whose expression was associated with recurrent SCNAs and/or DNA methylation (Table S2 and Figures S2B–S2E). For example, the expression of PVT1 (8q24.21) was highly correlated with SCNA 8q (rho = 0.65, false discovery rate [FDR] = 6×10^{-8}) and this lncRNA was among the most differentially expressed transcripts in poor-prognosis lncRNA clusters 3 and 4 versus clusters 1 and 2. Both LINC00152 and PVT1 were among a small set of differentially expressed M3-associated lncRNAs that were

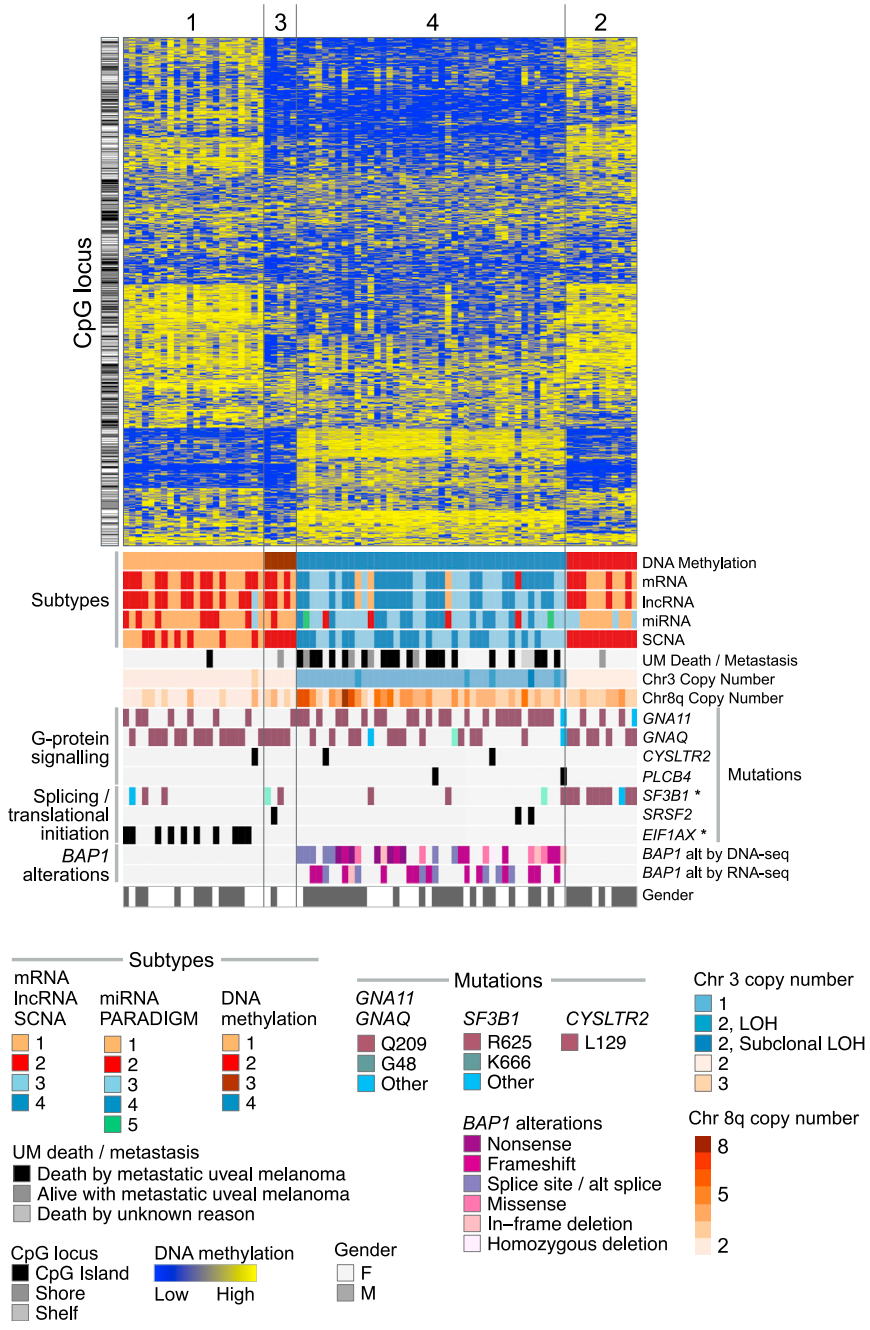


Figure 2. DNA Methylation Landscape in Primary UM

Unsupervised clustering of DNA methylation data, with the heatmap showing beta values ordered by DNA methylation clusters. CpG locus types (island, shore, and shelf) are indicated at the left border. Covariate tracks show unsupervised clusters for four other genomic data types, clinical outcomes, chromosome 3 and 8q copy number status, specific gene alterations, and gender. *SF3B1* and *EIF1AX* mutations were statistically associated with the clusters (*$p < 0.01$, Fisher's exact test). LOH, loss of heterozygosity.

significantly influenced by DNA methylation (Table S2 and Figure S2E). Increased LINC00152 expression has been reported in solid tumors and is correlated with cell migration, invasion, and proliferation (Pang et al., 2014). PVT1 has been shown to be oncogenic through multiple mechanisms, including stabilization of MYC protein levels (Colombo et al., 2015).

CYSLTR2, which is recurrently mutated in ~3% of primary UM, showed markedly low expression in mRNA cluster 1 versus all other clusters (Figure S2F), suggesting possible roles for both *CYSLTR2* expression and mutation. Transcripts with the highest fold changes in mRNA cluster 4 included immune genes and genes localized to 8q (Figure S2F). LncRNAs and mRNAs that were differentially abundant between SCNA- and transcription-based subtypes are shown in Figures S2A, S2F–S2H.

The miRNA Expression Landscape Is Concordant with Transcriptional UM Subsets

MicroRNA sequencing (miRNA-seq) data identified four consensus clusters, with a two-sample outlier group in which cancer-associated miRNAs were differentially abundant (e.g., miR-9, -21, -182/3, -375; Figure S3A). The four main miRNA clusters were clearly associated with M3 and its DNA methylation state, and were less concordant with the mRNA and lncRNA subtypes than these were with each other (Figures S3B and S3C). Consistent with Worley et al. (2008), miR-199a-3p/5p, miR-199b-3p, and let-7b-5p were more highly expressed in the M3-enriched miRNA cluster 3 (Figure S3D). In addition, miR-486-5p and miR-451a were abundant in miRNA cluster 3, while cluster-4 tumors showed higher expression of miR-142, -150, -21, -29b, -146b, and -155. While miRNAs localized to Xq27.3 were abundant in subtype 1, the association between gender and subtypes was not significant ($p = 0.77$, Fisher's exact test).

Many cancer-associated miRNAs (Schoenfeld, 2014) were differentially expressed between clusters. For example, expression of the oncomiR miR-21-5p was ~4-fold greater in miRNA cluster 4 (Figure S3D), consistent with *MIR21* DNA hypomethylation (Figure S3E). Expression of 39 other miRNAs was influenced by DNA methylation (Table S2). Expression of certain miRNAs was influenced by SCNA; miR-30d and miR-151a expression was correlated with 8q SCNA (Figures S3E–S3G), and M3-UM had lower expression of a number of chromosome 3 miRNAs, including let-7g, miR-28, and miR-191. Differential miRNA-mRNA targeting relationships were inferred between miRNA clusters 3 and 4 (Figures S3H–S3I).

miRNA cluster 4 corresponded to M3-UM with immune infiltration (Figure S3A), suggesting that expression of a number of

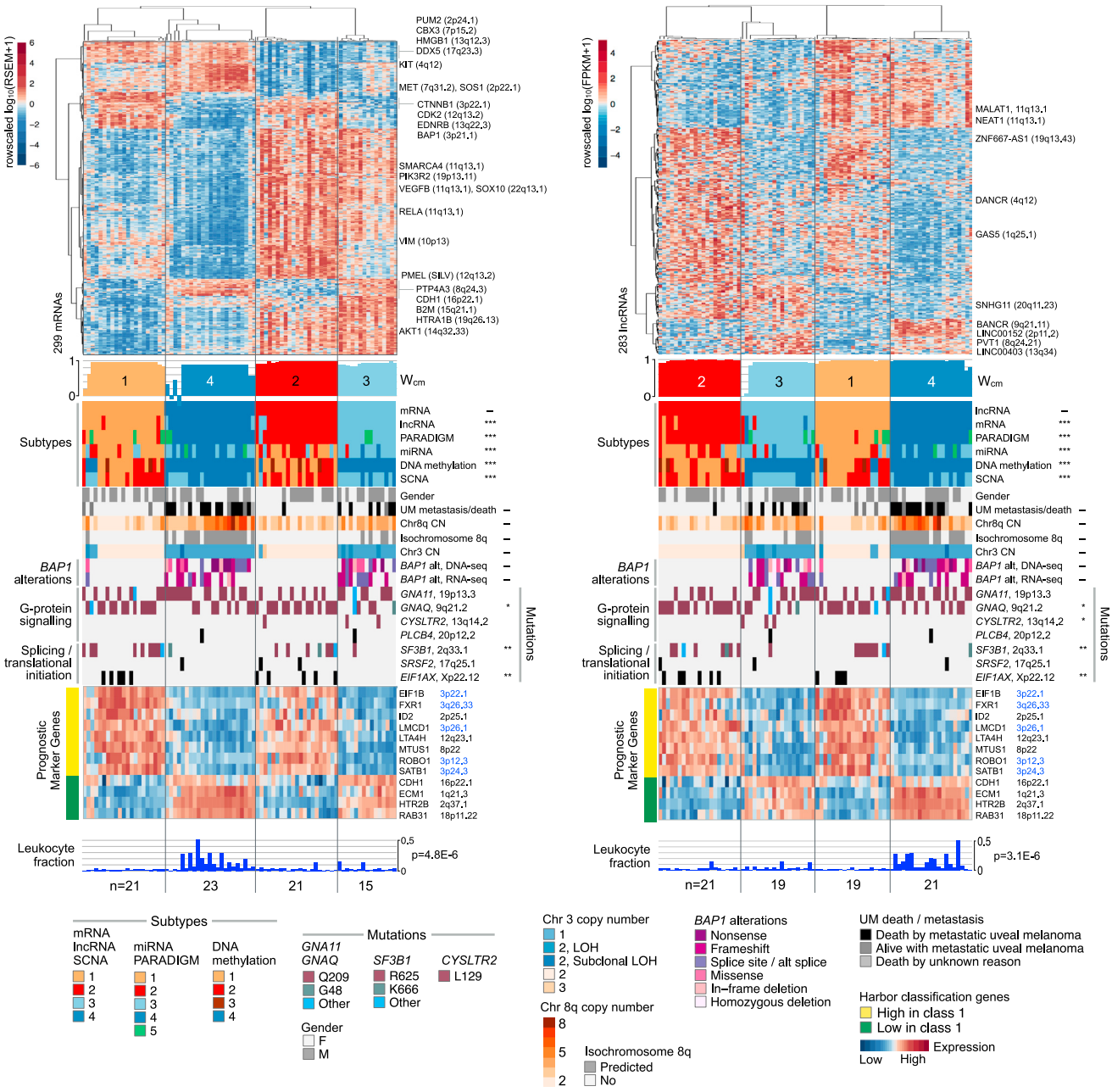


Figure 3. Gene Expression Patterns in UM

The upper heatmap shows unsupervised consensus clustering for RNA-seq data of mRNA (left) or lncRNA (right) expression. Covariate annotation tracks show selected genomic and clinical features. The lower heatmap displays the expression profiles of 12 genes used in a prognostic test for the risk of developing metastasis (Harbour, 2014), with blue text highlighting genes that are on chromosome 3. High-risk primary tumors show low expression of eight of these genes and high expression of four genes (yellow versus green panels at the left). *BAP1* structural alterations that include alternative splicing and rearrangements were detected by assembly of RNA-seq and DNA-seq data. Leukocyte fraction was estimated from DNA methylation data. LOH, loss of heterozygosity. *, **, *** p value < 0.1, 0.01, 10^{-10}, Fisher's Exact or Chi-square test.

See also Figures S2 and S3; Table S2.

miRNAs may be associated with the promotion of an immune environment that plays a significant role in aggressive UM.

Characteristics of Immune-Infiltrated UM

By both DNA methylation and RNA-seq analyses we inferred that a CD8 T cell infiltrate was present in ~30% of M3-UM while nearly absent in D3-UM, and found that genes involved in interferon-γ signaling (*IFNG*, *IFNGR1*, and *IRF1*), T cell invasion (*CXCL9* and *CXCL13*), cytotoxicity (*PRF1* and *GZMA*), and immunosuppression (*IDO1*, *TIGIT*, *IL6*, *IL10*, and *FOXP3*) were coexpressed with *CD8A* (Figure 4A).

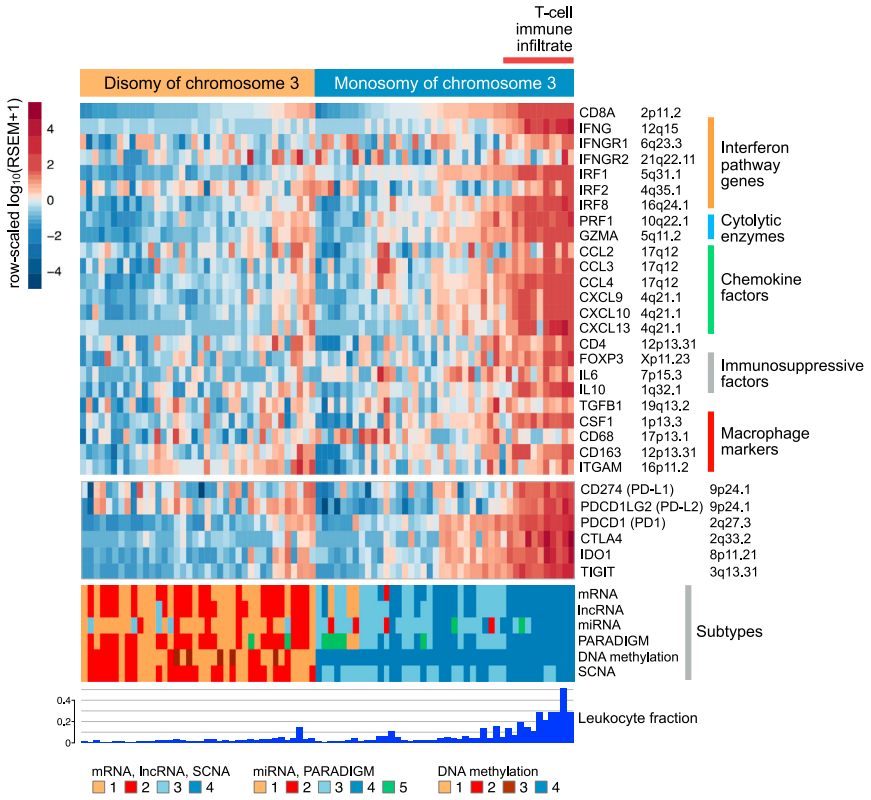


Figure 4. Immune Gene Expression in M3- versus D3-UM

Heatmap for 80 primary UM, highlighting mRNA expression levels of key immunological genes that represent the interferon-γ pathway, T cell cytolytic enzymes, chemokine factors, immunosuppressive factors, and macrophage markers, as well as individual immune checkpoint blockade genes (*CD274*, *PDCD1LG2*, *PDCD1*, *CTLA4*, *IDO1*, and *TIGIT*). Samples were separated by D3 versus M3 status, and sorted from lowest (left) to highest (right) CD8A expression level. Covariate tracks show mRNA, lncRNA, miRNA, PARADIGM, DNA methylation, and SCNA clusters. Leukocyte fraction was estimated from DNA methylation data. See also Figure S4.

Consistent with human leukocyte antigen (HLA) gene expression correlating with the presence of an inflammatory infiltrate (Maat et al., 2008), we found that HLA expression was higher in M3-UM and correlated with CD8A expression (Figure S4A). Furthermore, in 50 UM with low-pass whole-genome sequencing data we identified 11 structural variations in HLA genes (Figure S4B) in which differential HLA expression was observed in D3-UM versus M3-UM ($p = 0.015$, Fisher's exact test).

Pathways and Regulators Are Differentially Active between UM Subsets

We analyzed RNA (PARADIGM and MARINa algorithms) and protein (reverse-phase protein array [RPPA]) expression to identify activated signaling pathways and regulators in the UM subsets. PARADIGM-inferred pathway levels resolved four major groups of samples, with a smaller ($n = 7$) more heterogeneous group (Figure 5A). In PARADIGM cluster-4 cases, 95% of which were also transcription-based cluster 4, DNA damage repair/response (DDR) was active, as was MYC signaling and HIF1a, consistent with an upregulated hypoxia response. Multiple immune-related transcription factors were relatively active, including JAK2-STAT1/3 and JUN-FOS, consistent with the elevated levels of immune-related genes in these poor-prognosis M3 tumors. PARADIGM cluster-3 cases, 93% of which were transcription-based cluster 3, showed higher activities of key transcription factors FOXA1 and FOXM1, as well as elevated levels of MAPK and AKT, indicating high cellular cycling and cell proliferation. Thus, although the two subsets of poor-prognosis M3/*BAP1*-aberrant UM shared the same global DNA methylation profile, they had markedly distinct cellular signaling profiles.

Noting that SCNA-based and transcription-based clusters were largely but incompletely concordant (Figures 1, 3, and 5), we compared differential PARADIGM signaling and MARINa regulator activities between clusters (Figures S5A–S5C). For both transcription- and SCNA-based clusters, DDR, HIF1a, and MYC signaling were more active in cluster 4 than in cluster 3. However, the mediators of immune signaling observed in transcription cluster 4 were not identified for SCNA clusters (Figures S2F–S2G and S5D), suggesting a biological basis for the incomplete concordance between transcription- and SCNA-based clustering.

Given the strong correlation between M3 and 8q gain (Figure 1A), the oncogenic transcription factor MYC (8q24.21) has been postulated to play a role in UM progression (McCarthy et al., 2016; van den Bosch et al., 2012). MYC can either activate or repress its gene targets, depending on its complexes (e.g., with MAX and/or MIZ1) (Kress et al., 2015). PARADIGM showed highly differential activation of MYC/MAX targets across the cohort (Figure 5A). Unexpectedly, both PARADIGM clusters 1 (mostly D3/8q-normal tumors) and 4 (all poor-prognosis M3/8q-gain tumors) displayed high MYC/MAX complex activity levels, despite differing most in 8q levels. In contrast, MYC/MAX/MIZ complex targets were most represented in PARADIGM clusters 4 and 5 (88% M3/8q-gain tumors). Thus, activities for MYC/MAX/MIZ, but not MYC/MAX, corresponded with M3/8q-gain status.

Sufficient tissue material was available from 11 UM samples, five M3/*BAP1*-aberrant versus six D3/*SF3B1*-mutant, to generate RPPA data. As expected, BAP1 protein levels were lower in M3/*BAP1*-aberrant cases. M3/*BAP1*-aberrant UM had a higher ($p = 0.017$) DDR pathway score than D3/*SF3B1* R625-mutant UM (Figure 5B and Table S3). This is consistent with PARADIGM pathway results; with *in vitro* data indicating a role for BAP1 in homologous recombination DDR (Eletr et al., 2013; Yu et al., 2014); and with each of the M3/*BAP1*-aberrant UM evaluated in the RPPA analysis having evidence of isochromosome 8q gain, which can be mediated through inefficient repair of homologous recombination.

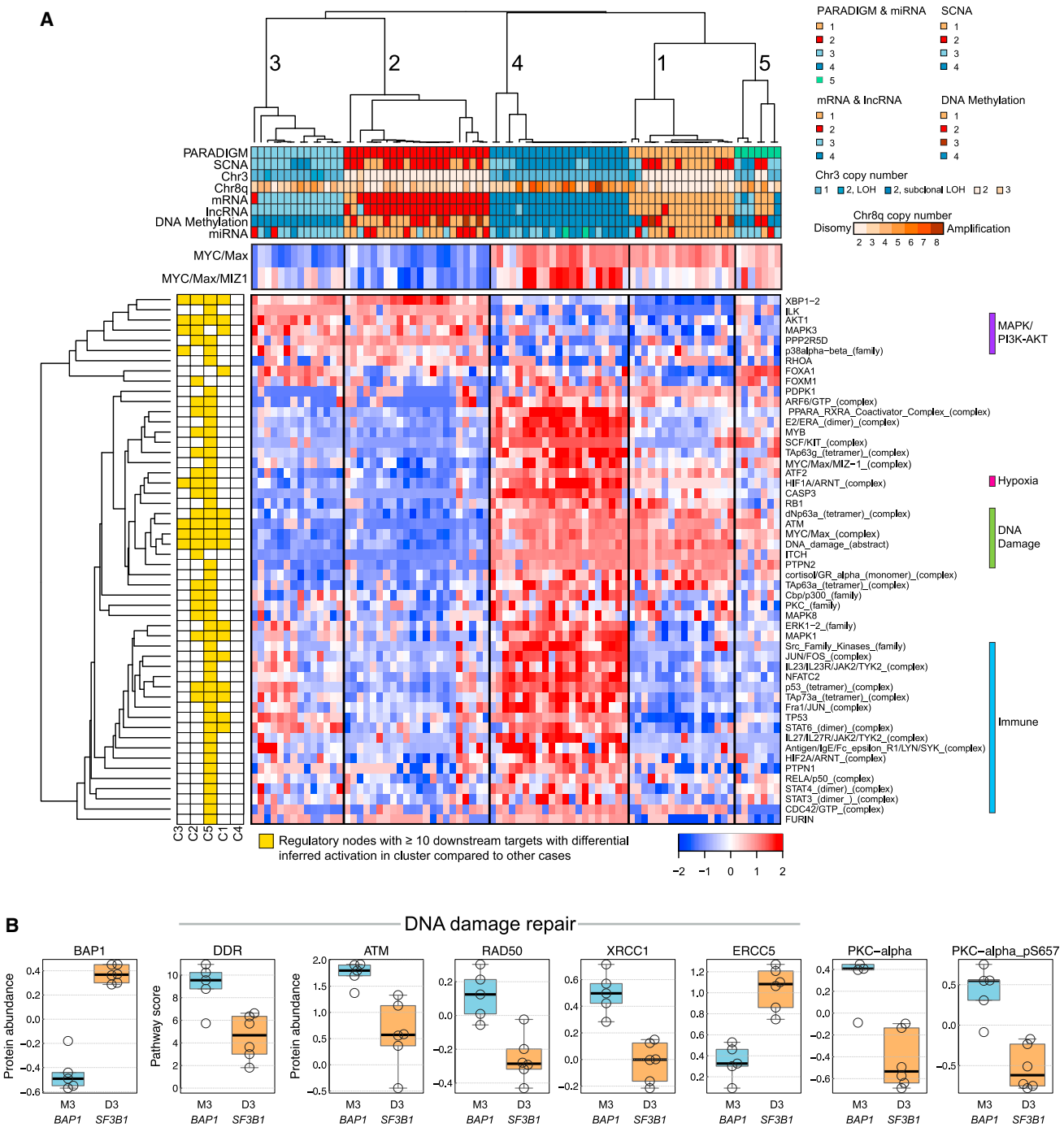
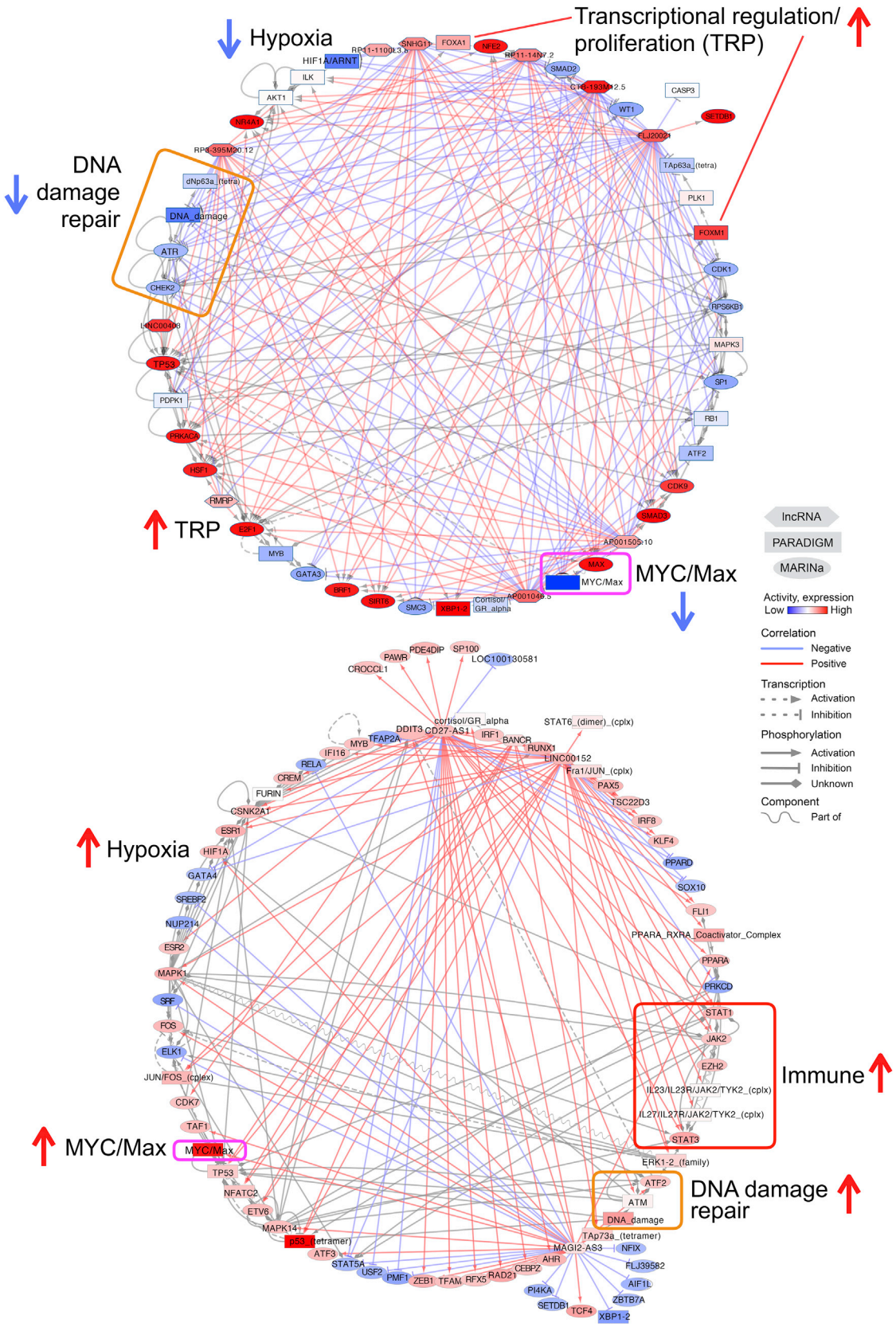


Figure 5. Integrative Pathway Analysis of UM

(A) Heatmap of hierarchically clustered PARADIGM inferred pathway levels (IPLs) for 80 primary UMs. Samples are clustered into five groups (top horizontal track). Below this are cluster memberships for other platforms, and for chromosome 3 and 8q copy number, then IPL profiles for the MYC/MAX and MYC/MAX/MIZ1 complexes. The main heatmap shows PARADIGM features or nodes that have at least ten downstream regulatory targets and are differentially active in one-cluster-versus-other comparisons; the annotation panel to the left indicates the cluster(s) in which a node satisfies these conditions. The vertical colored bars on the right highlight sets of pathway nodes that belong to common biological processes: MAPK/PI3K-AKT (purple), hypoxia (magenta), DNA damage repair/response (green), and immune response (blue). LOH, loss of heterozygosity.

(B) Distributions of DDR pathway score and abundance for selected proteins, from RPPA data for M3/*BAP1*-aberrant versus D3/*SF3B1*-mutant UM (n = 11). PKC-α_pS657 denotes PKC-α phosphorylated at S657. Box plots show median values and the 25th to 75th percentile range in the data, i.e., the IQR. Whiskers extend 1.5 times the IQR. Circles show all data values.

See also Figure S5 and Table S3.



(legend on next page)

All of the samples tested by RPPA harbored an activating *GNAQ/11* mutation, and protein kinase C (PKC) isoforms are downstream effectors of activated mutant *GNAQ/11* (Wu et al., 2012). Protein levels for both total PKC-α, activated phospho-PKC-α (S657), and phospho-PKC-δ (S664) were markedly higher in M3/*BAP1*-aberrant UM compared with D3/*SF3B1* R625 UM, indicating that activated mutant *GNAQ/11* signaling may be enhanced in M3/*BAP1*-aberrant UM.

Because the roles of lncRNAs (Hon et al., 2017; Nguyen and Carninci, 2016) in UM largely remain to be clarified, we compared correlations of lncRNA abundance with PARADIGM pathway activities and MARINa regulator activities in the M3/*BAP1*-abberant lncRNA transcriptional clusters 3 and 4 (Figure 6 and Table S4). In cluster 3, LINC00403, RMRP, and SNHG11, and uncharacterized lncRNAs such as RP11-14N7.2 and CTB-193M12.5, were correlated with activated transcriptional regulators of proliferation (e.g., FOXM1, FOXA1, E2F1), low MYC/MAX complex pathway activation, diminished HIF1A/ARNT complex activity, and low DDR pathway activity. In cluster 4, LINC00152, BANCR, MAGI2-AS3, and CD27-AS1 were positively correlated with immune-associated pathway nodes and regulators of JAK-STAT and cytokine mediators, as well as mediators of DDR, MYC/MAX, and HIF1α activity.

Correlation of Distinct Biological Subsets with Clinical Prognosis in UM

As expected, M3-UM patients had a significantly worse prognosis than D3-UM (Figures 7A and S6A). While limited by the duration of follow-up, we observed that features known to be prognostic (i.e., histological type, closed connective tissue loops, and tumor-associated macrophage infiltration) were also prognostic in our cohort (Figure S6B).

As all M3-UM shared the same global DNA methylation profile (Figure 2), M3 and DNA methylation cluster 4 had identical Kaplan-Meier curves (Figure 7A). SCNA clusters 3 and 4, wholly comprising M3-UM cases, had different UM metastasis (i.e., the time interval from primary UM diagnosis to development of distant UM metastasis) ($p = 0.002$). Consistent with mRNA and lncRNA clusters 3 and 4 largely overlapping SCNA clusters 3 and 4 (Figures 1 and 3), differences in UM metastasis for transcriptional clusters trended similarly.

We then sought to identify genes whose expression was associated with differential time to UM metastasis (Figure S7). We identified 111 mRNAs and 23 lncRNAs in our TCGA cohort that were both differentially abundant in M3 SCNA clusters 3 versus 4 ($|\text{fold change}| > 2$ and 1.5, respectively; FDR < 0.05), and associated with UM metastasis in M3 cases (95% confidence interval [CI] on the hazard ratio [HR] either less than or greater than 1.0) (Figures S2H, and S7; Tables S5 and S6). For mRNAs and lncRNAs in the TCGA that were more abundant in SCNA cluster 4, most HR were above 1.0 (Figures S7A–S7C). Thirty-five of the differentially abundant mRNAs and three lncRNAs were also associated with UM metastasis in an independent cohort (Laurent et al., 2011) (Figures S7C–S7E, Table S6). Eighteen (69%) of the 26 genes with HR 95% CI > 1.0 in both cohorts (i.e., with higher gene expression associated with shorter UM metastasis) were on 8q (Figure S7C). Despite localizing to 8q, the expression of *ENPP2* (8q24.12) was associated with a low HR in both cohorts (0.30 and 0.36, respectively), consistent with our unbiased analysis that showed *ENPP2* DNA methylation to be anti-correlated with its transcript expression (Spearman $\rho = -0.81$) (Table S2). Four of the 12 genes with HR 95% CI < 1.0 were associated with recurrent SCNA losses in 3p (*PPARG*, *SYN2*), 6q (*NEDD9*), and 8p (*SLC7A2*).

DISCUSSION

Our integrated, multidimensional molecular and computational investigation into UM provides insights that have mechanistic, prognostic, and therapeutic implications. The analysis divided primary UM tumors into four molecular groups, subdividing poor-prognosis M3-UM and better-prognosis D3-UM into two subgroups each (Figure 7B). We show that poor-prognosis M3-UM is associated with a distinct global DNA methylation pattern that differs from the pattern observed in D3-UM, suggesting that *BAP1* aberrancy may result in metastasis-prone DNA methylation state. M3-UM cases, despite sharing a characteristic global DNA methylation profile, were divided by SCNA-based and transcription-based analyses into two subgroups that have different biological pathway profiles and clinical outcomes.

Prior studies have shown poorer clinical outcomes to be associated with higher chromosome 8q copy number (Caines et al., 2015; Cassoux et al., 2014; Versluis et al., 2015). Given the proposed role of BAP1 in DNA damage repair/response (DDR) (Ismail et al., 2014; Yu et al., 2014), and the upregulated DDR pathway activity by both transcription- and protein-based pathway analyses, these data suggest that loss of BAP1 function may result in inefficient DDR, and may play a role in isochromosome 8q formation observed in all SCNA cluster 4 and one-fourth of SCNA cluster 3 M3-UM samples; however, studies to confirm this hypothesis are beyond the scope of TCGA.

Although expression of the *MYC* oncogene on 8q24 has been implicated in mediating the effect of 8q copy number gain in UM, our analysis reveals a more complicated scenario in which MYC/MAX complex targets were highly activated in UM with (SCNA cluster 4) or without (SCNA cluster 1) 8q gain. In contrast, the MYC/MAX/MIZ1 complex targets were most prominently activated only in samples with 8q gain, suggesting that other processes, in addition to copy number gain, e.g., post-transcriptional alterations, may also be relevant to MYC signaling in these UM subtypes.

The lncRNA *PVT1* locus is adjacent to the *MYC* locus and is coamplified with *MYC* in UM with elevated 8q copy number. Our data indicate convergent genomic (copy number) and epigenetic (DNA methylation) mechanisms of *PVT1* regulation in UM.

Figure 6. Pathway and Regulators that were Differentially Active in Transcriptional Subtypes 3 and 4

Correlation network for lncRNA clusters 3 (top) and 4 (bottom), showing PARADIGM pathway features, (hierarchical) MARINa regulators, and lncRNAs. Red and blue lines indicate Spearman correlations ($|\text{rho}| > 0.5$) between the expression of a differentially expressed lncRNA and inferred activity of a differentially active PARADIGM or MARINa feature. The color of each node reflects differential expression for a lncRNA, and relative activity for a PARADIGM/MARINa feature (red for overexpressed/active, blue for underexpressed/inactive). See also Table S4.

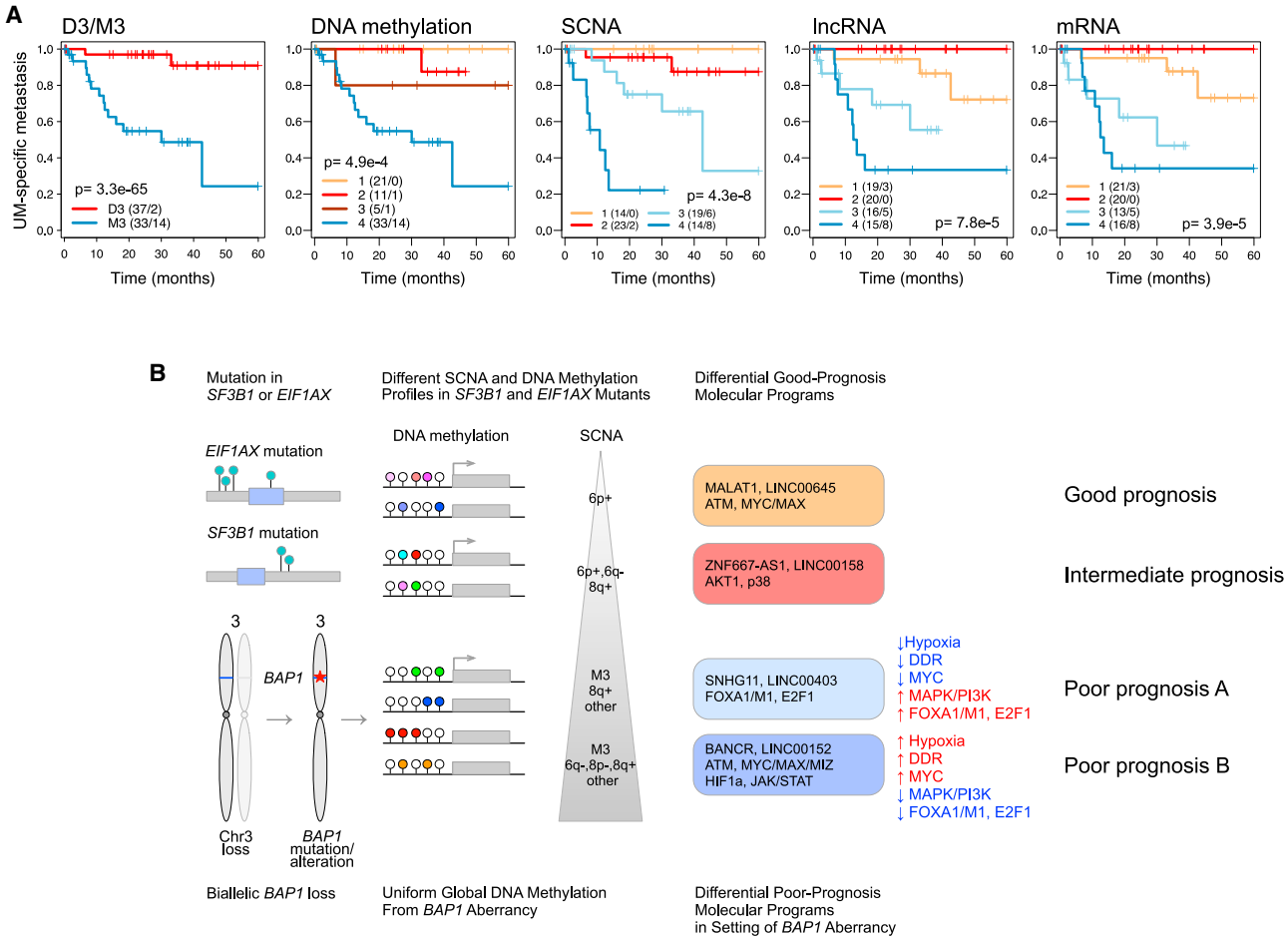


Figure 7. Good-Prognosis D3-UM and Poor-Prognosis M3-UM Separate into Distinct Biological Subsets

(A) Kaplan-Meier plots and log-rank p values for the clinical event of UM metastasis for M3- versus D3-UM, then for unsupervised clusters for DNA methylation, SCNA, lncRNA, and mRNA. The number of cases and events in a cluster are shown on the plots. Median event times for clusters 3 and 4 were 10.8 versus 42.6 months for SCNA (p = 0.002, p = 0.01 with a Bonferroni correction [BC]); 13.0 versus >30 months for lncRNA (p = 0.19, p = 1.0 with BC); and 13.5 versus 30.0 months for mRNA (p = 0.43, p = 1.0 with BC).

(B) Schematic of D3-UM and M3-UM molecular prognosis subtypes. D3-UM tumors with *EIF1AX* versus *SF3B1* mutations, which are known to be associated with low and intermediate risk of developing UM metastasis, respectively, correlated with distinct DNA methylation and SCNA profiles. D3-UM tumors also separated into two groups by transcription (mRNA, lncRNA, and miRNA) profile analysis. Loss of chromosome 3, followed by *BAP1* alteration, results in bilallelic *BAP1* loss. M3/*BAP1* aberrancy is associated with a global DNA methylation profile that is not observed in D3-UM. Despite all M3/*BAP1*-aberrant UM sharing this common DNA methylation pattern, these tumors divide into two groups by SCNA and transcription profiles, with distinct pathway features indicative of hypoxia, DDR, MYC/MAX signaling, and proliferation.

See also Figures S6 and S7; Tables S5 and S6.

Overall, our observations for *PVT1* in M3-UM are consistent with it being highly regulated by DNA methylation in renal cell carcinoma (Posa et al., 2016), acting as an independent oncogene and enhancing MYC protein levels/activity (Tseng et al., 2014). In addition, we identified other coding and non-coding genes that are associated with recurrent SCNA in UM and are candidates for further functional studies.

Not observed in our cohort, due to relatively short follow-up times, was the association between D3-UM with an *EIF1AX* versus *SF3B1* mutation and low versus intermediate risk of developing metastatic disease compared with M3-UM (Yavuzyigitoglu et al., 2016). The distinct SCNA and DNA methylation profiles we observe in *EIF1AX*- versus *SF3B1*-mutant D3-UM may contribute to the different prognoses associated with these mutually exclusive mutations.

We ultimately identified *BAP1* alterations in ~85% of M3-UM, consistent with the initial report using Sanger sequencing (Harbour et al., 2010). While next-generation sequencing (NGS) has become the standard for detecting germline and somatic *BAP1* alterations in both research and clinical settings, more than half of the *BAP1* alterations were initially missed by NGS mutation detection algorithms used in our study, and the identification of additional *BAP1* alterations required assembly-based methods. These results suggest that longer and more complex gene alterations in *BAP1*, and other genes, may be detectable only by methods that include sequence assembly.

Almost all of our UM harbored mutually exclusive hotspot mutations in *GNAQ*, *GNA11*, *CYSLTR2*, or *PLCB4*, suggesting that constitutively activated G-protein signaling plays a central role in early UM development. Furthermore, neither *CYSLTR2* nor *PLCB4* mutations preferentially localized to a specific subset of UM, consistent with mutations in these genes functioning like *GNAQ/11* mutations to drive tumorigenesis without initiating metastasis. Mutant-activated GNAQ/11 signal through PKC-α, and we show that M3/*BAP1*-aberrant tumors had elevated total and activated PKC-α (and -δ) protein levels. Thus, *BAP1* aberrancy may enhance the effector function of PKC downstream of mutant-activated GNAQ/11. These data suggest both an association between early and later genetic events in metastasis-prone UM, and that inhibiting activated PKC isoforms may require targeting downstream effects of *BAP1* aberrancy.

We identified the splicing factor gene *SRSF2* as an SMG in 4% of our UM cohort, expanding the landscape of functional spliceosome alterations in UM. We showed that UM with *SRSF2* or *SF3B1* mutations have mutation-specific mis-splicing that affects elongation initiation factors and signaling gene transcripts that are known to play a role in tumorigenesis. Previous genetic studies had identified nearly mutually exclusive mutations in *SF3B1* and *EIF1AX* in UM (Alsafadi et al., 2016; Harbour et al., 2013; Martin et al., 2013). In our cohort, UM with *SF3B1* mutations were enriched in SCNA clusters 2 and 3, while virtually absent in UM with the lowest and highest levels of aneuploidy (clusters 1 and 4 respectively). UM with *SRSF2* mutations harbored neither *EIF1AX* nor *SF3B1* mutations, and, like all but one *SF3B1*-mutated case, were observed only in SCNA clusters 2 and 3.

In many cancers an immune infiltrate within the tumor is typically associated with a better prognosis and with response to immunotherapy (Lee et al., 2016). In primary UM, in contrast, marker-specific immunohistochemistry has demonstrated that a dense infiltrate of leukocytes (Bronkhorst et al., 2012; Ksander et al., 1998) or macrophages (Bronkhorst et al., 2011; Maat et al., 2008) is associated with M3 and a poor prognosis. In our cohort, immune infiltrates were highly correlated with upregulation of chemotactic signals (e.g., CXCL9 and CXCL13) and of stimulators and targets (e.g., IFNG and HLA) that are essential in T cell-mediated immune therapies. Also in contrast with other cancers, an increased HLA class I expression has been associated with a worse prognosis in UM (de Lange et al., 2015), and is considered a tumor-escape mechanism from natural killer cell-mediated cytotoxicity in blood (Jager et al., 2002). The increased HLA class I expression in poor-prognosis UM is likely induced by infiltrating cytotoxic T cells (van Essen et al., 2016); however, the molecular immune profile of these tumors is consistent with a chronically inflamed milieu in which either T cells are more immunosuppressive (regulatory T cells) and/or cytotoxic T cells have been rendered dysfunctional (Bronkhorst et al., 2012). Notably, the immune checkpoint inhibitors IDO1 and TIGIT, which can limit the efficacy of T cell killing of cancer cells, were among the most highly expressed mRNAs in CD8-enriched M3-UM. These findings may, in part, explain the clinical observations suggesting that single-agent anti-CTLA-4 or anti-PD1 immune checkpoint inhibitors have low efficacy in patients with metastatic UM (Kelderman et al., 2013), and that agents targeting IDO1 and/or TIGIT, which are currently in clinical trials, may help overcome immune suppression in UM (Dougall et al., 2017; Manieri et al., 2017).

Pathway profiling showed that relative activity of cellular processes such as DDR, hypoxia, MYC signaling, and MAPK/AKT programs differentiated subgroups within both M3-UM and D3-UM. These results suggest that different UM subsets may require specific targeted strategies to achieve efficacy. DDR-modulating agents, anti-hypoxia drugs, direct or indirect anti-MYC therapeutics, and compounds that target these pathways are currently being investigated in human clinical trials.

This retrospective study suggests that probe-based or NGS-based copy number data should support a DNA-based clinical assay that assigns a high-risk M3-UM sample to one of two groups (SCNA subtypes 3 versus 4), which have different median times to UM metastasis. Such an approach would have the advantage of also identifying isodisomy 3 tumors, which are not detected by fluorescence *in situ* hybridization or array comparative genomic hybridization, and which have a similar prognosis to M3-UM tumors. In addition, we identified coding and non-coding genes that were differentially expressed between M3-UM SCNA subtypes 3 versus 4 and associated with UM metastasis. We showed that a number of these transcripts, particularly certain 8q transcripts, are associated with M3-UM metastasis in an independent cohort.

Developing a clinically relevant classifier will require prospective evaluation of copy number and/or gene expression data in tumors with similar clinical-pathological features to identify patients with higher- versus lower-risk M3-UM, and to validate the differential UM metastasis intervals observed in this study. Such a classifier could influence the frequency of metastatic surveillance, prioritize high-risk patients for more aggressive/earlier adjuvant clinical trials, provide more precise UM metastasis data for the design of clinical trials and use of historical controls, and offer information to patients that may assist them in medical and personal choices. As no effective adjuvant therapy has yet been developed for UM, a prospective analysis of characterizing these two molecular subtypes relative to UM metastasis is especially timely and important.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - *BAP1* Terminology
  - Biospecimen and Clinical Data Processing
  - Whole Exome Sequencing (WES)
  - SNP-based Copy Number Analysis
  - RNA Sequencing
  - Non-Coding RNA Sequencing
  - DNA Methylation
  - Low-Pass Whole Genome Sequencing
  - Reverse Phase Protein Arrays (RPPA)
  - Microbial Detection

- Regulome Explorer
- cBioPortal Visualization
- PARADIGM Integrated Pathway Analysis
- MARINa/hMARINa Analysis of Regulator Activity
- Identifying Pathway Features Differentiating lncRNA Clusters
- lncRNA Pathway Regulator Correlation Networks
- Relationship of Fold Change between TCGA SCNA Clusters 3 vs. 4, and Association with Time to Metastasis in TCGA and Laurent Monosomy 3 Cases

● QUANTIFICATION AND STATISTICAL ANALYSIS
● DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and six tables and can be found with this article online at http://dx.doi.org/10.1016/j.ccell.2017.07.003.

CONSORTIA

Mohamed H. Abdel-Rahman, Rehan Akbani, Adrian Ally, J. Todd Auman, Ozgun Babur, Miruna Balasundaram, Saianand Balu, Christopher Benz, Rameen Beroukhim, Inanc Birol, Tom Bodenheimer, Jay Bowen, Reanne Bowlby, Christopher A. Bristow, Denise Brooks, Rebecca Carlsen, Colleen M. Cebulla, Matthew T. Chang, Andrew D. Cherniack, Lynda Chin, Juok Cho, Eric Chuah, Sudha Chudamani, Carrie Cibulskis, Kristian Cibulskis, Leslie Cope, Sarah E. Coupland, Ludmila Danilova, Timothy Defreitas, John A. Demchok, Laurence Desjardins, Noreen Dhalla, Bita Esmaeli, Ina Felau, Martin L. Ferguson, Scott Frazer, Stacey B. Gabriel, Julie M. Gastier-Foster, Nils Gehlenborg, Mark Gerken, Jeffrey E. Gershenwald, Gad Getz, Ewan A. Gibb, Klaus G. Griewank, Elizabeth A. Grimm, D. Neil Hayes, Apurva M. Hegde, David I. Heiman, Carmen Helsel, Julian M. Hess, Katherine A. Hoadley, Shital Hobensack, Robert A. Holt, Alan P. Hoyle, Xin Hu, Carolyn M. Hutter, Lisa Iype, Martine J. Jager, Stuart R. Jefferys, Corbin D. Jones, Steven J.M. Jones, Cyriac Kandoth, Katayoon Kasaian, Jaegil Kim, Patrick K. Kimes, Melanie Kucherlapati, Raju Kucherlapati, Eric Lander, Michael S. Lawrence, Alexander J. Lazar, Semin Lee, Kristen M. Leraas, Tara M. Lichtenberg, Pei Lin, Jia Liu, Wenbin Liu, Laxmi Lolla, Yiling Lu, Yussanne Ma, Harshad S. Mahadeshwar, Odette Mariani, Marco A. Marra, Michael Mayo, Sam Meier, Shaowu Meng, Matthew Meyerson, Piotr A. Mieczkowski, Gordon B. Mills, Richard A. Moore, Lisle E. Mose, Andrew J. Mungall, Karen L. Mungall, Bradley A. Murray, Rashi Naresh, Michael S. Noble, Junna Oba, Angeliki Pantazi, Michael Parfenov, Peter J. Park, Joel S. Parker, Alexander Penson, Charles M. Perou, Todd Pihl, Robert Pilarski, Alexei Protopopov, Amie Radenbaugh, Karan Rai, Nilsa C. Ramirez, Xiaojia Ren, Sheila M. Reynolds, Jeffrey Roach, A. Gordon Robertson, Sergio Roman-Roman, Jason Roszik, Sara Sadeghi, Gordon Saksena, Xavier Sastre, Dirk Schadendorf, Jacqueline E. Schein, Lynn Schoenfield, Steven E. Schumacher, Jonathan Seidman, Sahil Seth, Geetika Sethi, Margi Sheth, Yan Shi, Carol Shields, Juliann Shih, Ilya Shmulevich, Janae V. Simons, Arun D. Singh, Payal Sipahimalani, Tara Skelly, Heidi Sofia, Matthew G. Soloway, Xingzhi Song, Marc-Henri Stern, Joshua Stuart, Huandong Sun, Qiang Sun, Angela Tam, Donghui Tan, Jiabin Tang, Ming Tang, Roy Tarnuzzer, Barry S. Taylor, Nina Thiessen, Vesteinn Thorsson, Kane Tse, Vladislav Uzunangelov, Umadevi Veluvolu, Roel G.W. Verhaak, Doug Voet, Vonn Walter, Yunhu Wan, Zhining Wang, John N. Weinstein, Matthew D. Wilkerson, Michelle D. Williams, Lisa Wise, Scott E. Woodman, Tina Wong, Ye Wu, Liming Yang, Lixing Yang, Christina Yau, Jean C. Zenklusen, Hailei Zhang, Jiashan Zhang, Erik Zmuda.

AUTHOR CONTRIBUTIONS

The Cancer Genome Atlas research network contributed collectively to this work. Special thanks go out to the following network members who made substantial contributions. Supervision: B.E., C.K., A.G.R., and S.E.W. Data and analysis coordinator: S.E.W. Manuscript coordinator: S.E.W. NIH project coordinator: I.F. Formal analysis of DNA sequence: J.M.H. Low-pass DNA: M.K. and C.A.B. DNA methylation: L.D. mRNA: V.W., C.A.B., K.A.H., A.P., K.L.M., and A.G.R. miRNA: A.G.R. LncRNA: E.A.G. and A.G.R. RNA fusion: R.G.W.V. Copy number: J.S., M.T.C., and A.D.C. Mutual exclusivity: O.B. and M.T. RPPA: R.A., O.B., S.E.W., G.B.M., and A.G.R. Pathways: C.Y., V.U., C.B., O.B., and M.T. Regulome explorer: L.I. Clinical data: T.M.L., J.O., J.E.G., L.S., C.M.C., K.G.G., M.D.W., M.J.J., S.E.C., B.E., and S.E.W. Tissue resources: M.-H.S., S.R.-R., S.E.C., B.E., C.M.C., and M.H.A. Pathology review: L.S., A.J.L., M.H.A., M.D.W., and S.E.C. Manuscript writing: A.G.R., J.O., J.S., C.Y., E.A.G., K.L.M., J.M.H., V.U., V.W., L.D., T.M.L., M.K., P.K.K., M.T., A.D.C., C.B., L.S., M.H.A., S.R.-R., M.-H.S., C.M.C., M.D.W., M.J.J., S.E.C., B.E., C.K., and S.E.W.

ACKNOWLEDGMENTS

We are grateful to all patients and families who contributed to this study. We thank Dr. Jasmine Francis, Dr. Amy Schefler and Dr Helen Kalirai for following through with the requirements for sample submission. This work was supported by the following grants from the NIH: U54 HG003273, U54 HG003067, U54 HG003079, U24 CA143799, U24 CA143835, U24 CA143840, U24 CA143843, U24 CA143845, U24 CA143848, U24 CA143858, U24 CA143866, U24 CA143867, U24 CA143882, U24 CA143883, U24 CA144025, P30 CA016672, P50 CA083639, and K08 EY022672 (C.M.C.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. E.A.G. is an employee of GenomeDx Biosciences. A.D.C. declares research funding from Bayer AG. J.E.G. has an advisory role in Castle Biosciences and Merck. K.G.G. holds the patent for WO2011130691: GNA11 and GNAQ exon 4 mutations in melanoma.

Received: December 14, 2016
Revised: April 24, 2017
Accepted: July 9, 2017
Published: August 14, 2017

REFERENCES

Abdel-Rahman, M.H., Pilarski, R., Cebulla, C.M., Massengill, J.B., Christopher, B.N., Boru, G., Hovland, P., and Davidorf, F.H. (2011). Germline BAP1 mutation predisposes to uveal melanoma, lung adenocarcinoma, meningioma, and other cancers. J. Med. Genet. *48*, 856–859.

Akbani, R., Ng, P.K., Werner, H.M., Shahmoradgoli, M., Zhang, F., Ju, Z., Liu, W., Yang, J.Y., Yoshihara, K., Li, J., et al. (2014). A pan-cancer proteomic perspective on The Cancer Genome Atlas. Nat. Commun. *5*, 3887.

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Borresen-Dale, A.L., et al. (2013). Signatures of mutational processes in human cancer. Nature *500*, 415–421.

Alsafadi, S., Houy, A., Battistella, A., Popova, T., Wassef, M., Henry, E., Tirode, F., Constantinou, A., Piperno-Neumann, S., Roman-Roman, S., et al. (2016). Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. Nat. Commun. *7*, 10615.

Alvarez, M.J., Shen, Y., Giorgi, F.M., Lachmann, A., Ding, B.B., Ye, B.H., and Califano, A. (2016). Functional characterization of somatic mutations in cancer using network-based inference of protein activity. Nat. Genet. *48*, 838–847.

Anders, S., Reyes, A., and Huber, W. (2012). Detecting differential usage of exons from RNA-seq data. Genome Res. *22*, 2008–2017.

Aytes, A., Mitrofanova, A., Lefebvre, C., Alvarez, M.J., Castillo-Martin, M., Zheng, T., Eastham, J.A., Gopalan, A., Pienta, K.J., Shen, M.M., et al. (2014). Cross-species regulatory network analysis identifies a synergistic interaction between FOXM1 and CENPF that drives prostate cancer malignancy. Cancer Cell *25*, 638–651.

Babur, O., Gonen, M., Aksoy, B.A., Schultz, N., Ciriello, G., Sander, C., and Demir, E. (2015). Systematic identification of cancer driving signaling pathways based on mutual exclusivity of genomic alterations. Genome Biol. *16*, 45.

Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J.M., Delano, D., Zhang, L., Schroth, G.P., Gunderson, K.L., et al. (2011). High density DNA methylation array with single CpG site resolution. Genomics *98*, 288–295.

Blum, E.S., Yang, J., Komatsubara, K.M., and Carvajal, R.D. (2016). Clinical management of uveal and conjunctival melanoma. Oncology (Williston Park) *30*, 29–32, 34–43, 48.

Bronkhorst, I.H., Ly, L.V., Jordanova, E.S., Vrolijk, J., Versluis, M., Luyten, G.P., and Jager, M.J. (2011). Detection of M2-macrophages in uveal melanoma and relation with survival. Invest. Ophthalmol. Vis. Sci. *52*, 643–650.

Bronkhorst, I.H., Vu, T.H., Jordanova, E.S., Luyten, G.P., Burg, S.H., and Jager, M.J. (2012). Different subsets of tumor-infiltrating lymphocytes correlate with macrophage influx and monosomy 3 in uveal melanoma. Invest. Ophthalmol. Vis. Sci. *53*, 5370–5378.

Caines, R., Eleuteri, A., Kalirai, H., Fisher, A.C., Heimann, H., Damato, B.E., Coupland, S.E., and Taktak, A.F. (2015). Cluster analysis of multiplex ligation-dependent probe amplification data in choroidal melanoma. Mol. Vis. *21*, 1–11.

Cancer Genome Atlas Research Network. (2012). Comprehensive genomic characterization of squamous cell lung cancers. Nature *489*, 519–525.

Cancer Genome Atlas Research Network. (2014a). Comprehensive molecular characterization of gastric adenocarcinoma. Nature *513*, 202–209.

Cancer Genome Atlas Research Network. (2014b). Comprehensive molecular profiling of lung adenocarcinoma. Nature *511*, 543–550.

Cancer Genome Atlas Research Network. (2014c). Integrated genomic characterization of papillary thyroid carcinoma. Cell *159*, 676–690.

Cancer Genome Atlas Research Network. (2015). Genomic classification of cutaneous melanoma. Cell *161*, 1681–1696.

Carter, S.L., Cibulskis, K., Helman, E., McKenna, A., Shen, H., Zack, T., Laird, P.W., Onofrio, R.C., Winckler, W., Weir, B.A., et al. (2012). Absolute quantification of somatic DNA alterations in human cancer. Nat. Biotechnol. *30*, 413–421.

Cassoux, N., Rodrigues, M.J., Plancher, C., Asselain, B., Levy-Gabriel, C., Lumbroso-Le Rouic, L., Piperno-Neumann, S., Dendale, R., Sastre, X., Desjardins, L., and Couturier, J. (2014). Genome-wide profiling is a clinically relevant and affordable prognostic test in posterior uveal melanoma. Br. J. Ophthalmol. *98*, 769–774.

Challis, D., Yu, J., Evani, U.S., Jackson, A.R., Paithankar, S., Coarfa, C., Milosavljevic, A., Gibbs, R.A., and Yu, F. (2012). An integrative variant analysis suite for whole exome next-generation sequencing data. BMC Bioinformatics *13*, 8.

Chattopadhyay, C., Kim, D.W., Gombos, D.S., Oba, J., Qin, Y., Williams, M.D., Esmaeli, B., Grimm, E.A., Wargo, J.A., Woodman, S.E., and Patel, S.P. (2016). Uveal melanoma: from diagnosis to treatment and the science in between. Cancer *122*, 2299–2312.

Chen, K., Wallis, J.W., McLellan, M.D., Larson, D.E., Kalicki, J.M., Pohl, C.S., McGrath, S.D., Wendl, M.C., Zhang, Q., Locke, D.P., et al. (2009). BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. Nat. Methods *6*, 677–681.

Chou, C.H., Chang, N.W., Shrestha, S., Hsu, S.D., Lin, Y.L., Lee, W.H., Yang, C.D., Hong, H.C., Wei, T.Y., Tu, S.J., et al. (2016). miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. Nucleic Acids Res. *44*, D239–D247.

Chu, J., Sadeghi, S., Raymond, A., Jackman, S.D., Nip, K.M., Mar, R., Mohamadi, H., Butterfield, Y.S., Robertson, A.G., and Birol, I. (2013). BioBloom tools: fast, accurate and memory-efficient host species sequence screening using bloom filters. Bioinformatics *30*, 3402–3404.

Chu, A., Robertson, G., Brooks, D., Mungall, A.J., Birol, I., Coope, R., Ma, Y., Jones, S., and Marra, M.A. (2016). Large-scale profiling of microRNAs for The Cancer Genome Atlas. Nucleic Acids Res. *44*, e3.

Cibulskis, K., McKenna, A., Fennell, T., Banks, E., DePristo, M., and Getz, G. (2011). ContEst: estimating cross-contamination of human samples in next-generation sequencing data. Bioinformatics *27*, 2601–2602.

Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S., and Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat. Biotechnol. *31*, 213–219.

Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly (Austin) *6*, 80–92.

Colombo, T., Farina, L., Macino, G., and Paci, P. (2015). PVT1: a rising star among oncogenic long noncoding RNAs. Biomed. Res. Int. *2015*, 304208.

Costello, M., Pugh, T.J., Fennell, T.J., Stewart, C., Lichtenstein, L., Meldrim, J.C., Fostel, J.L., Friedrich, D.C., Perrin, D., Dionne, D., et al. (2013). Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. Nucleic Acids Res. *41*, e67.

Coupland, S.E., Lake, S.L., Zeschnigk, M., and Damato, B.E. (2013). Molecular pathology of uveal melanoma. Eye (Lond) *27*, 230–242.

Dabney, A.R. (2006). ClaNC: point-and-click software for classifying microarrays to nearest centroids. Bioinformatics *22*, 122–123.

Damato, B., Dopierala, J.A., and Coupland, S.E. (2010). Genotypic profiling of 452 choroidal melanomas with multiplex ligation-dependent probe amplification. Clin. Cancer Res. *16*, 6083–6092.

de Lange, M.J., van Pelt, S.I., Versluis, M., Jordanova, E.S., Kroes, W.G., Ruivenkamp, C., van der Burg, S.H., Luyten, G.P., van Hall, T., Jager, M.J., et al. (2015). Heterogeneity revealed by integrated genomic analysis uncovers a molecular switch in malignant uveal melanoma. Oncotarget *6*, 37824–37835.

Diener-West, M., Reynolds, S.M., Agugliaro, D.J., Caldwell, R., Cumming, K., Earle, J.D., Hawkins, B.S., Hayman, J.A., Jaiyesimi, I., Kirkwood, J.M., et al. (2005). Second primary cancers after enrollment in the COMS trials for treatment of choroidal melanoma: COMS Report No. 25. Arch. Ophthalmol. *123*, 601–604.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Dougall, W.C., Kurtulus, S., Smyth, M.J., and Anderson, A.C. (2017). TIGIT and CD96: new checkpoint receptor targets for cancer immunotherapy. Immunol. Rev. *276*, 112–120.

Eletr, Z.M., Yin, L., and Wilkinson, K.D. (2013). BAP1 is phosphorylated at serine 592 in S-phase following DNA damage. FEBS Lett. *587*, 3906–3911.

Faith, J.J., Hayete, B., Thaden, J.T., Mogno, I., Wierzbowski, J., Cottarel, G., Kasif, S., Collins, J.J., and Gardner, T.S. (2007). Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. PLoS Biol. *5*, e8.

Fisher, S., Barry, A., Abreu, J., Minie, B., Nolan, J., Delorey, T.M., Young, G., Fennell, T.J., Allen, A., Ambrogio, L., et al. (2011). A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. Genome Biol. *12*, R1.

Forbes, S.A., Tang, G., Bindal, N., Bamford, S., Dawson, E., Cole, C., Kok, C.Y., Jia, M., Ewing, R., Menzies, A., et al. (2010). COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. Nucleic Acids Res. *38*, D652–D657.

Gaujoux, R., and Seoighe, C. (2010). A flexible R package for nonnegative matrix factorization. BMC Bioinformatics *11*, 367.

Gonzalez-Angulo, A.M., Hennessy, B.T., Meric-Bernstam, F., Sahin, A., Liu, W., Ju, Z., Carey, M.S., Myhre, S., Speers, C., Deng, L., et al. (2011). Functional proteomics can define prognosis and predict pathologic complete response in patients with breast cancer. Clin. Proteomics *8*, 11.

Harbour, J.W. (2014). A prognostic test to predict the risk of metastasis in uveal melanoma based on a 15-gene expression profile. Methods Mol. Biol. *1102*, 427–440.

Harbour, J.W., Onken, M.D., Roberson, E.D., Duan, S., Cao, L., Worley, L.A., Council, M.L., Matatall, K.A., Helms, C., and Bowcock, A.M. (2010). Frequent mutation of BAP1 in metastasizing uveal melanomas. Science *330*, 1410–1413.

Harbour, J.W., Roberson, E.D., Anbunathan, H., Onken, M.D., Worley, L.A., and Bowcock, A.M. (2013). Recurrent mutations at codon 625 of the splicing factor SF3B1 in uveal melanoma. Nat. Genet. *45*, 133–135.

Hon, C.C., Ramilowski, J.A., Harshbarger, J., Bertin, N., Rackham, O.J., Gough, J., Denisenko, E., Schmeier, S., Poulsen, T.M., Severin, J., et al. (2017). An atlas of human long non-coding RNAs with accurate 5′ ends. Nature *543*, 199–204.

Hornbeck, P.V., Zhang, B., Murray, B., Kornhauser, J.M., Latham, V., and Skrzypek, E. (2014). PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Res. *43*, D512–D520.

Hu, J., He, X., Baggerly, K.A., Coombes, K.R., Hennessy, B.T., and Mills, G.B. (2007). Non-parametric quantification of protein lysate arrays. Bioinformatics *23*, 1986–1994.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. *37*, 1–13.

Huang, H., Liu, Y., Yuan, M., and Marron, J.S. (2015). Statistical significance of clustering using soft thresholding. J. Comput. Graph Stat. *24*, 975–993.

Ismail, I.H., Davidson, R., Gagne, J.P., Xu, Z.Z., Poirier, G.G., and Hendzel, M.J. (2014). Germline mutations in BAP1 impair its function in DNA double-strand break repair. Cancer Res. *74*, 4282–4294.

Jager, M.J., Hurks, H.M., Levitskaya, J., and Kiessling, R. (2002). HLA expression in uveal melanoma: there is no rule without some exception. Hum. Immunol. *63*, 444–451.

Johansson, P., Aoude, L.G., Wadt, K., Glasson, W.J., Warrier, S.K., Hewitt, A.W., Kiilgaard, J.F., Heegaard, S., Isaacs, T., Franchina, M., et al. (2016). Deep sequencing of uveal melanoma identifies a recurrent mutation in PLCB4. Oncotarget *7*, 4624–4631.

Johnson, D.B., Roszik, J., Shoushtari, A.N., Eroglu, Z., Balko, J.M., Higham, C., Puzanov, I., Patel, S.P., Sosman, J.A., and Woodman, S.E. (2016). Comparative analysis of the GNAQ, GNA11, SF3B1, and EIF1AX driver mutations in melanoma and across the cancer spectrum. Pigment Cell Melanoma Res *29*, 470–473.

Johnson, C.P., Kim, I.K., Esmaeli, B., Amin-Mansour, A., Treacy, D.J., Carter, S.L., Hodis, E., Wagle, N., Seepo, S., Yu, X., et al. (2017). Systematic genomic and translational efficiency studies of uveal melanoma. PLoS One *12*, e0178189.

Ju, Z., Liu, W., Roebuck, P.L., Siwak, D.R., Zhang, N., Lu, Y., Davies, M.A., Akbani, R., Weinstein, J.N., Mills, G.B., and Coombes, K.R. (2015). Development of a robust classifier for quality control of reverse-phase protein arrays. Bioinformatics *31*, 912–918.

Kalirai, H., Dodson, A., Faqir, S., Damato, B.E., and Coupland, S.E. (2014). Lack of BAP1 protein expression in uveal melanoma is associated with increased metastatic risk and has utility in routine prognostic testing. Br. J. Cancer *111*, 1373–1380.

Katz, Y., Wang, E.T., Silterra, J., Schwartz, S., Wong, B., Thorvaldsdottir, H., Robinson, J.T., Mesirov, J.P., Airoldi, E.M., and Burge, C.B. (2015). Quantitative visualization of alternative exon expression from RNA-seq data. Bioinformatics *31*, 2400–2402.

Kelderman, S., van der Kooij, M.K., van den Eertwegh, A.J., Soetekouw, P.M., Jansen, R.L., van den Brom, R.R., Hospers, G.A., Haanen, J.B., Kapiteijn, E., and Blank, C.U. (2013). Ipilimumab in pretreated metastatic uveal melanoma patients. Results of the Dutch Working group on Immunotherapy of Oncology (WIN-O). Acta Oncol. *52*, 1786–1788.

Khurana, E., Fu, Y., Chen, J., and Gerstein, M. (2013). Interpretation of genomic variants using a unified biological network approach. PLoS Comput. Biol. *9*, e1002886.

Kim, E., Ilagan, J.O., Liang, Y., Daubner, G.M., Lee, S.C., Ramakrishnan, A., Li, Y., Chung, Y.R., Micol, J.B., Murphy, M.E., et al. (2015). SRSF2 mutations contribute to myelodysplasia by mutant-specific effects on exon recognition. Cancer Cell *27*, 617–630.

Koopmans, A.E., Verdijk, R.M., Brouwer, R.W., van den Bosch, T.P., van den Berg, M.M., Vaarwater, J., Kockx, C.E., Paridaens, D., Naus, N.C., Nellist, M., et al. (2014). Clinical significance of immunohistochemistry for detection of BAP1 mutations in uveal melanoma. Mod. Pathol. *27*, 1321–1330.

Korn, J.M., Kuruvilla, F.G., McCarroll, S.A., Wysoker, A., Nemesh, J., Cawley, S., Hubbell, E., Veitch, J., Collins, P.J., Darvishi, K., et al. (2008). Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. Nat. Genet. *40*, 1253–1260.

Kress, T.R., Sabo, A., and Amati, B. (2015). MYC: connecting selective transcriptional control to global RNA production. Nat. Rev. Cancer *15*, 593–607.

Ksander, B.R., Geer, D.C., Chen, P.W., Salgaller, M.L., Rubsamen, P., and Murray, T.G. (1998). Uveal melanomas contain antigenically specific and non-specific infiltrating lymphocytes. Curr. Eye Res. *17*, 165–173.

Lachmann, A., Xu, H., Krishnan, J., Berger, S.I., Mazloom, A.R., and Ma'ayan, A. (2010). ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. Bioinformatics *26*, 2438–2444.

Laurent, C., Valet, F., Planque, N., Silveri, L., Maacha, S., Anezo, O., Hupe, P., Plancher, C., Reyes, C., Albaud, B., et al. (2011). High PTP4A3 phosphatase expression correlates with metastatic risk in uveal melanoma patients. Cancer Res. *71*, 666–674.

Lawrence, M.S., Stojanov, P., Mermel, C.H., Robinson, J.T., Garraway, L.A., Golub, T.R., Meyerson, M., Gabriel, S.B., Lander, E.S., and Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. Nature *505*, 495–501.

Lee, N., Zakka, L.R., Mihm, M.C., Jr., and Schatton, T. (2016). Tumour-infiltrating lymphocytes in melanoma prognosis and cancer immunotherapy. Pathology *48*, 177–187.

Lefebvre, C., Rajbhandari, P., Alvarez, M.J., Bandaru, P., Lim, W.K., Sato, M., Wang, K., Sumazin, P., Kustagi, M., Bisikirska, B.C., et al. (2010). A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. Mol. Syst. Biol. *6*, 377.

Leiserson, M.D., Wu, H.T., Vandin, F., and Raphael, B.J. (2015). CoMEt: a statistical approach to identify combinations of mutually exclusive alterations in cancer. Genome Biol. *16*, 160.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589–595.

Li, J., and Tibshirani, R. (2013). Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-Seq data. Stat. Methods Med. Res. *22*, 519–536.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079.

Maat, W., Ly, L.V., Jordanova, E.S., de Wolff-Rouendaal, D., Schalij-Delfos, N.E., and Jager, M.J. (2008). Monosomy of chromosome 3 and an inflammatory phenotype occur together in uveal melanoma. Invest. Ophthalmol. Vis. Sci. *49*, 505–510.

Manieri, N.A., Chiang, E.Y., and Grogan, J.L. (2017). TIGIT: a key inhibitor of the cancer immunity cycle. Trends Immunol. *38*, 20–28.

Manning, G., Whyte, D.B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002). The protein kinase complement of the human genome. Science *298*, 1912–1934.

Marsico, A., Huska, M.R., Lasserre, J., Hu, H., Vucicevic, D., Musahl, A., Orom, U., and Vingron, M. (2013). PROmiRNA: a new miRNA promoter recognition method uncovers the complex regulation of intronic miRNAs. Genome Biol. *14*, R84.

Martin, M., Masshofer, L., Temming, P., Rahmann, S., Metz, C., Bornfeld, N., van de Nes, J., Klein-Hitpass, L., Hinnebusch, A.G., Horsthemke, B., et al. (2013). Exome sequencing identifies recurrent somatic mutations in EIF1AX and SF3B1 in uveal melanoma with disomy 3. Nat. Genet. *45*, 933–936.

Mayakonda, A., and Koeffler, H.P. (2016). Maftools: efficient analysis, visualization and summarization of MAF files from large-scale cohort based cancer studies. bioRxiv. http://dx.doi.org/10.1101/052662.

McCarroll, S.A., Kuruvilla, F.G., Korn, J.M., Cawley, S., Nemesh, J., Wysoker, A., Shapero, M.H., de Bakker, P.I., Maller, J.B., et al. (2008). Integrated detection and population-genetic analysis of SNPs and copy number variation. Nat. Genet. *40*, 1166–1174.

McCarthy, C., Kalirai, H., Lake, S.L., Dodson, A., Damato, B.E., and Coupland, S.E. (2016). Insights into genetic alterations of liver metastases from uveal melanoma. Pigment Cell Melanoma Res *29*, 60–67.

McPherson, A., Hormozdiari, F., Zayed, A., Giuliany, R., Ha, G., Sun, M.G., Griffith, M., Heravi Moussavi, A., Senz, J., Melnyk, N., et al. (2011). deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data. Plos Comput. Biol. *7*, e1001138.

Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., Beroukhim, R., and Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. *12*, R41.

Moore, A.R., Ceraudo, E., Sher, J.J., Guan, Y., Shoushtari, A.N., Chang, M.T., Zhang, J.Q., Walczak, E.G., Kazmi, M.A., Taylor, B.S., et al. (2016). Recurrent activating mutations of G-protein-coupled receptor CYSLTR2 in uveal melanoma. Nat. Genet. *48*, 675–680.

Mullokandov, G., Baccarini, A., Ruzo, A., Jayaprakash, A.D., Tung, N., Israelow, B., Evans, M.J., Sachidanandam, R., and Brown, B.D. (2012). High-throughput assessment of microRNA activity and function using microRNA sensor and decoy libraries. Nat. Methods *9*, 840–846.

Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. Nat. Methods *12*, 453–457.

Nguyen, Q., and Carninci, P. (2016). Expression specificity of disease-associated lncRNAs: toward personalized medicine. Curr. Top. Microbiol. Immunol. *394*, 237–258.

Olshen, A.B., Venkatraman, E.S., Lucito, R., and Wigler, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics *5*, 557–572.

Pang, Q., Ge, J., Shao, Y., Sun, W., Song, H., Xia, T., Xiao, B., and Guo, J. (2014). Increased expression of long intergenic non-coding RNA LINC00152 in gastric cancer and its clinical significance. Tumour Biol. *35*, 5441–5447.

Petryszak, R., Burdett, T., Fiorelli, B., Fonseca, N.A., Gonzalez-Porta, M., Hastings, E., Huber, W., Jupp, S., Keays, M., Kryvych, N., et al. (2014). Expression Atlas update—a database of gene and transcript expression from microarray- and sequencing-based functional genomics experiments. Nucleic Acids Res. *42*, D926–D932.

Poon, H., Quirk, C., DeZiel, C., and Heckerman, D. (2014). Literome: PubMed-scale genomic knowledge base in the cloud. Bioinformatics *30*, 2840–2842.

Posa, I., Carvalho, S., Tavares, J., and Grosso, A.R. (2016). A pan-cancer analysis of MYC-PVT1 reveals CNV-unmediated deregulation and poor prognosis in renal carcinoma. Oncotarget *7*, 47033–47041.

Van Raamsdonk, C.D., Bezrookove, V., Green, G., Bauer, J., Gaugler, L., O'Brien, J.M., Simpson, E.M., Barsh, G.S., and Bastian, B.C. (2009). Frequent somatic mutations of GNAQ in uveal melanoma and blue naevi. Nature *457*, 599–602.

Van Raamsdonk, C.D., Griewank, K.G., Crosby, M.B., Garrido, M.C., Vemula, S., Wiesner, T., Obenauf, A.C., Wackernagel, W., Green, G., Bouvier, N., et al. (2010). Mutations in GNA11 in uveal melanoma. N. Engl. J. Med. *363*, 2191–2199.

Radenbaugh, A.J., Ma, S., Ewing, A., Stuart, J.M., Collisson, E.A., Zhu, J., and Haussler, D. (2014). RADIA: RNA and DNA integrated analysis for somatic mutation detection. PLoS One *9*, e111516.

Reinius, L.E., Acevedo, N., Joerink, M., Pershagen, G., Dahlen, S.E., Greco, D., Soderhall, C., Scheynius, A., and Kere, J. (2012). Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. PLoS One *7*, e41361.

Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S.D., Mungall, K., Lee, S., Okada, H.M., Qian, J.Q., et al. (2010). De novo assembly and analysis of RNA-seq data. Nat. Methods *7*, 909–912.

Saunders, C.T., Wong, W.S., Swamy, S., Becq, J., Murray, L.J., and Cheetham, R.K. (2012). Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics *28*, 1811–1817.

Schoenfield, L. (2014). Uveal melanoma: a pathologist's perspective and review of translational developments. Adv. Anat. Pathol. *21*, 138–143.

Sedgewick, A.J., Benz, S.C., Rabizadeh, S., Soon-Shiong, P., and Vaske, C.J. (2013). Learning subgroup-specific regulatory interactions and regulator independence with PARADIGM. Bioinformatics *29*, i62–70.

Shabalin, A.A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics *28*, 1353–1358.

Shen, R., and Seshan, V. (2016). FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. Nucleic Acids Res. *44*, e131.

Shen, Y., Wan, Z., Coarfa, C., Drabek, R., Chen, L., Ostrowski, E.A., Liu, Y., Weinstock, G.M., Wheeler, D.A., Gibbs, R.A., and Yu, F. (2010). A SNP discovery method to assess variant allele probability from next-generation resequencing data. Genome Res. *20*, 273–280.

Shen, S., Park, J.W., Lu, Z.X., Lin, L., Henry, M.D., Wu, Y.N., Zhou, Q., and Xing, Y. (2014). rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proc. Natl. Acad. Sci. USA *111*, E5593–E5601.

Shields, C.L., Ganguly, A., Bianciotto, C.G., Turaka, K., Tavallali, A., and Shields, J.A. (2010). Prognosis of uveal melanoma in 500 cases using genetic testing of fine-needle aspiration biopsy specimens. Ophthalmology *118*, 396–401.

Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J., and Birol, I. (2009). ABySS: a parallel assembler for short read sequence data. Genome Res. *19*, 1117–1123.

Singh, A.D., Turell, M.E., and Topham, A.K. (2011). Uveal melanoma: trends in incidence, treatment, and survival. Ophthalmology *118*, 1881–1885.

Smigielski, E.M., Sirotkin, K., Ward, M., and Sherry, S.T. (2000). dbSNP: a database of single nucleotide polymorphisms. Nucleic Acids Res. *28*, 352–355.

Tetzlaff, M.T., Pattanaprichakul, P., Wargo, J., Fox, P.S., Patel, K.P., Estrella, J.S., Broaddus, R.R., Williams, M.D., Davies, M.A., Routbort, M.J., et al. (2015). Utility of BRAF V600E immunohistochemistry expression pattern as a surrogate of BRAF mutation status in 154 patients with advanced melanoma. Hum. Pathol. *46*, 1101–1110.

Thorvaldsdottir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform *14*, 178–192.

Tibes, R., Qiu, Y., Lu, Y., Hennessy, B., Andreeff, M., Mills, G.B., and Kornblau, S.M. (2006). Reverse phase protein array: validation of a novel proteomic technology and utility for analysis of primary leukemia specimens and hematopoietic stem cells. Mol. Cancer Ther. *5*, 2512–2521.

Torres-Garcia, W., Zheng, S., Sivachenko, A., Vegesna, R., Wang, Q., Yao, R., Berger, M.F., Weinstein, J.N., Getz, G., and Verhaak, R.G. (2014). PRADA: pipeline for RNA sequencing data analysis. Bioinformatics *30*, 2224–2226.

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat. Biotechnol. *31*, 46–53.

Triche, T.J., Jr., Weisenberger, D.J., Van Den Berg, D., Laird, P.W., and Siegmund, K.D. (2013). Low-level processing of Illumina Infinium DNA Methylation BeadArrays. Nucleic Acids Res. *41*, e90.

Tschentscher, F., Husing, J., Holter, T., Kruse, E., Dresen, I.G., Jockel, K.H., Anastassiou, G., Schilling, H., Bornfeld, N., Horsthemke, B., et al. (2003). Tumor classification based on gene expression profiling shows that uveal melanomas with and without monosomy 3 represent two distinct entities. Cancer Res. *63*, 2578–2584.

Tseng, Y.Y., Moriarity, B.S., Gong, W., Akiyama, R., Tiwari, A., Kawakami, H., Ronning, P., Reuland, B., Guenther, K., Beadnell, T.C., et al. (2014). PVT1 dependence in cancer with MYC copy-number increase. Nature *512*, 82–86.

UniProt Consortium (2014). UniProt: a hub for protein information. Nucleic Acids Res. *43*, D204–D212.

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics *43*, 11.10.1–11.10.33.

van den Bosch, T., van Beek, J.G., Vaarwater, J., Verdijk, R.M., Naus, N.C., Paridaens, D., de Klein, A., and Kilic, E. (2012). Higher percentage of FISH-determined monosomy 3 and 8q amplification in uveal melanoma cells relate to poor patient prognosis. Invest. Ophthalmol. Vis. Sci. *53*, 2668–2674.

van Essen, T.H., van Pelt, S.I., Bronkhorst, I.H., Versluis, M., Nemati, F., Laurent, C., Luyten, G.P., van Hall, T., van den Elsen, P.J., van der Velden, P.A., et al. (2016). Upregulation of HLA expression in primary uveal melanoma by infiltrating leukocytes. PLoS One *11*, e0164292.

Versluis, M., de Lange, M.J., van Pelt, S.I., Ruivenkamp, C.A., Kroes, W.G., Cao, J., Jager, M.J., Luyten, G.P., and van der Velden, P.A. (2015). Digital PCR validates 8q dosage as prognostic tool in uveal melanoma. PLoS One *10*, e0116371.

Virgili, G., Gatta, G., Ciccolallo, L., Capocaccia, R., Biggeri, A., Crocetti, E., Lutz, J.M., Paci, E., and Group, E.W. (2007). Incidence of uveal melanoma in Europe. Ophthalmology *114*, 2309–2315.

Wang, K., Singh, D., Zeng, Z., Coleman, S.J., Huang, Y., Savich, G.L., He, X., Mieczkowski, P., Grimm, S.A., Perou, C.M., et al. (2010). MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. Nucleic Acids Res. *38*, e178.

Weis, E., Shah, C.P., Lajous, M., Shields, J.A., and Shields, C.L. (2006). The association between host susceptibility factors and uveal melanoma: a meta-analysis. Arch. Ophthalmol. *124*, 54–60.

Wilkerson, M.D., and Hayes, D.N. (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics *26*, 1572–1573.

Woodman, S.E. (2012). Metastatic uveal melanoma: biology and emerging treatments. Cancer J. *18*, 148–152.

Worley, L.A., Long, M.D., Onken, M.D., and Harbour, J.W. (2008). Micro-RNAs associated with metastasis in uveal melanoma identified by multiplexed microarray profiling. Melanoma Res. *18*, 184–190.

Wu, T.D., and Watanabe, C.K. (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics *21*, 1859–1875.

Wu, X., Li, J., Zhu, M., Fletcher, J.A., and Hodi, F.S. (2012). Protein kinase C inhibitor AEB071 targets ocular melanoma harboring GNAQ mutations via effects on the PKC/Erk1/2 and PKC/NF-kappaB pathways. Mol. Cancer Ther. *11*, 1905–1914.

Xi, R., Hadjipanayis, A.G., Luquette, L.J., Kim, T.M., Lee, E., Zhang, J., Johnson, M.D., Muzny, D.M., Wheeler, D.A., Gibbs, R.A., et al. (2011). Copy number variation detection in whole-genome sequencing data using the Bayesian information criterion. Proc. Natl. Acad. Sci. USA *108*, E1128–E1136.

Yang, L., Luquette, L.J., Gehlenborg, N., Xi, R., Haseley, P.S., Hsieh, C.H., Zhang, C., Ren, X., Protopopov, A., Chin, L., et al. (2013). Diverse mechanisms of somatic structural variations in human cancer genomes. Cell *153*, 919–929.

Yavuzyigitoglu, S., Koopmans, A.E., Verdijk, R.M., Vaarwater, J., Eussen, B., van Bodegom, A., Paridaens, D., Kilic, E., de Klein, A., and Rotterdam Ocular Melanoma Study, G. (2016). Uveal melanomas with SF3B1 mutations: a distinct subclass associated with late-onset metastases. Ophthalmology *123*, 1118–1128.

Yu, H., Pak, H., Hammond-Martel, I., Ghram, M., Rodrigue, A., Daou, S., Barbour, H., Corbeil, L., Hebert, J., Drobetsky, E., et al. (2014). Tumor suppressor and deubiquitinase BAP1 promotes DNA double-strand break repair. Proc. Natl. Acad. Sci. USA *111*, 285–290.

Zack, T.I., Schumacher, S.E., Carter, S.L., Cherniack, A.D., Saksena, G., Tabak, B., Lawrence, M.S., Zhsng, C.Z., Wala, J., Mermel, C.H., et al. (2013). Pan-cancer patterns of somatic copy number alteration. Nat. Genet. *45*, 1134–1140.

Zhang, L., Wei, Q., Mao, L., Liu, W., Mills, G.B., and Coombes, K. (2009). Serial dilution curve: a new method for analysis of reverse phase protein array data. Bioinformatics *25*, 650–654.

Zhang, J., Lieu, Y.K., Ali, A.M., Penson, A., Reggio, K.S., Rabadan, R., Raza, A., Mukherjee, S., and Manley, J.L. (2015). Disease-associated mutation in SRSF2 misregulates splicing by altering RNA-binding affinities. Proc. Natl. Acad. Sci. USA *112*, E4726–E4734.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
RPPA antibodies	RPPA Core Facility, MD Anderson Cancer Center; Gonzalez-Angulo et al., 2011	https://www.mdanderson.org/research/research-resources/core-facilities/functional-proteomics-rppa-core.html See Table S3.
Biological Samples		
Primary tumour samples	Multiple tissue source sites, processed through the Biospecimen Core Resource	See Methods: Experimental Model and Subject Details
Critical Commercial Assays		
Genome-Wide Human SNP Array 6.0	ThermoFisher Scientific	Cat: 901153
Infinium HumanMethylation450 BeadChip Kit	Illumina	Cat: WG-314-1002
EZ-96 DNA Methylation Kit	Zymo Research	Cat: D5004
Illumina Barcoded Paired-End Library Preparation Kit	Illumina	https://www.illumina.com/techniques/sequencing/ngs-library-prep.html
TruSeq RNA Library Prep Kit	Illumina	Cat: RS-122-2001
TruSeq PE Cluster Generation Kit	Illumina	Cat: PE-401-3001
Phusion High-Fidelity PCR Master Mix with HF Buffer	New England Biolabs	Cat: M0531L
VECTASTAIN Elite ABC HRP Kit (Peroxidase, Standard)	Vector Lab	Catalog: PK-6100
Deposited Data		
Raw and processed clinical, array and sequence data.	Genomic Data Commons	https://portal.gdc.cancer.gov/legacy-archive
Digital pathology images	Genomic Data Commons Cancer Digital Slide Archive	https://gdc-portal.nci.nih.gov/legacy-archive/ http://cancer.digitalslidearchive.net/
Software and Algorithms		
ABSOLUTE	Carter et al., 2012	http://archive.broadinstitute.org/cancer/cga/absolute
ABySS	Simpson et al., 2009	http://www.bcgsc.ca/platform/bioinfo/software/abyss/
Array-Pro Analyzer	Media Cybernetics, Washington DC	
Atlas-SNP, Atlas2 Suite	Challis et al., 2012	https://sourceforge.net/p/atlas2
BioBloomTools (BBT)	Chu et al., 2013	http://www.bcgsc.ca/platform/bioinfo/software/biobloomtools/
Birdseed	Korn et al., 2008	http://archive.broadinstitute.org/mpg/birdsuite/birdseed.html
BreakDancer	Chen et al., 2009	http://breakdancer.sourceforge.net/
BWA, BWA-backtrack	Li and Durbin, 2010	http://bio-bwa.sourceforge.net/
Circular Binary Segmentation	Olshen et al., 2004	
ClaNC	Dabney, 2006	http://www.stat.tamu.edu/~adabney/clanc/
CoMEt	Leiserson et al., 2015	http://compbio.cs.brown.edu/projects/comet/
ConsensusClusterPlus	Wilkerson and Hayes, 2010	http://bioconductor.org/packages/release/bioc/html/ConsensusClusterPlus.html
ContEst	Cibulskis et al., 2011	http://archive.broadinstitute.org/cancer/cga/contest
Cufflinks	Trapnell et al., 2013	https://cole-trapnell-lab.github.io/cufflinks/
Cytoscape		http://www.cytoscape.org/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
DEXSeq	Anders et al., 2012	http://www.bioconductor.org/packages/release/bioc/html/DEXSeq.html
deFuse	McPherson et al., 2011	http://compbio.bccrc.ca/software/defuse/
EGC.tools (v1.4.11)	NA	https://github.com/uscepigenomecenter/EGC.tools
FACETS	Shen and Seshan, 2016	https://github.com/mskcc/facets
DNA methylation background correction	Triche et al., 2013	
Genome Analysis Toolkit	Van der Auwera et al., 2013	https://software.broadinstitute.org/gatk/
GISTIC, GISTIC2	Mermel et al., 2011	http://archive.broadinstitute.org/cancer/cga/gistic
GMAP	Wu and Watanabe, 2005	http://research-pub.gene.com/gmap/
Heatmap.plus	NA	https://CRAN.R-project.org/package=heatmap.plus
ImageMagick		www.imagemagick.org
In Silico Admixture Removal (ISAR)	Zack et al., 2013	
Integrative Genomics Viewer (IGV)	Thorvaldsdottir et al., 2013	http://software.broadinstitute.org/software/igv/
Maftools	Mayakonda and Koeffler, 2016	https://bioconductor.org/packages/release/bioc/html/maftools.html
MapSplice	Wang et al., 2010	http://www.netlab.uky.edu/p/bioinfo/MapSplice/
MARINa (MAster Regulator INference Algorithm), ssMARINa	Lefebvre et al., 2010 Aytes et al., 2014	http://califano.c2b2.columbia.edu/marina
MatrixEQTL	Shabalin, 2012	https://cran.r-project.org/web/packages/MatrixEQTL
Meerkat	Yang et al., 2013	http://compbio.med.harvard.edu/Meerkat/
methylumi (v2.10.0)	NA	https://www.bioconductor.org/packages/release/bioc/html/methylumi.html
MuTect	Cibulskis et al., 2013	http://archive.broadinstitute.org/cancer/cga/mutect
MuTect2	Van der Auwera et al., 2013	https://software.broadinstitute.org/gatk/documentation/tooldocs/current/org_broadinstitute_gatk_tools_walkers_cancer_m2_MuTect2.php
Mutex	Babur et al., 2015	
MutSig2CV	Lawrence et al., 2014	http://archive.broadinstitute.org/cancer/cga/mutsig
NMF	Gaujoux and Seoighe, 2010	https://cran.r-project.org/web/packages/NMF/
NovoAlign	NA	http://www.novocraft.com/
PARADIGM	Sedgewick et al., 2013	http://sbenz.github.io/Paradigm/
Picard		https://broadinstitute.github.io/picard/
RADIA	Radenbaugh et al., 2014	https://github.com/aradenbaugh/radia/
Strelka	Saunders et al., 2012	https://sites.google.com/site/strelkasomaticvariantcaller/
SnpEff	Cingolani et al., 2012	http://snpeff.sourceforge.net/
PRADA	Torres-Garcia et al., 2014	https://sourceforge.net/projects/prada/
pheatmap	NA	https://cran.r-project.org/web/packages/pheatmap/
rMATS	Shen et al., 2014	http://rnaseq-mats.sourceforge.net/
RSEM	Li and Dewey, 2011	https://deweylab.github.io/RSEM/
Samr	Li and Tibshirani, 2013	https://cran.r-project.org/web/packages/samr
Samtools	Li et al., 2009	http://samtools.sourceforge.net/
SigClust	Huang et al., 2015	https://cran.r-project.org/web/packages/sigclust
STAR	Dobin et al., 2013	https://github.com/alexdobin/STAR
SuperCurve	Ju et al., 2015, Zhang et al., 2009	http://bioinformatics.mdanderson.org/Software/supercurve/
Tran-ABySS	Robertson et al., 2010	http://www.bcgsc.ca/platform/bioinfo/software/trans-abyss
VIPER	Alvarez et al., 2016	http://califano.c2b2.columbia.edu/viper
Ziggurat Deconstruction	Mermel et al., 2011	

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Other		
ChEA database	Lachmann et al., 2010	
Firehose, FireBrowse	The Broad Institute, Cambridge MA	https://gdac.broadinstitute.org/ http://firebrowse.org/
Laurent microarray expression data, GEO: GSE22138	Laurent et al., 2011	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE22138
Literome	Poon et al., 2014	
Multinet	Khurana et al., 2013	
PhosphositePlus	Hornbeck et al., 2014	http://www.phosphosite.org
Regulome Explorer	Institute for Systems Biology, Seattle WA	http://explorer.cancerregulome.org/

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact Scott E. Woodman (swoodman@mdanderson.org).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Tumor and whole blood samples were obtained from patients at contributing centers, with informed consent from their local Institutional Review Boards (IRBs, see below). Biospecimens were processed centrally, and DNA and RNA were distributed to TCGA analysis centers. In total, 80 evaluable primary tumors with associated clinicopathologic data were assayed on at least one molecular-profiling platform.

TCGA Project Management has collected necessary human subjects documentation to ensure the project complies with 45-CFR-46 (the ''Common Rule''). The program has obtained documentation from every contributing clinical site to verify that IRB approval has been obtained to participate in TCGA. Such documented approval may include one or more of the following:

- An IRB-approved protocol with Informed Consent specific to TCGA or a substantially similar program. In the latter case, if the protocol was not TCGA-specific, the clinical site PI provided a further finding from the IRB that the already-approved protocol is sufficient to participate in TCGA.
- A TCGA-specific IRB waiver has been granted.
- A TCGA-specific letter that the IRB considers one of the exemptions in 45-CFR-46 applicable. The two most common exemptions cited were that the research falls under 46.102(f)(2) or 46.101(b)(4). Both exempt requirements for informed consent, because the received data and material do not contain directly identifiable private information.
- A TCGA-specific letter that the IRB does not consider the use of these data and materials to be human subjects research. This was most common for collections in which the donors were deceased.

METHOD DETAILS

BAP1 Terminology

In our cohort, BAP1 mRNA levels were lower in M3 tumors than in D3 tumors. We used the terms ''*BAP1*-aberrant'' and ''*BAP1* aberrancy'' to refer to cases in which we detect *BAP1* sequence alterations (i.e. DNA-seq or RNA-seq variants, which may be short, long, or complex), and/or decreased mRNA expression. We say ''and/or'' because, while *BAP1* alterations in the setting of M3 typically result in decreased BAP1 mRNA expression, we detected no *BAP1* alterations in 7 of 42 M3 tumors in our cohort. It is possible that *BAP1* alterations were present in these cases, but our approaches failed to detect them; alternatively, BAP1 with unaltered sequence may be epigenetically modulated in these cases.

Biospecimen and Clinical Data Processing

Patient and Sample Characteristics

Eight academic medical centers provided primary UM tumor samples, matched blood for germline DNA, and clinicopathologic data from 121 enucleated UM patients under IRB-approved protocols. 80 primary UM from six centers passed all quality-control measures and had data from all molecular analytic platforms except reverse phase protein array (RPPA), for which data were derived from 12 primary UM. Eleven of these 12 cases had *BAP1*, *SF3B1* or *EIF1AX* mutations; V4-A9EH did not, and was removed from further analysis (Table S1). Patients who had been treated prior to tissue procurement with systemic chemotherapy or

focal radiotherapy were excluded. Enrollment criteria required tumors to consist of at least 200 mg of fresh frozen tissue, and DNA-matched normal (blood) controls to be available. A top-slide of the frozen tumor was cut to confirm the pathology characteristics, including adequate cellularity and percentage of necrosis. The presence of adequate amounts and quality of DNA and RNA isolated from the specimens was confirmed, resulting in 80 patients enrolled. The clinical data collected included patient age, sex, race, ethnicity, height, weight, tumor anatomic location (choroid, ciliary body, iris), iris color, tumor clinical dimensions, tumor pathology dimensions, clinical and pathologic AJCC staging, history of prior and synchronous malignancies, new malignancies including development of local and systemic metastases, date of UM treatment, date of diagnosis with metastasis, date of death, cause of death and date of last contact.

Histologic Evaluation of Uveal Melanoma

A panel of five histopathologists with expertise in ocular pathology and melanoma evaluated digital slides via Biopathology Center's Virtual Imaging for Pathology, Education & Research application (VIPER) for the 80 UM. Slides consisted of hematoxylin- and eosin-stained sections from the formalin-fixed paraffin embedded tumors scanned at 200x or 400x magnification. Histomorphologic features evaluated included tumor extent (ciliary body involvement and extrascleral extension), cytologic features including cell morphology (percent spindle and percent epithelioid cells) and degree of pigmentation, and the presence of associated inflammatory components for both tumor-infiltrating lymphocytes and tumor-infiltrating macrophages. Inflammation was characterized as focal vs. diffuse or mild vs. moderate vs. heavy, according to the distribution and density of inflammatory infiltrate within the tumor. The number of mitoses was determined within a 2 mm^2 area, with the mitotic index grouped as low (0-5 mitoses), intermediate (6-10 mitoses), and high (>10 mitoses). Group discussions and slide reviews resulted in consensus determinations for the above features.

Clinical Outcome Analysis

All clinical outcome events were calculated using the time interval in days from the date of the pathologic diagnosis of primary tumor to either the date of documented metastatic disease (UM metastasis), death (UM survival), or last follow-up, censored to 5 years. Kaplan-Meier (KM) analysis for UM-specific death (n=77), UM-specific metastasis (n=70), and UM-specific metastasis or death (n=77) was performed using the survival R package, and log-rank testing was used to compare curves. For genes, sample groups with low vs. high expression were generated by thresholding at the median expression level.

Whole Exome Sequencing (WES)

Library Construction

Libraries were constructed using the protocol described in (Fisher et al., 2011) with several modifications. First, initial genomic DNA input into shearing was reduced from 3 μg to 100 ng in 50 μL of solution. Second, for adapter ligation, Illumina paired end adapters were replaced with palindromic forked adapters with unique 8 base index sequences embedded within the adapter. These index sequences enable pooling of libraries prior to sequencing. Third, custom sample preparation kits from Kapa Biosciences were used for all enzymatic steps of the library construction process.

Following sample preparation, libraries were quantified using PicoGreen, normalized to equal concentration, and pooled by equal volume. Library pools were then quantified using a SYBR Green-based qPCR assay, using PCR primers complementary to the P5/P7 ends of the adapters (kit from Kapa Biosciences). After qPCR quantification, library pools were normalized to 2 nM, denatured using 0.2 N NaOH, and diluted to 20 pM, the working concentration for cluster amplification and sequencing. Denatured library pools were spread across the number of sequencing lanes required to achieve target coverage for all samples.

Sequencing

Cluster amplification and sequencing of denatured templates were performed according to Illumina protocols using HiSeq 2000 v3 cluster amplification kits, v3 flow cells, v3 Sequencing-by-Synthesis kits, Multiplexing Sequencing Primer kits, and the latest version of Illumina's RTA software. 76bp paired end reads, with additional cycles added to read molecular index sequences, were performed on Illumina HiSeq 2000 sequencers.

Alignment and QC

Reads were aligned using BWA-backtrack (Li and Durbin, 2010) to assembly hg19/GRCh37; alignments were processed through the Picard pipeline, which finds and excludes PCR and optical duplicate reads, identifies sites likely harboring strand-specific 8-oxoguanine lesions (Costello et al., 2013), and provides overall library QC metrics to identify problematic samples (none were excluded). Sample contamination levels were estimated using ContEst (Cibulskis et al., 2011), which estimates cross-sample contamination by looking at the distribution of common germline SNP sites. No samples exceeded the maximum contamination threshold of 4%.

Somatic Mutation Calling and Filtering

At the Broad Institute, somatic single nucleotide variants (sSNVs) were identified from tumor-normal paired alignments using MuTect (Cibulskis et al., 2013), which identifies variants unique to the tumor sample by contrasting alignment pileups at each genomic position. Somatic insertions or deletions (sINDELs) were identified using Indelocator, which similarly uses pileups to identify tumor-specific variants. In addition, regions hypothesized to harbor longer sINDEL events (on the order of 50-100 bases, as in *BAP1*) were called using MuTect2 (Van der Auwera et al., 2013). This performs local reassembly according to haplotype structure to better call events that are not trivially associated with pileups, and dramatically reduces the number of false positives due to alignment errors. This resulted in 2,699 sSNV calls and 2,636 sINDEL calls, for 5,335 total calls.

At the Baylor College of Medicine (BCM), mutations in BAM files were detected as follows: Atlas-SNP (Shen et al., 2010) of the Atlas2 Suite (Challis et al., 2012) was run to list all sSNVs. This list was further filtered by removing variant alleles observed in fewer than 4 reads, or present at a variant allele fraction (VAF) of less than 4%. The VAF in the normal had to be less than 1% of the VAF in the

tumor. At least one read had to have a mapping quality of Q20 or better, and the variant had to lie in the central portion of the read. In addition, reads had to support the variant allele in both forward and reverse orientations. COSMIC variants were exempted from the above filters. sINDELs were discovered by similar processing except that the initial list was generated by Atlas-Indel of the Atlas2 Suite (https://sourceforge.net/p/atlas2), and indels must have been observed in at least 10 reads, with a VAF of 15% or more. All variants were compared to a panel of normal genomes and matching variants removed because they were likely germline alleles or recurrent artifacts. Further filtering was done by removing variants with fewer than 2 reads in the normal, tumor VAF 5% or less, or genes with greater than 2 variants for the same sample.

At the University of California Santa Cruz (UCSC), sSNVs were identified by RADIA (RNA AND DNA Integrated Analysis), a method that combines the patient matched normal and tumor DNA whole exome sequencing (WES) data with the tumor RNA-seq data for somatic mutation detection (Radenbaugh et al., 2014). The inclusion of the RNA-seq data in RADIA increases the power to detect somatic mutations, especially at low DNA allelic frequencies. RADIA classifies somatic mutations into 3 categories depending on the read support from the DNA and RNA: 1) DNA calls - mutations with high support in the DNA, 2) RNA Confirmation calls - mutations with high support in both the DNA and RNA, 3) RNA Rescue calls - mutations with high support in the RNA but weak support in the DNA. In the UM cohort, RADIA made 1,955 DNA calls, 399 RNA Confirmation calls, and 59 RNA Rescue calls.

At the BC Cancer Agency's Genome Sciences Centre (BCGSC), Strelka (v1.0.6) (Saunders et al., 2012) was used to identify sSNVs and sINDELs (up to ~22 bp long) from the exome sequencing data for tumors and blood normals. All parameters were set to defaults, with the exception of "isSkipDepthFilters", which was set to 1 in order to skip depth filtration, given the higher coverage in exome datasets. The variants were subsequently annotated using SnpEff (Cingolani et al., 2012), and the COSMIC (v61) (Forbes et al., 2010) and dbSNP (v137) (Smigielski et al., 2000) databases.

Calls generated at the Broad Institute were merged with the calls from BCM, UCSC, and BCGSC. Calls were included in a consensus set if they were called by either the Broad or by two or more of the four participating centers. This resulted in an additional 215 variants not called by the Broad. Consensus calls were filtered through a panel-of-normals, which encodes the distribution of allelic coverage at each genomic position across thousands of normal exomes. This filters out both recurrent sequencing/alignment artifacts and rare germline variants missed during paired tumor-normal calling. By filtering sites exhibiting recurrently high nonreference read counts, we dramatically reduced the number of calls to 2,699, mostly by reducing sINDELs called by Indelocator, which are often false positives due to recurrent alignment artifacts.

Significantly Mutated Genes

This filtered set of calls was analyzed for significantly mutated genes using the MutSig2CV suite (Lawrence et al., 2014). This uses three tests to infer significantly mutated genes: abundance, which classifies whether a gene's observed mutation rate is significantly elevated relative to its expected background mutation rate; clustering, which looks for genes harboring recurrently mutated loci; and conservation, which looks for genes whose mutations are significantly enriched in evolutionarily conserved sites. Each of these tests returns a p-value for every gene, which are Fisher-combined and false discovery rate (FDR)-corrected via Benjamini-Hochberg. Genes were considered "significant" if their FDR value was below 0.1.

Validation Analysis

Calls in significantly mutated genes were subject to Fluidigm validation. Samples were initially aligned with BWA-backtrack, but an inability to properly align reads spanning long deletions led to realigning all samples with NovoAlign (www.novocraft.com), which properly gapped these reads. Mutations were validated by comparing allelic fractions in the whole exome alignments with allelic fractions in the validation alignments; mutations that fell outside of the expected beta-binomial distribution of deviation were rejected. In addition, recurrent sites found in the deep coverage validation data but not present in the lower coverage exomes were added to the final set of calls.

Mutual Exclusivity and Concurrence of Mutations

To generate the alteration matrix, we first ranked all genes based on their MutSig p-value (M), GISTIC p-value (G) and expression verification p-value of copy number changes (E). We aggregated these scores using Gene score = Min(M, Max(G, E)), which uses either a mutation frequency score or a copy number alteration score, whichever is more significant, and tempers G with E when the latter is less significant.

The top 500 genes on this ranked list were selected in an alteration matrix for the Mutex algorithm (Babur et al., 2015), after filtering out genes that had only a single alteration in the cohort. We used gene copy number alterations only if they were also verified with gene expression change compared to other samples. As Mutex parameters, we used 10,000 iterations for first-level randomized runs, and 5 as the maximum exclusive set size. We selected result groups with scores smaller than 0.05. This identified the groups *CYSLTR2, GNAQ, and GNA11; ABR, GNAQ,* and *GNA11*; *SF3B1, BAP1,* and *EIF1AX*; and several genes amplified at 8q: *E2F5, MYLB1, GGH, LYN, LRRCC1, UBE2V2, CEBPD*, and *CSPP1*.

Limiting Mutex to the recurrent Q209 hotspots in *GNAQ* and *GNA11* led to the detection of new groups *PLCB4, GNAQ, GNA11,* and *CYSLTR2*; *DEPDC5, GNAQ, GNA11,* and *CYSLTR2*; *UTRN*, *GNAQ, GNA11,* and *CYSLTR2*; *ABR, GNAQ, GNA11,* and *PLCB4*; and the same prior group of genes amplified at 8q. *PLCB4* and *DEPDC5* were recovered here because they were co-occurrent with the rarer *GNAQ*/*GNA11* hotspot at position 183 in 1 and 2 samples respectively.

The CoMEt algorithm (Leiserson et al., 2015) was used to detected groups of mutually exclusively mutated genes, by running it on the UM mutation list using arguments -t4 -k4 -N1000000 -np 100.

Identifying Mutation Signatures

Using maftools 0.99.34 (Mayakonda and Koeffler, 2016) and NMF 0.20.6, somatic nucleotide substitutions across the cohort and their trinucleotide sequence contexts, were decomposed into three distinct mutation signatures, that each correlate to three validated signatures (15, 19 and 1A) (Alexandrov et al., 2013). Correlations to the validated signatures were weak (r = 0.61, 0.57, and 0.76), and none of these three signatures is described as UV-mediated.

SNP-based Copy Number Analysis

DNA from each of the 80 tumor and 80 normal samples were hybridized to Affymetrix SNP 6.0 arrays using protocols at the Genome Analysis Platform of the Broad Institute, as previously described (McCarroll et al., 2008). Briefly, from raw CEL files, Birdseed was used to infer a preliminary copy number at each probe locus (Korn et al., 2008). For each tumor, genome-wide copy number estimates were refined using tangent normalization, in which tumor signal intensities are divided by signal intensities from the linear combination of all normal samples that are most similar to the tumor. This linear combination of normal samples tends to match the noise profile of the tumor better than any set of individual normal samples, thereby reducing the contribution of noise to the final copy number profile. Individual copy number estimates then were segmented using Circular Binary Segmentation (Olshen et al., 2004). As part of this process of copy number assessment and segmentation, regions corresponding to germline copy number alterations were removed by applying filters generated from either the 80 UM blood normals, or the larger cohort of blood normals in the TCGA ovarian cancer analysis. Segmented copy number profiles for tumor and matched control DNAs were analyzed using Ziggurat Deconstruction, an algorithm that parsimoniously assigns a length and amplitude to the set of inferred copy number changes underlying each segmented copy number profile (Mermel et al., 2011). Allelic copy number, whole genome doubling, subclonality, and purity and ploidy estimates were calculated using the ABSOLUTE and FACET algorithms (Carter et al., 2012; Shen and Seshan, 2016). For samples with ABSOLUTE-corrected copy number, CBS-derived segmented copy number values were re-centered using the In Silico Admixture Removal (ISAR) procedure (Zack et al., 2013). Significant focal copy number alterations were identified from ISAR-corrected segmented data using GISTIC 2.0.225. For copy number based clustering, tumors were clustered based on thresholded copy number at reoccurring alteration peaks from GISTIC analysis (all_lesions.conf_99.txt file). Clustering was done in R based on Manhattan distance using Ward's method. Isochromosome status (e.g. for isochromosome 8q) was inferred from allelic copy number profiles from the ABSOLUTE algorithm. Specifically, for any metacentric chromosome, a potential isochromosome was reported if the modal integer copy number of the major allele for one arm (e.g. q, or long arm) was at least two greater than the modal integer copy number of the minor allele of the opposite arm (e.g. p, or short arm).

RNA Sequencing

RNA Library Construction, Sequencing, and Analysis

One μg of total RNA was converted to mRNA libraries using the Illumina mRNA TruSeq kit (RS-122-2001 or RS-122-2002) following the manufacturer's directions. Libraries were sequenced 48x7x48bp on the Illumina HiSeq 2000 as previously described (Cancer Genome Atlas Research Network, 2012). FASTQ files were generated by CASAVA. RNA reads were aligned to the hg19 genome assembly using MapSplice 0.7.4 (Wang et al., 2010). Gene expression was quantified for the transcript models corresponding to the TCGA GAF 2.1, using RSEM (Li and Dewey, 2011), and were normalized within each sample to a fixed upper quartile. For further details on this processing, refer to Description file at the NCI GDC data portal under the V2_MapSpliceRSEM workflow (https://gdc-portal.nci.nih.gov/). Quantification of genes, transcripts, exons and junctions can also be found at the GDC Data Portal.

Unsupervised Clustering

For clustering, a set of 1,981 genes that were both highly expressed and had highly variable expression values were identified. The 0.75 quantile of mean(RSEM) values was used as a threshold for highly expressed genes, while the 0.9 quantile of variance(RSEM) values was used as a threshold to identify genes with highly variable expression values. After median centering the log10(RSEM+1) values by gene, consensus clustering was applied using the ConsensusClusterPlus R package (Wilkerson and Hayes, 2010) with partitioning around medoids (PAM), a Spearman correlation-based distance, and 10,000 subsamples with a 0.85 random gene fraction. Output from ConsensusClusterPlus along with gene expression heatmaps, principal component analysis, and silhouette plots suggested four expression subtypes: cluster 1 (n = 22), cluster 2 (n = 21), cluster 3 (n = 15), and cluster 4 (n = 22) (Figure 3A). ClaNC (Dabney, 2006) was used to identify genes whose expression patterns characterized the subtypes. The statistical significance of the differences in gene expression patterns present in the subtypes was assessed with the SigClust R package (Huang et al., 2015) using 1,000 permutations, the default covariance estimation method, and the 1,981 clustering genes.

Differential Expression Analysis

The samr R package (Li and Tibshirani, 2013) was used to identify genes that were differentially expressed in the RNA subtypes using 1,000 permutations and a q-value threshold of 0.05. We then used the DAVID annotation database (Huang da et al., 2009) to identify pathways that were enriched for differentially expressed genes.

Structural Rearrangements, Emphasizing BAP1

To identify structural rearrangements, including longer indels, we assembled the 48-bp paired-end read RNA-seq data for each sample using the *de novo* assembler ABySS v.1.3.4 (Simpson et al., 2009), and analyzed the resulting assembly with Tran-ABySS v.1.4.8 (Robertson et al., 2010). To address how variations in transcript abundance influence assembly, for each library we generated sets of assemblies using every second k-mer length between 24 and 48 bp, then generated a working contig set by merging the contigs from all of the library's k-mer assemblies. Each merged assembly was used as input into Trans-ABySS, which identifies indels and

alternative splicing events by using GMAP (Wu and Watanabe, 2005) to compare the *de novo* contigs to the human reference genome and to multiple sets of transcript models. Structural contig variant events that do not match the reference but fulfill specific alignment and filtering criteria are reported in the analysis results. Events identified in *BAP1* were manually reviewed in RNA-seq data in the Integrative Genomics Viewer (IGV) (Thorvaldsdottir et al., 2013) and were compared in IGV with the sample's exome data to help verify Trans-ABySS rearrangement calls. M3 samples that did not have *BAP1* rearrangements called by the above RNA-seq analysis were individually reviewed in IGV to identify potential events. In VD-AA8O, when a homozygous deletion within *BAP1* was reported from DNA data but not by our analysis of RNA-seq data, we used deFuse (McPherson et al., 2011) on the RNA-seq data to confirm the deletion.

Gene Fusion Detection

RNA-seq data supports detecting structural variants, including alternate splicing, intra-chromosomal fusions, and inter-chromosomal fusions. We used two algorithms to identify gene fusions: MapSplice (Wang et al., 2010) and PRADA (Torres-Garcia et al., 2014). PRADA uses BWA (Li and Durbin, 2010) to extract all best alignments per read from a dual (genome and transcriptome) reference file. After this initial mapping, the alignment coordinates of reads that mapped to the transcriptome are transformed into coordinates on the genome. Mapped reads whose best alignments have multiple genomic coordinates are removed. Quality scores are recalibrated using the Genome Analysis Toolkit (Van der Auwera et al., 2013). Index files are generated using Samtools (Li et al., 2009) and duplicate reads are flagged using Picard. The PRADA fusion module detects fusion transcripts by identifying discordant read pairs and junction-spanning reads. Discordant read pairs are paired read ends that map uniquely to different protein-coding genes with orientations consistent with formation of a sense–sense chimera. Junction-spanning reads are detected by constructing a sequence database that holds all possible exon-exon junctions that match the 3′ end of one gene fused to 5′ end of a second gene. Unmapped reads aligned to the database of all hypothetical exon junctions created by using the Ensembl transcriptome reference. Only reads for which the mate pair maps to either of the two fusion partner genes are considered as fusion transcripts. In this study, we extracted fusions with (1) at least two discordant read pairs, (2) at least one junction spanning read and (3) without high gene homology between each fusion gene partner (BLAST E-value > 0.001). Next, we applied the concept of mutation allele fraction to RNA sequencing data, and calculated the ratio of junction-spanning reads to the total number of reads crossing over the junction point in the reference transcript. We used the transcript allele fraction (TAF) to exclude artifacts that depend on highly expressed transcripts. We included fusion transcripts showing TAF > 0.01 of both genes in our fusion list. In addition, we filtered out fusions that are found in normal TCGA samples.

Briefly, MapSplice_2_0_beta_7_21 identifies fusion candidates as any two segments of a read alignment that were (1) separated by a gap longer than 200,000 nt, or (2) were on different chromosomes, or (3) were on different strands, or (4) mapped to discordant locations (i.e. the apparent direction of transcription changes between the segments). To decrease false positives, these candidates were further filtered by manual review and visually examining predicted fusion events of special interest utilizing a novel realignment and visualization utility. For each predicted fusion, this visualization tool generates a contiguous synthetic genomic reference sequence across the fusion junction. This region includes the sequence from both the donor and acceptor sides of a putative fused transcript, plus flanking genome sequence immediately adjacent to the predicted genomic fusion loci. An attempt is then made to (re) align all reads from the RNA-seq experiment that predicted the fusion, to the synthetic fusion reference sequences. All the reads that map to one of the synthetic fusion loci (including flanking regions) are collected into one BAM file, those reads that support the fusion are also copied into a second more exclusive BAM file. This second file contains only reads directly supporting the fusion junction, either by spanning it or comprising a mate pair that bridges the junction even though neither read spans it. These BAM files together with the synthetic fusion sequences can be loaded into IGV (Thorvaldsdottir et al., 2013) for the purposes of visualizing the predicted fusion events as well as its read alignments. Visualizing predicted fusions in this way provides an opportunity for the application of human pattern recognition skills to the task of filtering fusions through direct qualitative inspection of the predicted variant and its bridging and spanning supporting reads, within the context of its surrounding genomic sequence and transcript models.

Splicing Factor Mutants

All tumor RNA-seq data was realigned using STAR 2.4.1d (Dobin et al., 2013) in multi-sample two-pass mode, removing splice junctions covered by less than 10 unique reads across all samples. After realignment, splice junctions for which neither splice site was present in Gencode v19 and those connecting two genes were removed. Differentially-used splice junctions were identified with DEXSeq 1.17.6 (Anders et al., 2012) using all samples.

Splicing defects associated with mutations in splicing factors *SRSF2* or *SF3B1* were identified with rMATS 3.0.9 (Shen et al., 2014). Two samples with in-frame deletions in the *SRSF2* linker sequence between the functional RRM and RS domains were compared with five randomly chosen control samples that had no somatic mutations in spliceosomal genes. Eighteen samples with *SF3B1* missense mutations in HEAT domains were compared to 20 control samples. To increase sensitivity to novel splice junctions in the *SF3B1* comparison, a custom annotation was created from mutant and control samples with Cufflinks 2.2.1 using default parameters (Trapnell et al., 2013).

We used Sashimi plots (Katz et al., 2015) to visualize splicing changes in RNA-seq data, across sets of samples that had, or lacked, particular mutations.

Effect of Immune Marker Genes on mRNA Consensus Clustering

We used a subtraction approach to determine the effect of immune marker genes on the mRNA four-cluster solution. We removed the expression data from 513 genes that define different immune cell types (Newman et al., 2015) from the RSEM data and repeated the clustering analysis using the same parameters that were applied in the original analysis. Manual review of PCA plots, gene expression

heatmaps, and silhouette plots strongly suggested a stable four-class solution that was highly concordant with the original four-class solution. Only one sample had different class labels after subtraction, and the sample that changed clusters was a strongly atypical cluster member in its initial cluster, based on silhouette widths calculated from the consensus memberships.

Non-Coding RNA Sequencing

RNA-seq Read Mapping for lncRNAs

RNA-seq reads were aligned to the human reference genome (GRCh38/hg38) and transcriptome (Ensembl v82) using STAR 2.4.2a (Dobin et al., 2013). STAR was run with the following parameters: minimum / maximum intron sizes were set to 30 and 500,000, respectively; noncanonical, unannotated junctions were removed; maximum tolerated mismatches was set to 10; and the outSAMstrandField intron motif option was enabled. The Cuffquant command included with Cufflinks 2.2.1 (Trapnell et al., 2013) was used to quantify the read abundances per sample, with fragment bias correction and multiread correction enabled, and all other options set to default. To calculate normalized abundance as fragments per kilobase of exon per million fragments mapped (FPKM), the Cuffnorm command was used with default parameters. From the FPKM matrix for the 80 tumor samples, we extracted 8,167 genes with "lincRNA" and "processed_transcript" Ensembl biotypes.

miRNA Sequencing

We generated miRNA sequencing data using methods described previously (Chu et al., 2016), except that 1ug of total RNA (at 250ng/uL) was used as input instead of messenger RNA-depleted RNA. Briefly, we aligned reads to the GRCh37/hg19 reference human genome, assigned read count abundances to miRBase v16 stem-loops and mature strands, and assigned miRBase v20 5p and 3p mature strand names to MIMAT accession IDs. While we used only reads with exact-match alignments in calculating miRNA abundances, BAM files available from the Genomics Data Commons (https://gdc.cancer.gov/) include all sequence reads.

Unsupervised Clustering

We extracted 356 lncRNAs that were robustly expressed (mean FPKM $\geq$ 1) and highly variable across the n = 80 tumor cohort ($\geq$ 95th FPKM variance percentile) from the matrix of 8,167 lncRNAs (above). Groups of samples with similar abundance profiles were identified by unsupervised consensus clustering with ConsensusClusterPlus (CCP) 1.20.0. Calculations were performed using Spearman correlations, partitioning around medoids (PAM) and 10,000 iterations. From solutions with 2, 3, 4 and 5 clusters we selected a four-cluster solution after assessing consensus membership heatmaps and dendrograms, CCP clustering metrics, KM plots, and clustering results from other platforms. To visualize typical vs. atypical cluster members, we used the R cluster package to calculate a profile of silhouette widths (W_{cm}) from the consensus membership matrix. To generate a heatmap we used a SAM (Li and Tibshirani, 2013) (samr v2.0) multiclass analysis with an FPKM input matrix and an FDR threshold of 0.05 to identify lncRNAs whose abundance varied across the clusters. For lncRNAs with larger SAM scores, a q-value of $\leq$ 0.01, and a mean FPKM $\geq$ 5, we set the columns of the FPKM data matrix to the heatmap order, transformed each row of the matrix by $\log_{10}$(FPKM + 1), then used the pheatmap R package (v1.0.2) to scale and cluster only the rows, using a Pearson distance metric and Ward clustering.

For miRNA sequencing data we used unsupervised non-negative matrix factorization (NMF) consensus clustering (v0.20.5) in R 3.1.2, with default settings (Gaujoux and Seoighe, 2010). The input was a reads-per-million (RPM) data matrix for the 303 (25% of 1212 miRBase v16) most-variant 5p or 3p mature strands. After running a rank survey for between 2 and 15 clusters with 50 iterations per solution, we identified a clustering solution for more detailed work by assessing profiles of the cophenetic correlation coefficient and the average W_{cm} (see below), KM plots, and clinical covariate associations, then performed a 500-iteration run to generate the final clustering result. To visualize typical vs. atypical cluster members, we used the R cluster package to calculate a W_{cm} from the final NMF consensus membership matrix. To generate a clustering heatmap we used a SAM (Li and Tibshirani, 2013) (samr v2.0) multiclass analysis with an RPM input matrix and an FDR threshold of 0.05 to identify mature strands whose abundance varied across the clusters. For mature strands with larger SAM scores and a mean RPM $\geq$ 25, we set the columns of the RPM data matrix to the heatmap order, transformed each row of the matrix by $\log_{10}$(RPKM + 1), then used the pheatmap R package (v0.7.7 or v1.0.2) to scale and cluster only the rows, using a Pearson distance metric and Ward clustering. The RPM filtering acknowledged that more abundant miRNAs are more likely to be influential (Mullokandov et al., 2012).

Differentially Abundant mRNAs, lncRNAs and miRNAs

We identified mRNAs, lncRNAs and miRNAs that were differentially abundant unsupervised clusters using unpaired two-class SAM analyses (samr v2.0), with an RSEM, FPKM and RPM input matrix and an FDR threshold of 0.05. For miRNA figures we retained miRNAs with a mean RPM > 50 in at least one of the two groups being compared. Unfiltered results are available in supplemental files (https://tcga-data.nci.nih.gov/docs/publications/uvm_2016).

LncRNAs/miRNAs Influenced by Copy Number

To determine lncRNAs whose abundance was influenced by somatic copy number alterations (SCNA), we used MatrixEQTL v2.1.1 (Shabalin, 2012) to calculate Spearman correlations (FDR < 0.05) between a) FPKM for the 713 noncoding genes that had an FPKM of at least 1.0 in at least 10 of the tumor samples and b) Gencode v20-based GISTIC2 real-valued (i.e. unthresholded) 'all_data_by_gene' SCNA. We used IGV v2.3.60 with 'seg' data to generate a global heatmap of SCNA with samples ordered by the four-cluster unsupervised clustering solution, and to generate whole-chromosome graphics of SCNA at a gene, sorting the heatmap by copy number amplification at the gene.

Similarly, for miRNAs, we used MatrixEQTL v2.1.1 to calculate FDR-thresholded Spearman correlations between a) normalized (RPM) abundance for the subset of pre-miRNAs (i.e. stem-loops) that had an RPM of at least 1.0 in at least 10 of the tumor samples, and b) GISTIC2 'all_data_by_gene' SCNA data.

Covariates Associated with Unsupervised Clusters

We compared unsupervised clusters to clinical and molecular covariates by calculating contingency table association p-values using R, with a Chi-square or Fisher exact test for categorical data, and a Kruskal-Wallis test for real-valued data.

Differential miRNA Targeting

To identify potential differential miRNA-mRNA targeting effects between miRNA clusters 3 and 4, we used SAM 2-class unpaired analyses (Li and Tibshirani, 2013) to identify gene-level mRNAs and miRNAs that were differentially abundant between these clusters (FDR < 0.05). From these, we then identified miRNA-mRNA pairs that were inversely differential between the clusters and had functional validation publications (using evidence types like luciferase reporter, qPCR, and Western blots) that indicated direct miRNA targeting, as reported by miRTarBase v6.0 (Chou et al., 2016). We displayed the resulting network with Cytoscape 3.4.0, coloring nodes to reflect positive and negative fold changes between the two miRNA-based clusters. Boxplots were generated in R using default settings. Each box spans the 25th to 75th percentile range in the data, i.e. the interquartile range (IQR), and shows a line at the median value. Whiskers extend 1.5 times the IQR from the box extent.

Testing the Influence of Gender on miRNA Subtyping

Given the strong differential expression of Xq27.3 miRNAs between miRNA subtypes, we assessed whether gender may have influenced the miRNA subtypes. Patient gender was statistically unassociated with subtypes on all molecular platforms. While some of the ~300 miRNAs used for unsupervised clustering were differentially abundant between genders in our cohort, most of the 20 that had the largest gender-based fold changes localized to non-sex chromosomes, and none were in Xq27.3.

DNA Methylation

Sample Preparation and Hybridization

The Illumina Infinium HM450 array (Bibikova et al., 2011) was used with standard protocols. Briefly, genomic DNA (1,000 ng) for each sample was treated with sodium bisulfite, recovered using the Zymo EZ DNA methylation kit (Zymo Research, Irvine, CA) according to the manufacturer's specifications and eluted in 18 ul volume. After passing quality control, bisulfite-converted DNA samples were whole-genome amplified followed by enzymatic fragmentation and hybridized overnight to BeadChips followed by a locus-specific base extension with labeled nucleotides (cy3 and cy5). BeadArrays are scanned and the raw data are imported into custom R programs for pre-processing and calculation of DNA methylation beta value for each probe and sample. Quality control and probe exclusions were done using standard protocols, as previously described (Cancer Genome Atlas Research Network, 2014b).

Analytical Methods

We carried out unsupervised consensus clustering on the most variable 1% of CpG probes (3,859 of 385,857 probes), using the ConsensusClusterPlus (Wilkerson and Hayes, 2010) R package, with Euclidean distance and PAM. Solutions with between 2 and 7 clusters were evaluated for cluster stability, and for associations with clinical and molecular covariates.

To identify epigenetically silenced genes, we applied a method previously described (Cancer Genome Atlas Research Network, 2014c). Specifically, we first identified promoter CpG sites that meet several criteria: (a) at least 90% of normal samples should be clearly unmethylated ($\beta \leq 0.1$) at that site, (b) at least 5% of tumor samples should be clearly methylated ($\beta \geq 0.3$) and (c) a t-test comparing expression levels in methylated ($\beta \geq 0.3$) and unmethylated tumor samples ($\beta < 0.1$) should be significant at an FDR < 0.01. A gene was defined as epigenetically silenced if at least 25% of the promotor CpG sites met all of these criteria. A total of 120 adjacent normal tissue samples were used for this analysis, including 10 drawn at random from each of the 12 TCGA projects that include such normal samples: lung adenocarcinoma, breast invasive carcinoma, colon adenocarcinoma, endometrial carcinoma, and others (https://tcga-data.nci.nih.gov/docs/publications/).

We estimated leukocyte fraction using an approach described in (Carter et al., 2012). As a source of leukocyte DNA methylation levels, we used data for peripheral blood mononuclear cells (PBMC) from six healthy donors (Reinius et al., 2012) (GEO: GSE35069).

We identified 36 mRNAs, 65 lncRNAs and 94 miRNAs that were statistically associated with local DNA methylation. We required an 'epigenetically-controlled pattern', which consisted of a) BH-corrected p-values less than 0.05 for a Spearman correlation of miRNA/lncRNA abundance to beta for probes in promoter regions associated with the miRNAs (Marsico et al., 2013) and lncRNAs, and b) BH-corrected p-values less than 0.01 for a t-test of RPM between unmethylated ($\beta < 0.1$) and methylated ($\beta > 0.3$) samples.

Fisher's exact test was used to test for associations of DNA methylation clusters with clusters for SCNA, mRNA, lncRNA and miRNA, as well as with significantly mutation genes.

The analyses described above were done with R, using standard methods and custom scripts.

Low-Pass Whole Genome Sequencing

Library Construction

Approximately 500-700 ng of genomic DNA from fifty randomly selected tumor and matched normal pair samples were individually sheared into fragments of approximately 300 bp using an E220 Focused-ultrasonicator (Covaris). These fragments were made into paired-end libraries using KAPA Bios kits in a Sciclone NGS Workstation (Caliper/Perkin Elmer) according to manufacturers' protocols. Libraries were sequenced using an Illumina HiSeq 2000, one sample per lane, with a paired-end 2 x 51 bp setup. The average depth of coverage was approximately 4.9X, with a minimum of 1.56X and maximum at 8.17X. The average genome coverage was 89.05%, with a minimum of 71.87% and maximum of 92.12%. Raw data was converted to FASTA format, and the Burrows-Wheeler Aligner used to generate BAM files.

Structural Rearrangements Detected using BreakDancer and Meerkat

BreakDancer (Chen et al., 2009) and Meerkat (Yang et al., 2013) algorithms were used to detect structural variations. BreakDancer configuration files were created for each tumor/normal pair from BAM files using bam2cfg.pl. Insertions, deletions, inversions, inter and intra chromosomal translocations were predicted on the basis of read pairs with unexpected separation distances or orientations. The variants between tumor and normal configuration files were filtered to remove germline alterations. Data was then re-examined using the Meerkat algorithm, which required identifying at least two discordant read pairs, with one read covering the actual breakpoint junction. Variants from tumor genomes were filtered by those in normal genomes, and germline events were removed. Alterations found in simple or satellite repeats were also excluded from the output. The final Meerkat calls met one of two criteria: (i) the read identified to span the breakpoint junction hit the predicted breakpoint region uniquely, according to a BLAT (BLAST-like alignment tool) search, or (ii) the mate of the read spanning the breakpoint junction was mapped near the predicted breakpoint. BIC-seq was used to determine copy number alterations in the tumor genomes (Xi et al., 2011).

Exon Expression Graphs

RNA-seq–derived exon expression levels for genes with somatic structural alterations were visualized. The input file "UVM.rnaseqv2__illuminahiseq_rnaseqv2__unc_edu__Level_3__exon_quantification__data.data.txt" was obtained from Broad GDAC Firehose (2016_01_28 stddata Run, https://confluence.broadinstitute.org/display/GDAC/Dashboard-Stddata). Normal expression levels were quantified with "TCGA.hg19.June2011.gaf" (https://gdc.cancer.gov/about-data/data-harmonization-and-generation/gdc-reference-files). A standard Z-score was calculated for each exon of each gene on either side of a fusion by mean-centering the log2-transformed RPKM (reads per kilobase of transcript per million reads mapped) values and dividing by the standard deviation, visualizing high (red) and low (blue) relative to the tumor cohort average. Exons that had expression levels below one RPKM across 70% of the patient samples, were flagged as not expressed (gray). Exon expression graphs were built stepwise, initially taking the fusion coordinates and the reference genome to create an "exon/start/stop" table that was used to parse the RNA-seq input file. After verification and error checking, a final file was loaded in to R where the graphs were assembled. ImageMagick 6.9.1 (www.imagemagick.org) was used to visualize the results.

Reverse Phase Protein Arrays (RPPA)

RPPA Experiments and Data Processing

Protein was extracted using RPPA lysis buffer (1% Triton X-100, 50 mmol/L Hepes (pH 7.4), 150 mmol/L NaCl, 1.5 mmol/L MgCl2, 1 mmol/L EGTA, 100 mmol/L NaF, 10 mmol/L NaPPi, 10% glycerol, 1 mmol/L phenylmethylsulfonyl fluoride, 1 mmol/L Na3VO4, and aprotinin 10 ug/mL) from human tumors and RPPA was performed as described previously, using the SuperCurve v1.4.1 R package (Hu et al., 2007; Ju et al., 2015; Zhang et al., 2009). Lysis buffer was used to lyse frozen tumors by Precellys homogenization. Tumor lysates were adjusted to 1 μg/μL concentration as assessed by bicinchoninic acid assay (BCA) and boiled with 1% SDS. Tumor lysates were manually serial diluted in two-fold of 5 dilutions with lysis buffer. An Aushon Biosystems 2470 arrayer (Burlington, MA) printed 1,056 samples on nitrocellulose-coated slides (Grace Bio-Labs). Slides were probed with 220 validated primary antibodies (Table S3) followed by corresponding secondary antibodies. Signal was captured using a DakoCytomation-catalyzed system and DAB colorimetric reaction. Slides were scanned in a CanoScan 9000F. Spot intensities were analyzed and quantified using ArrayPro Analyzer (Media Cybernetics, Washington DC) to generate spot signal intensities (Level 1 data). The software SuperCurveGUI (Hu et al., 2007), available at http://bioinformatics.mdanderson.org/Software/supercurve, was used to estimate the EC50 values of the proteins in each dilution series (in log2 scale). Briefly, a fitted curve ("supercurve") was plotted with the signal intensities on the Y-axis and the relative log2 concentration of each protein on the X-axis using the non-parametric, monotone increasing B-spline model (Tibes et al., 2006). During the process, the raw spot intensity data were adjusted to correct spatial bias before model fitting. A QC score was calculated for each slide to help determine the quality of the slide: if the score was less than 0.8 on a 0-to-1 scale, the slide was dropped. In most cases, the staining was repeated to obtain a high quality score. If more than one slide was stained for an antibody, the slide with the highest QC score was used for analysis (Level 2 data). Protein measurements were corrected for loading as described (Gonzalez-Angulo et al., 2011; Hu et al., 2007) using median centering across antibodies (level 3 data, described later). In total, 220 antibodies and 12 UM samples were processed on the RPPA platform. Final selection of antibodies was also driven by the availability of high quality antibodies that consistently pass a strict validation process. These antibodies are assessed for specificity, quantification and sensitivity (dynamic range) in their application for protein extracts from cultured cells or tumor tissue. Antibodies are labeled as validated and use with caution based on degree of validation (Gonzalez-Angulo et al., 2011).

RPPAs were quantitated and processed (including normalization and load controlling) as described previously, using ArrayPro Analyzer software (Media Cybernetics, Washington DC) and SuperCurve v1.3, available at http://bioinformatics.mdanderson.org/OOMPA. Raw data (level 1), SuperCurve nonparameteric model fitting on a single array (level 2), and loading-corrected data (level 3) (Ju et al., 2015; Zhang et al., 2009) were deposited at the TCGA Data Coordinating Center (DCC).

Data Normalization

We performed median centering across all the antibodies for each sample to correct for sample loading differences. Those differences arise because protein concentrations are not uniformly distributed per unit volume. By observing the expression levels across

many different proteins in a sample, we can estimate differences in the total amount of protein in that sample vs. other samples. Subtracting the median protein expression level forces the median value to become zero, allowing us to compare protein expressions across samples. Those median-centered "level 3" RPPA data have been uploaded to the TCGA portal.

Antibodies that Were Differentially Abundant between the D3 and M3 Samples

Of the 12 samples that had RPPA data available, 6 were D3/*SF3B1* mutants and 5 were M3/*BAP1*-aberrant in a mutually exclusive manner. V4-A9EH was a D3 sample with no aberrations in either of the above genes, and was not analyzed further. We identified antibodies that were differentially abundant between the retained D3 and M3 samples using a Wilcoxon test in R v3.3.3, applying a Benjamini-Hochberg correction for multiple testing to the p values. Boxplots were generated in R using default settings. Each box spans the 25th to 75th percentile range in the data, i.e. the interquartile range (IQR), and shows a line at the median value. Whiskers extend 1.5 times the IQR from the box extent.

Antibody-Based Pathway Scores

Pathway scores were calculated with the method described in (Akbani et al., 2014).

Microbial Detection

The microbial detection pipeline is based on BioBloomTools (BBT, v1.2.4.b1), which is a Bloom filter-based method for rapidly classifying RNA-seq or DNA-seq read sequences (Chu et al., 2013). We generated 43 filters from 'complete' NCBI genome reference sequences of bacteria, viruses, fungi and protozoa, using 25-bp k-mers and a false positive rate of 0.02. We ran BBT in paired-end mode with a sliding window to screen FASTQ files from 80 tumor RNA-seq libraries (48-bp PE reads), 160 whole exome libraries (80 tumor and 80 blood normal libraries with 76-bp PE reads) and 102 whole genome libraries (51 tumor and 51 blood normal libraries with 51-bp PE reads). In a single-pass scan for each library, BBT categorized each read pair as matching the human filter, matching a unique microbial filter, matching more than one filter (multi-match), or matching neither human nor microbe (no-match). For each filter, we then calculated a RPM abundance metric as:

$$RPM = \frac{reads\ mapped\ to\ a\ microbe}{reads\ mapped\ to\ human} \cdot 10^6$$

Given the BBT read screening results, we elected not to test for viral genomic integration, using methods previously described (Cancer Genome Atlas Research Network, 2014a).

Regulome Explorer

To gain greater insight into the development and progression of UM, we have integrated all of the data types produced by TCGA and described in this paper into a single "feature matrix". From this single heterogeneous dataset, significant pairwise associations have been inferred using statistical analysis and can be visually explored in a genomic context using Regulome Explorer, an interactive web application (http://explorer.cancerregulome.org).

In addition to associations that are inferred directly from the TCGA data, additional sources of information and tools are integrated into the visualization for more extensive exploration (e.g., NCBI Gene, miRBase, the UCSC Genome Browser, etc).

Feature Matrix Construction

A feature matrix was constructed using all available clinical, sample, and molecular data for 80 unique patient/tumor samples. The clinical information includes features such as age and gender; while the sample information includes features derived from molecular data such as single-platform cluster assignments. The molecular data includes mRNA and miRNA expression levels (Illumina HiSeq data), protein levels (RPPA data), SCNA (derived from segmented Affymetrix SNP data as well as GISTIC regions of interest and arm-level values), DNA methylation levels (Illumina Infinium Methylation 450k array), and somatic mutations. For mRNA expression data, gene level RSEM values from RNA-seq were log2 transformed, and filtered to remove low-variability genes (bottom 25% removed, based on interdecile range). For miRNA expression data, the summed and normalized miRNA quantification files were log2 transformed, and filtered to remove low-variability miRNAs (bottom 25% removed, based on interdecile range). For methylation data, probes were filtered to remove the bottom 25% based on interdecile range. For somatic mutations, several binary mutation features indicating the presence or absence of a mutation in each sample were generated. Mutation types considered were synonymous, missense, nonsense and frameshift. Protein domains (InterPro) including any of these mutation types were annotated as such, with nonsense and frameshift annotations being propagated to all subsequent protein domains.

Pairwise Statistical Significance

Statistical association among the diverse data types in this study was evaluated by comparing pairs of features in the feature matrix. Hypothesis testing was performed by testing against null models for absence of association, yielding a p-value. P-values for the association between and among clinical and molecular data types were computed according to the nature of the data levels for each pair: categorical vs. categorical (Chi-square test or Fisher's exact test in the case of a 2 x 2 table); categorical vs. continuous (Kruskal-Wallis test) or continuous vs. continuous (probability of a given Spearman correlation value). Ranked data values were used in each case. To account for multiple-testing bias, p-values were adjusted using the Bonferroni correction.

Exploring Significant Associations Between Features

Regulome Explorer allows the user to interactively explore significant associations between various types of features: associations between molecular features (e.g. miRNA expression and gene expression), associations between molecular features and derived numeric features (e.g. purity scores), and associations between molecular features and categorical features such as clinical features or clusters derived from prior analysis (e.g. mRNA clusters).

cBioPortal Visualization

sSNV, sINDEL, SCNA, and mRNA expression data was imported into cBioPortal at Memorial Sloan Kettering Cancer Center, and made available for explorative analyses at http://www.cbioportal.org/study?id=uvm_tcga.

PARADIGM Integrated Pathway Analysis

Integrated Pathway Levels (IPLs)

mRNA expression, SCNA, and pathway interaction data for 80 UM samples were integrated using the PARADIGM software (Sedgewick et al., 2013). Briefly, this procedure infers integrated pathway levels (IPLs) for genes, complexes, and processes, using pathway interactions, and genomic and functional genomic data from each patient sample.

Normalized gene-level RSEM RNA-seq expression data and thresholded SCNA data (GISTIC2 all_thresholded.by_genes.txt) were obtained from Firehose. One was added to all expression values, which were then log2 transformed and median-centered across samples for each gene. The log2 transformed, median-centered mRNA data were rank-transformed based on the global ranking across all samples and all genes and discretized (+1 for values with ranks in the highest tertile, -1 for values with ranks in the lowest tertile, and 0 otherwise) prior to PARADIGM analysis.

Pathways were obtained in BioPax Level 3 format, and included the NCIPID and BioCarta databases from http://pid.nci.nih.gov and the Reactome database from http://reactome.org. Gene identifiers were unified by UniProt ID then converted to Human Genome Nomenclature Committee's HUGO symbols using mappings provided by HGNC (http://www.genenames.org). Altogether, 1,524 pathways were obtained. Interactions from all of these sources were then combined into a merged Superimposed Pathway (SuperPathway). Genes, complexes, and abstract processes (e.g. "cell cycle" and "apoptosis") were retained and are henceforth referred to collectively as pathway "features". The resulting pathway structure contained a total of 19,504 features, representing 7,369 proteins, 9,354 complexes, 2,092 families, 82 RNAs, 15 miRNAs and 592 abstract processes.

The PARADIGM algorithm infers an IPL for each feature that reflects the log likelihood of the probability that it is activated (vs. inactivated). PARADIGM IPLs of the 19,504 features within the SuperPathway are available on Synapse (syn4556715). An initial minimum variation filter (at least 1 sample with absolute activity > 0.05) was applied, resulting in 15,502 concepts (5,898 proteins, 7,307 complexes, 1,916 families, 12 mRNAs, 15 miRNAs and 354 abstract processes) with relative activities showing distinguishable variation across tumors (syn4556729) for use in our differential pathway regulator analysis.

Consensus Clustering of Inferred Pathway Activation

Consensus clustering based on the 3,852 most varying features (i.e. IPLs with variance within the highest quartile) was used to identify UM subtypes implicated from shared patterns of pathway inference. Consensus clustering was implemented with the ConsensusClusterPlus package in R (Wilkerson and Hayes, 2010). Specifically, median-centered IPLs were used to compute the squared Euclidean distance between samples, and this distance matrix was used as the input. Hierarchical clustering using the Ward's minimum variance method (i.e. ward inner linkage option) with 80% subsampling was performed over 1,000 iterations, and the final consensus matrix was clustered using average linkage. The number of clusters was selected by considering the relative change in the area under the empirical cumulative distribution function (CDF) curve as well as the average pairwise item-consensus within consensus clusters. We selected a 5-cluster solution, given that solutions with more clusters provided minimal change and decreased the within-cluster consensus.

Differential pathway regulators of each PARADIGM clusters were identified by comparing one cluster vs. all others using the t-test and Wilcoxon Rank sum test with a BH FDR correction. All 15,502 features passing the minimum variation feature were considered in this analysis; features deemed significant (FDR corrected $p < 0.05$) by both tests and showing an absolute difference in group means > 0.05 were selected. Interconnectivity between the selected pathway regulators within the PARADIGM SuperPathway was assessed, and regulatory hubs with $\geq$ 10 differentially activated downstream targets were identified and displayed in a heatmap using the heatmap.plus R package.

Pathway Features Differentiating lncRNA Clusters

Differential pathway regulators of each lncRNA cluster were identified using the t-test and Wilcoxon Rank Sum test with BH FDR correction in a one cluster vs. all others comparison. Only features deemed significant (FDR $p < 0.05$) by both tests and showing an absolute difference in group means > 0.05 were selected. Interconnectivity between these pathway regulators within the PARADIGM SuperPathway was assessed, and regulatory hubs with $\geq$ 10 differentially activated downstream targets were selected. There were a total of 49 PARADIGM differential pathway regulators identified across the four lncRNA clusters. The mean IPL of the selected regulatory hubs was computed within each cluster and scaled across clusters to a mean of 0 and a standard deviation of 1. The resulting scaled mean IPLs are shown in Figure S5B.

MARINa/hMARINa Analysis of Regulator Activity

MARINa (MAster Regulator INference Algorithm) (Lefebvre et al., 2010) and Hierarchical MARINa (hMARINa) were used to evaluate the activity of transcription factors (TFs) and kinases in 80 UM samples.

Creating a Curated Transcription Factor (TF) Regulome

A compendium of TFs and their targets (TF regulons) were created by combining information from four databases:

(i) SuperPathway (Sedgewick et al., 2013): This is the same interaction network used in the PARADIGM analysis (above). Only links that correspond to regulation at the transcriptional level were retained for MARINa and hMARINa use.
(ii) Literome (Poon et al., 2014): The network was filtered to include only transcription links in which the regulator is a known TF.
(iii) Multinet (Khurana et al., 2013): The network was reduced to links that correspond to regulation on transcriptional level.
(iv) ChEA (Lachmann et al., 2010): Data from the Gene Expression Atlas (Petryszak et al., 2014) was used to filter the inferred links in the ChEA database. Specifically, the context likelihood of relatedness (CLR) method (Faith et al., 2007) was used to compute a measure of association between every pair of genes. The top 10% of gene pairs based on the CLR score were intersected with the ChEA network and the overlapping pairs were added to the final combined network.

The combined network includes 72,915 transcriptional regulatory links between 6,735 regulators and their targets. Only regulators with at least 15 targets were considered in the final analysis, which resulted in a final network consisting of 419 TFs with 58,363 total targets (covering a set of 12,754 unique targets).

Creating a Curated Kinase Regulome

Proteins identified as kinases in Manning (Manning et al., 2002) or Uniprot (UniProt Consortium, 2014) were aggregated into a list of 546 kinases. Protein substrates were extracted from PhosphositePlus (Hornbeck et al., 2014) on March 7, 2015. Kinase-substrate interactions were retained if the kinase appeared in the Manning-Uniprot kinase list and the kinase was identified as a human protein in the PhosphositePlus database. The final compendium consisted of 5,388 links between 342 kinases and 2,260 unique substrates.

MARINa Estimate of TF Activity

MARINa regulator activity scores predict each TF's relative activity as a contrast between two cohorts of interest. The activity score is derived from a combined view of the expression levels of each TF's transcriptional targets (the TF regulon), based on the following steps:

(i) The TF regulon is split into positively- and negatively-regulated sets by measuring the Spearman correlation between the expression of the TF and that of each of its targets.
(ii) A t-statistic derived from the difference in gene expression between the two classes of interest is computed for each gene. All genes are ranked based on their t-statistics to produce a gene signature.
(iii) Each TF's activation and inhibition regulons are examined for enrichment in the high or low end of the ranked gene list. The rankings of the positively- and negatively-regulated genes are then combined and examined simultaneously.

A TF whose two target sets show consistent enrichment (i.e. the activated set is enriched for highly ranked genes and the inhibited set is enriched for lowly ranked ones, or vice versa) receives the highest/lowest activity scores respectively.

Hierarchical MARINa (hMARINa) Estimate of Kinase Activity

MARINa is well suited for the analysis of TF activity, because TF proteins are directly involved in changes in expression of their targets. Kinases, on the other hand, regulate their targets post-translationally. Since the expression levels of genes are often poorly correlated with the activity of the proteins they encode, mRNA represents a poor proxy to protein phosphorylation data. In the absence of the latter, the differential activity of a kinase can be estimated using a hierarchical approach (see schematic below) in which activities are computed at two successive levels:

(i) Level 1 activities are inferred for any regulator (TF or kinase) using single-sample MARINa (ssMARINa) (Aytes et al., 2014). ssMARINa infers these activities based on the expression of the regulator's targets within individual samples. Note that the kinase activity score from level 1 analysis is interpreted as an inference about whether kinase targets are "poised" to be regulated, assuming that increased protein levels would often require an increase in mRNA production as a prerequisite.
(ii) Level 2 activities for kinases are inferred by performing a MARINa analysis on the level 1 activities computed in the previous step, rather than the usual gene expression levels. For level 2, the kinase regulome is used in place of the TF regulome, and the targets of each kinase are restricted to those members that are themselves kinases or TFs, i.e. proteins with level 1 imputed activities.

Identifying Pathway Features Differentiating lncRNA Clusters

The lncRNA clusters were dichotomized into one-vs-rest binary comparisons. For each comparison, MARINa was run via the VIPER R package (http://www.bioconductor.org/packages/release/bioc/html/viper.html) (Alvarez et al., 2016); and hMARINa was performed by extending the functionality of the package. Level 3 mRNA data and the curated TF and kinase regulomes were used as inputs. Analysis was limited to TFs with at least 15 targets present in the expression data. Because the kinase regulome is much smaller than the TF regulome, cutoffs for minimum number of kinase substrates were reduced to 10 in the Level 1 analysis and 5 in the Level 2 one. All other settings were identical to those used for inferring TF activity.

Background models were computed by generating 1,000 label permutations. Significance was evaluated by computing p-values against the background distribution and applying a BH FDR correction. The final results provided activity estimates for 393 TFs and 62 kinases in each dichotomy of interest. MARINa features (TFs) with an FDR $\leq$ 0.10 were retained. Since the kinase regulome is significantly sparser than the TF one, the FDR cutoff for hMARINa features was relaxed to 0.15. A total of 113 MARINa and (h)MARINa differential pathway regulators were identified across the four lncRNA clusters. The differential activity for each of these regulators in each lncRNA cluster is shown in Figure S5C.

Statistically significant findings from the PARADIGM and (h)MARINa differential pathway regulator analyses were examined for consistency. For each cluster, pathway regulators with similar findings across the two methods were identified as ''consistent pathway features.'' An expanded definition also included protein complexes or families with components identified by both methods, genes within the same pathway showing complementary inferred activation patterns, as well as abstract processes linked to any of these consistent findings.

lncRNA Pathway Regulator Correlation Networks

The FPKM expression of every lncRNA was correlated with PARADIGM per-sample IPL levels, and with the TF and kinase activities produced by (h)MARINa, using per-sample ssMARINa activity scores. For each lncRNA cluster, correlations between differentially active regulators and lncRNAs were retained if all four of the following criteria were satisfied:

(i) The TF/kinase was identified as a differentially active pathway feature by PARADIGM or (h)MARINa for that cluster, as described above
(ii) The lncRNA had a mean FPKM $\geq$ 5
(iii) The lncRNA had a SAM multiclass FDR q-value $\leq$ 0.05 and the absolute value of its SAM contrast for the cluster was the largest compared to the absolute contrast values for all other clusters
(iv) The absolute value of the Spearman correlation coefficient between the lncRNA and the regulator in question was $\geq$ 0.5

The filtered lncRNA-pathway regulator network for lncRNA cluster 3 contains 188 correlations between 10 lncRNAs with 24 PARADIGM features and 21 (h)MARINa features. Similarly, the filtered lncRNA-pathway regulator network for lncRNA cluster 4 contains 709 correlations between 26 lncRNAs, 29 PARADIGM features and 70 (h)MARINa features. Figure 6 shows the correlation networks of selected regulators and their associated lncRNAs. For the full list of links, respectively, see (https://tcga-data.nci.nih.gov/docs/publications/uvm_2016).

The networks in Figure 6 are augmented by protein-protein interaction and transcriptional regulation links extracted from PhosphositePlus and the SuperPathway (see Curated TF Regulome and Curated Kinase Regulome sections). In addition, regulators that were identified as consistent pathway features by both methods were displayed using the shape of the method that showed higher differential activity. Both the lncRNA cluster 3 network and lncRNA cluster 4 network contain network nodes identified as a MARINa feature, but retain significant correlation links from both lncRNA-MARINa and lncRNA-PARADIGM comparisons.

Relationship of Fold Change between TCGA SCNA Clusters 3 vs. 4, and Association with Time to Metastasis in TCGA and Laurent Monosomy 3 Cases

We processed Laurent microarray expression data (GEO: GSE22138) (Laurent et al., 2011) to 23,520 expression records, using the probe with the highest cohort variance when a gene symbol had data for more than one microarray probe (e.g. for *CD44*, given 13 probes we used 229221_at; for *MALAT1*, given 12 probes we used 224559_at). We then used Ensembl v82 gene symbols and biotypes for 20,425 protein-coding genes and 8,167 lncRNAs or processed transcripts ('lncRNAs') to identify 17,525 expression records for coding genes and 1,227 records for lncRNAs in Laurent data.

Of the 63 Laurent cases with clinical data, we retained the 32 M3 cases. These included 22 with metastasis and 10 without, and had a median event time of 20.4 months. We identified genes that were variably expressed in these 32 samples, finding 13,142 coding gene records above a mean abundance of 2.0 and above the 25th percentile in variance, and 736 lncRNA records with a mean abundance above 1.5 and a variance above the 40th percentile.

Similarly, for 33 TCGA cases had metastasis data, thresholding 20,531 RSEM genes on the 40th percentile (50.3) and 50th variance percentile retained 12,319 variably expressed mRNAs. For 8,167 Ensembl v82 lncRNAs and processed transcripts ('lncRNA'), FPKMs for which we calculated from the mRNA sequence data, thresholding on the 80th mean FPKM percentile (0.087) and 75th variance percentile retained 1,634 variably-expressed lncRNAs.

To identify genes that were associated with time to metastasis in M3 cases, we censored time and status at 5 years for the 32 Laurent and 33 TCGA records. Then, for each of the above expressed Laurent and TCGA coding genes and lncRNAs, we used the median expression to separate cases into high- and low-expressed groups, and used the R survival v2.41-3 to calculate univariate KM log-rank p-values, and univariate Cox hazard ratios (HRs) with 95% confidence intervals.

We used SAM 2-class unpaired analyses (FDR < 0.05) to identify TCGA mRNAs and lncRNAs that were differentially abundant between TCGA unsupervised SCNA clusters 3 vs. 4, which were M3 cases.

We integrated results separately for RSEM/coding genes and lncRNAs, as follows. For TCGA data we merged M3 metastasis-association results with differentially abundant genes in SCNA 3 vs. 4, and assessed the relationship of fold change vs. HR. We then merged these results with Laurent M3 metastasis-association results, and identified genes that had concordant HRs and HR 95% confidence intervals in both cohorts.

QUANTIFICATION AND STATISTICAL ANALYSIS

Quantitative and statistical methods are noted above according to their respective technology and analytic approach.

DATA AND SOFTWARE AVAILABILITY

The data and analysis results are available and can be explored through the Genomic Data Commons (https://gdc.cancer.gov), the Broad Institute GDAC FireBrowse portal (http://gdac.broadinstitute.org), the Memorial Sloan Kettering Cancer Center cBioPortal (http://www.cbioportal.org), the Institute for Systems Biology Regulome Explorer (http://explorer.cancerregulome.org), and the UVM publication page (https://tcga-data.nci.nih.gov/docs/publications). Software tools used in this project are listed in the Key Resources Table.