



OPEN

## Low E2F2 activity is associated with high genomic instability and PARPi resistance

Jonathan P. Rennhack<sup>1</sup> & Eran R. Andrechek<sup>1,2</sup>✉

The E2F family, classically known for a central role in cell cycle, has a number of emerging roles in cancer including angiogenesis, metabolic reprogramming, metastasis and DNA repair. E2F1 specifically has been shown to be a critical mediator of DNA repair; however, little is known about DNA repair and other E2F family members. Here we present an integrative bioinformatic and high throughput drug screening study to define the role of E2F2 in maintaining genomic integrity in breast cancer. We utilized in vitro E2F2 ChIP-chip and over expression data to identify transcriptional targets of E2F2. This data was integrated with gene expression from E2F2 knockout tumors in an MMTV-Neu background. Finally, this data was compared to human datasets to identify conserved roles of E2F2 in human breast cancer through the TCGA breast cancer, Cancer Cell Line Encyclopedia, and CancerRx datasets. Through these methods we predict that E2F2 transcriptionally regulates mediators of DNA repair. Our gene expression data supports this hypothesis and low E2F2 activity is associated with a highly unstable tumor. In human breast cancer E2F2, status was also correlated with a patient's response to PARP inhibition therapy. Taken together this manuscript defines a novel role of E2F2 in cancer progression beyond cell cycle and could impact patient treatment.

Breast cancer remains the leading cause of cancer related deaths in women. This is largely due to two factors, metastasis and the heterogeneity of breast cancer. While metastasis to distal sites is responsible for mortality, the difficulty of treating a heterogeneous disease is one of the primary factors allowing that progression to occur. The heterogeneity of breast cancer is evident in several facets, including histological subtypes, progression and response to treatment. Underlying this diversity are the unique genomic alterations, methylation patterns and the resulting gene expression differences that are recognized in the PAM50 classification system<sup>1,2</sup>. Each of the subtypes present (Luminal A/B, Basal, Claudin Low, HER2 + ve and normal like) have a unique transcriptomic profile, resulting in the dysregulation in key proteins in breast cancer, including the alteration of the E2F family of transcription factors<sup>3-7</sup>.

The E2F family of transcription factors is composed of nine unique family members (E2F1, E2F2, E2F3a, E2F3b, E2F4, E2F5, E2F6, E2F7, and E2F8)<sup>8-10</sup>. Classically they have been divided into family members that activate transcription (E2F1, E2F2, and E2F3a) and repressors of transcription (E2F3b, E2F4, E2F5, E2F6, E2F7, and E2F8). These definitions have recently become less clear, with each family member functioning in both activating and repressing roles depending on the tissue and developmental context<sup>11,12</sup>.

The activators specifically exist in a bound state to Rb at regions containing an E2F motif in a quiescent cell and transcription of controlled genes is repressed. However, upon a growth signal such as myc, cyclin D levels rise which in turn lead to the activation of cyclin dependent kinases, such as CDK4 and 6. Rb is phosphorylated by CDK4 and 6 and is released from the E2F complex at the gene promoter. This allows transcriptional machinery to bind the region and allows for increased transcription of E2F controlled genes<sup>11,13-15</sup>.

The E2F family members are highly conserved with E2F2 containing 46% homology to E2F1<sup>16</sup>. As a result, family members bind a conserved motif with gene specificity contributed by cofactors<sup>12,17,18</sup>. With regards to the development of the lens in an Rb null embryo E2F1, E2F2, and E2F3 loss all have different phenotypes indicating a unique role for each family member<sup>19</sup>. This has also been shown to be true in cancer, with the loss of each E2F family member having independent effects on metastasis in a variety of mouse models<sup>3,4,6</sup>.

While, this role is not governed by the DNA binding sequence, evidence shows that protein co-factors control the specificity of the E2F family members. Work has shown the binding of co-factors such as YY1 and RYBP to direct E2F2 to specific sites not bound by E2F1 including Cdc6<sup>20</sup>. Due to the binding of YY1, RYBP and

<sup>1</sup>Department of Physiology, Michigan State University, East Lansing, MI, USA. <sup>2</sup>Department of Physiology, Michigan State University, 2194 BPS Building, 567 Wilson Road, East Lansing, MI 48824, USA. ✉email: Andrech1@msu.edu

other cofactors E2F2 has been shown to perform many independent functions from other E2F family members including key roles in the development of inflammation in rheumatoid arthritis through STAT1 and PI3K<sup>21</sup>. Interestingly, both of these pathways have important roles beyond cell cycle and indicate the involvement of the E2Fs and E2F2 specifically beyond cell division.

The role of the E2F family has widely been described in cell cycle where the members regulate the G1/S checkpoint in response to Cyclin D levels<sup>22,23</sup>. However, beyond the G1/S checkpoint de-regulation E2Fs have a number of emerging roles in cancer<sup>24</sup>. This includes roles in other aspects of cancer progression including angiogenesis<sup>25</sup>, metabolic reprogramming<sup>26</sup>, and apoptosis<sup>27,28</sup>. Indeed, numerous accounts detail the role of the activators in metastasis of human breast cancer as well as mouse models of the disease<sup>3,4,6,29–31</sup>.

An additional emerging role for the E2Fs has been in the regulation of genomic stability. Specifically, the role of E2F1 has been well defined with both transcriptional and non-transcriptional roles in DNA repair<sup>32</sup>. In response to DNA damage, E2F1 undergoes post translational phosphorylation by ATM<sup>33</sup>, leading to protein stabilization and increased expression of repair proteins. In addition to the transcriptional role in DNA repair, E2F1 is physically recruited to sites of damage. During cases of double stranded breaks<sup>34</sup> or UV damage<sup>35</sup> it was observed that E2F1 formed foci with other damage induced proteins at the site of DNA damage. It has been shown the E2F1 is required for the efficient recruitment of other repair proteins including XPA/XPC<sup>35</sup> and NBS1<sup>36</sup>. Due to the well-defined role of E2F1 in instability and repair, we wanted to understand the specific role of other activator E2Fs in this process. A role for E2F2 specifically has been identified but it is poorly described. It has also been shown that E2F2 is transcriptionally upregulated in response to DNA damage and has been shown to complex with Rad51 and sites of DNA damage in neuronal cells<sup>37</sup>. It is unclear how this is preserved in the context of cancer or the specific mechanism of repair.

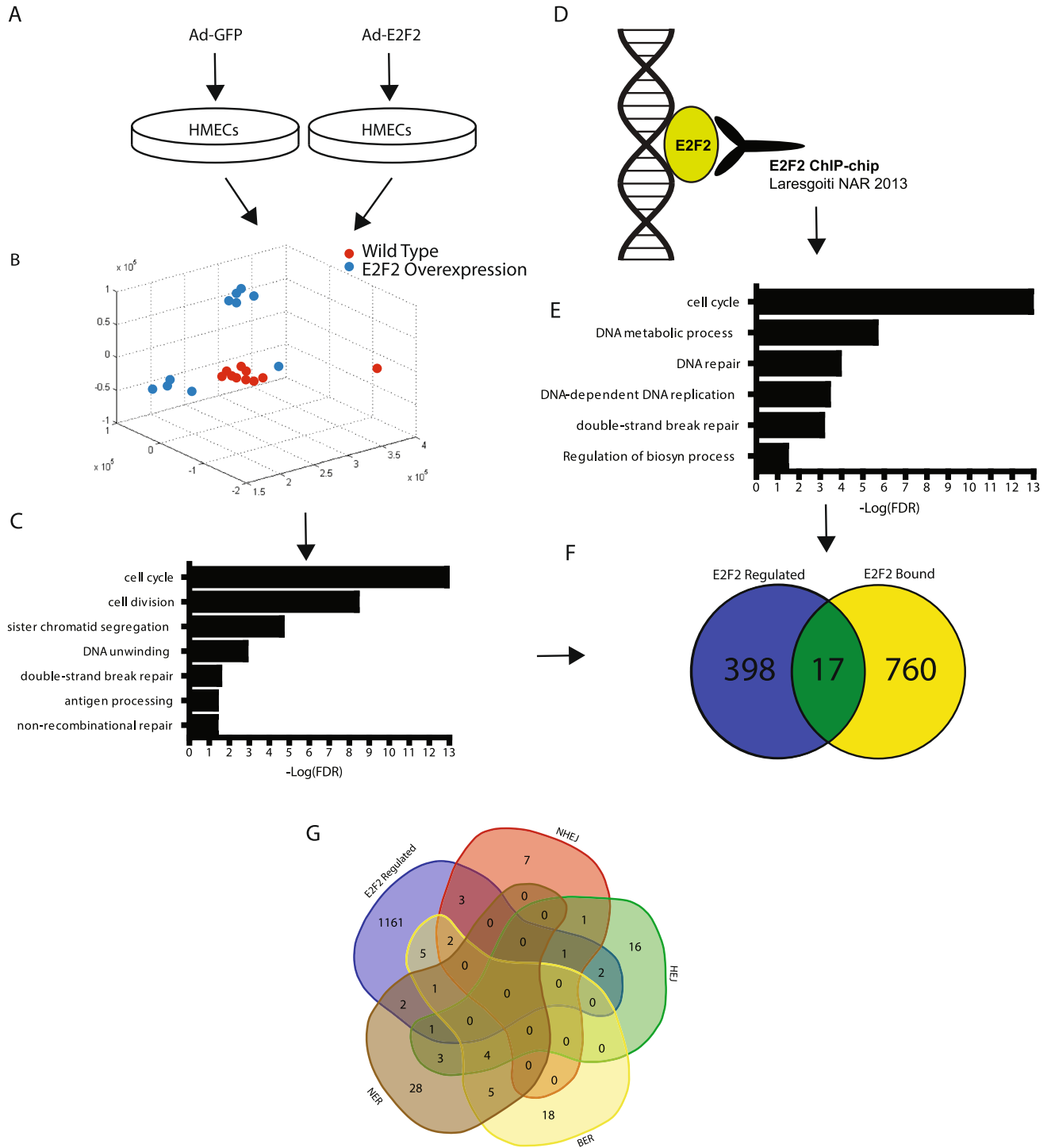
Within the context of cancer, the amplification of the centrosome within a cell leads to defects in cellular segregation and DNA replication, which in turn leads to the single nucleotide variants, copy number alterations, and translocations characteristics. Importantly, activator E2Fs have been shown to be associated with centrosome amplification<sup>38</sup>. It is through this amplification of the centrosome that it is believed E2Fs contribute to the DNA instability associated with their misregulation. However, the mechanism and specific E2Fs involved in this process remains undefined. Specifically, it is unclear what role E2F1, E2F2, and E2F3a play in this process. Together, there is an emerging role for the activator E2Fs role in maintaining genomic integrity, but only the role of E2F1 has been well defined. We believe E2F2 to be of interest and have an important role in DNA repair due to its specific governance of important repair associated proteins such as STAT1. Here we present an integrative bioinformatic analysis leveraging a number of public datasets to define a key role for E2F2 in maintaining genomic integrity. Through the use of cell lines, mouse models, and human samples, we have identified that low E2F2 activity level is associated with tumors containing high levels of genomic instability. Furthermore, E2F2 status has potential clinical implications. Indeed, tumors with high E2F2 activity have an increased sensitivity to cell cycle targeted chemotherapy as well as targeted PARP inhibitors.

## Results

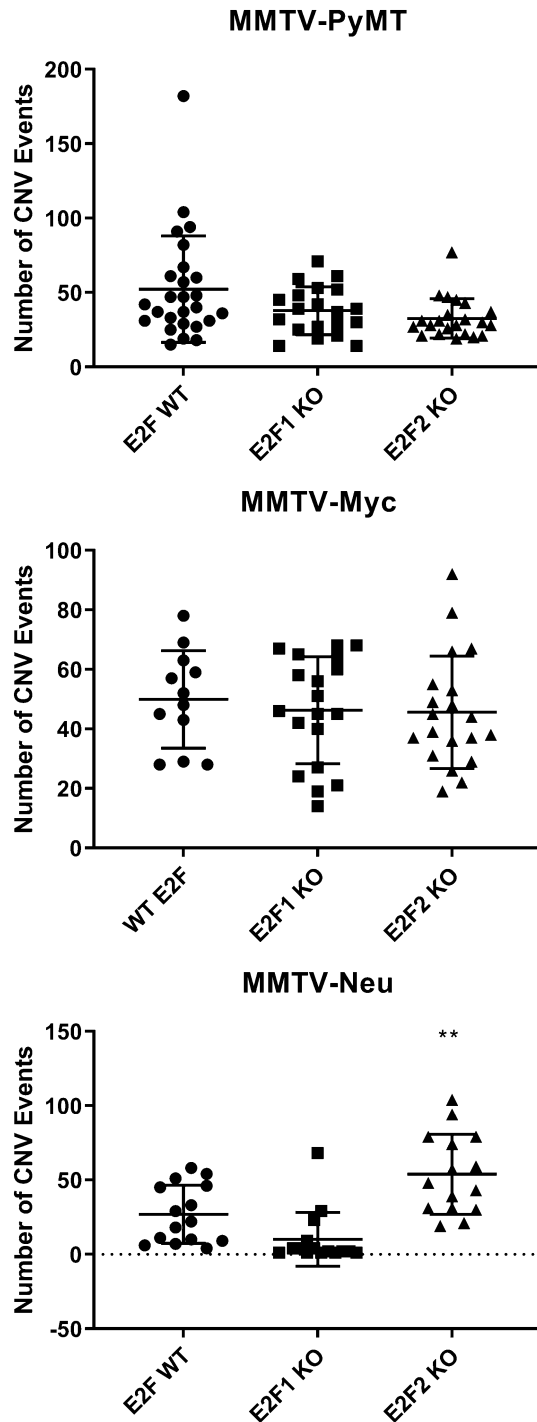
Based on the published literature for E2Fs in non-cell cycle roles, we hypothesized that E2F2 had key activities other than the traditional role in cell cycle. To test this hypothesis, we used principle components analysis on public gene expression data from mouse epithelial fibroblasts infected with adenoviral delivered GFP compared with adenoviral delivered E2F2 (Fig. 1A)<sup>39</sup>. This analysis revealed a consistent gene expression profile associated with over expression of E2F2 (Fig. 1B). We used Significance Analysis of Microarrays (SAM)<sup>40</sup> analysis to identify significantly ( $q < 0.05$ ) overexpressed genes with the infection of Ad-E2F2 relative to GFP. E2F2 induction allowed for the identification of overrepresented gene ontology groups through the use of PANTHER analysis (Fig. 1C) in the E2F2 active mouse embryonic fibroblasts. As expected, this uncovered over-representation of cell cycle proteins. Interestingly we also identified a number of repair associated gene ontology groups, including double-stranded break repair and non-recombinational repair.

To identify genes and pathways directly regulated by E2F2, we utilized publicly available E2F2 ChIP-Chip data (Fig. 1D)<sup>41</sup> from T lymphocytes isolated from 4-week-old C57B16:129SV mice. This analysis utilized a mouse ref-seq promoter (−2000 to +500 base pairs from the transcriptional start site) to identify E2F2 bound regions. This revealed numerous gene promoters bound by E2F2 across the genome. When target genes were analyzed with PANTHER and GATHER, the ontologies were consistent with the E2F2 overexpression data. Indeed, cell cycle and DNA damage repair gene ontology groups were overrepresented, including DNA repair and double-stranded break repair (Fig. 1E) in the T lymphocytes. Surprisingly due to the different nature of the lines used in the analysis, MEFs vs T lymphocytes, we identified consistent gene ontology groups being overrepresented. However, we did not identify a large overlap in the genes identified in the overexpression of genes with the induction of E2F2 in the MEFs vs the E2F2 bound genes in the T-lymphocytes (Fig. 1F). This may indicate that the pathways are regulated by both direct and indirect/downstream targets of E2F2. To determine if there was bias towards one particular repair pathway, we identified overlap between E2F2 regulated genes (gene expression and ChIP-chip) and each of the major repair pathways including Non-Homologous End Joining, Homologous End Joining, Base Excision repair, and nucleotide excision repair (Table S1). This analysis illustrated that E2F2 regulates proteins in each repair pathway (Fig. 1G).

To determine if there is a role for E2F2 in DNA repair processes in the in vivo setting, we utilized publicly available wildtype, E2F1, and E2F2 knockout transcriptome data within multiple genetically engineered mouse models including MMTV-PyMT, MMTV-Myc, and MMTV-Neu. This data was generated from endpoint bulk microarray analysis. We utilized ACE copy number prediction<sup>42</sup> to identify copy number changes within each model. MMTV-PyMT (Fig. 2A) and MMTV-Myc (Fig. 2B) did not show a significant change in number of copy number variations with the loss of E2F1 or E2F2 compared to wildtype. However, MMTV-neu showed a



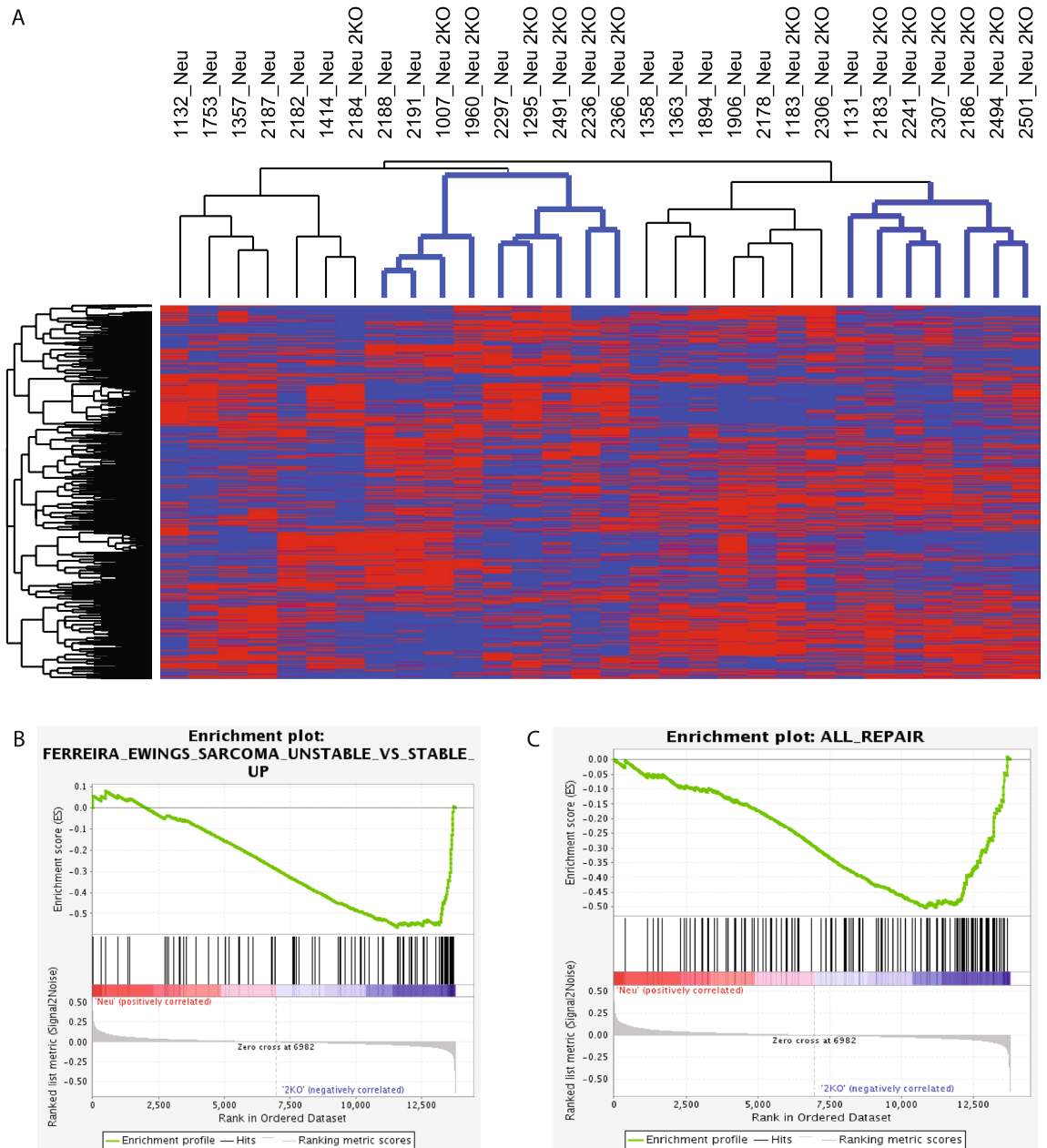
**Figure 1.** E2F2 target genes are enriched for DNA repair associated proteins. To understand genes regulated by E2F2 we utilized E2F2 overexpression data to compare the expression profile of HMECs infected with Ad-GFP and Ad-E2F2 (A). A graphic representation of the first three principle components reveals a consistent transcriptional response associated with E2F2 overexpression through displaying MEFs plus Ad-GFP (blue) and MEFs plus Ad-E2F2 (red) (B). Genes overexpressed with the addition of E2F2 as identified by SAM show a significant ( $FDR < 0.05$ ) overrepresentation in key gene groups including cell cycle and repair (C). ChIP-CHIP of E2F2 binding genes show binding across the genome (D). The binding targets show a significant ( $FDR < 0.05$ ) overrepresentation in a cancer related gene groups (E). A Venn Diagram showing genes predicted to be regulated by E2F2 from ChIP-Chip binding and over expression analysis show a small overlap in genes (F). E2F2 overexpressed or bound genes (E2F2 regulated) are shown to play a role in major repair pathways including Non-Homologous End Joining, Homologous End Joining, Base Excision Repair, and Nucleotide Excision Repair (G).



**Figure 2.** E2F2 loss is associated with higher number of copy number alterations in the MMTV-Neu mouse model. Copy number analysis obtained through the ACE algorithm for MMTV-PyMT tumor (A), MMTV-Myc (B) E2F2 knockout tumor and MMTV-Neu (C) models with E2F WT compared with E2F1 and E2F2 null models. This is shown to be significant across the entire genome with the E2F2 null samples having a significant increase in number of copy number event and the E2F1 null samples having a decrease in the number of copy number alterations in only the MMTV-Neu model (\*\* $P < 0.01$ ).

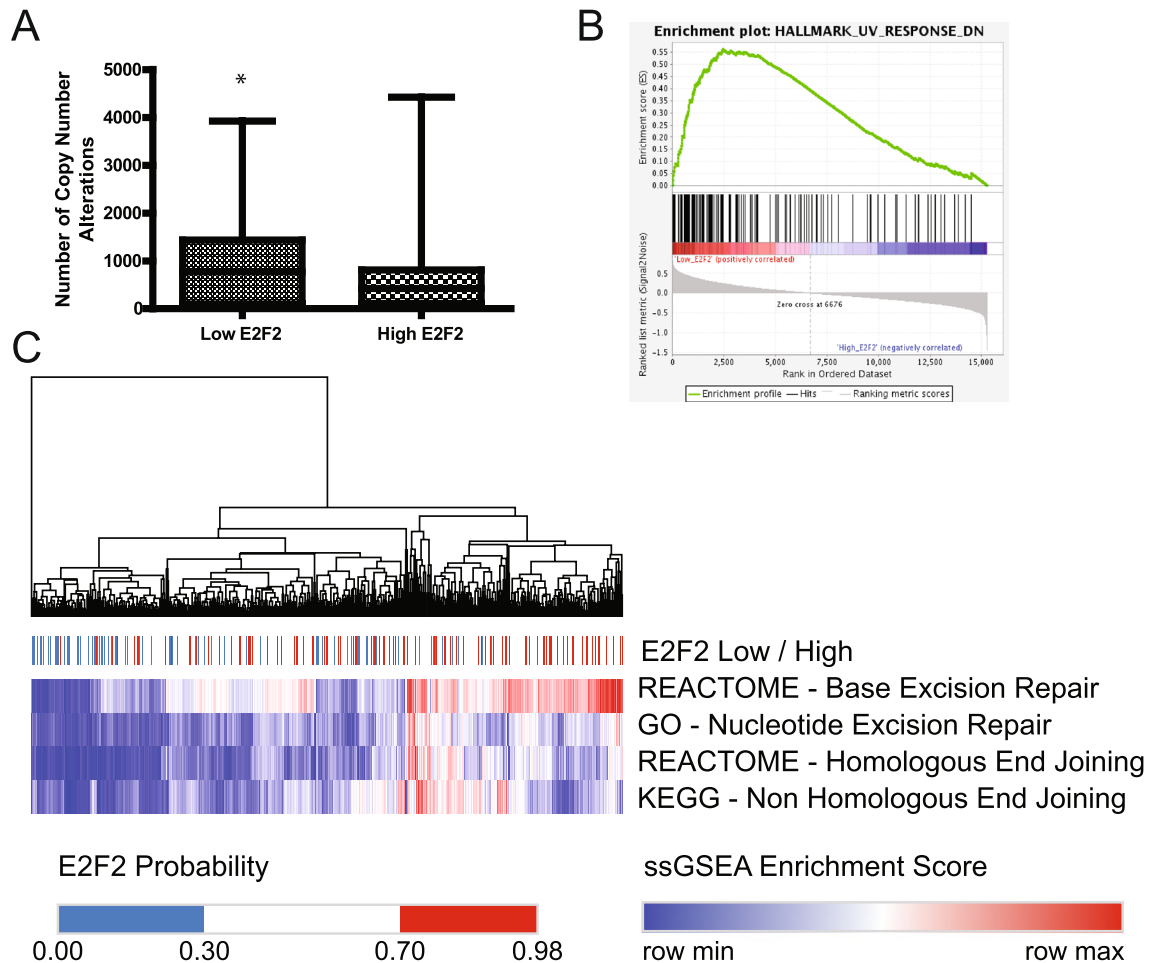
significant gain in the number of copy number alterations with the loss of E2F2 that was not seen with the loss of E2F1 (Fig. 2C).

To continue understanding the cause of this increased copy number variation we further investigated the transcriptomic data of the MMTV-Neu wildtype compared to the E2F2 null tumors. Unsupervised clustering



**Figure 3.** Loss of E2F2 is associated with an enrichment of genomic instability markers. E2F2 loss is associated with consistent transcriptional responses as revealed by unsupervised hierarchical clustering. MMTV-Neu or MMTV-Neu E2F2 knockout samples are arranged by column and unique genes by row. Gene expression values are represented by color from low (blue) to high (red) as indicated by the color bar. It is revealed that 2 main clusters contain an overrepresentation of MMTV-Neu, E2F2 KO samples (blue) (A). Consistent gene expression sets are enriched with the loss of E2F2 as identified by gene set enrichment analysis. This includes Ewings sarcoma unstable vs stable up (B) and all repair gene sets (C).

identified a consistent transcriptional profile with E2F2 loss in the MMTV-Neu mouse model (Fig. 3A). This revealed five major clusters, three primarily composed of MMTV-Neu E2F2 knockout samples and two clustered primary populated with MMTV-Neu E2F2 wildtype samples. This demonstrated a unique gene expression profile associated with the loss of E2F2, which was unique relative to the MMTV-Neu E2F2 wildtype background. To explore the enriched cellular processes in this data, we used Gene Set Enrichment Analysis (GSEA) comparing MMTV-Neu tumors with and without E2F2. This revealed, similar to the in vitro data, that E2F2 null samples were enriched for both the instability gene set (Fig. 3B) and the repair gene set (Fig. 3C). We do not believe this to be in conflict with the in vitro data with an upregulated gene expression signature for repair not necessarily correlating with an increase in the repair process and may be a compensation mechanism of a deficiency of repair processes in which E2F2 target genes are upregulated by other E2F family member; however, repair is unable to occur.



**Figure 4.** Low E2F2 activity is associated with decrease repair gene expression in the TCGA breast cancer cohort. The TCGA cohort reveals that low E2F2 activity is associated with higher number of genes with copy number alterations (A). GSEA analysis shows an enrichment of the Hallmark\_UV\_Reponse\_down geneset ( $p < .05$ ) (B). Unsupervised hierarchical clustering of ssGSEA data for TCGA breast cancer patients for each labeled geneset. Scores are color coordinated with low enrichment in blue and high enrichment scores in red. Samples are labeled according to low (blue) or high (red) E2F2 status (C).

While we have identified a potential role for E2F2 associated instability in mouse tumors, this role has not previously been examined in human breast cancer. Consistent with the mouse data we identified E2F2 high HER2 positive breast cancer to have worse overall survival (Fig. S1) as determined by log rank P-value across the KMplot dataset<sup>43,44</sup>. This is in contrast to all other subtypes of breast cancer which showed no correlation with overall survival. To address a potential role for E2F2 in breast cancer, an E2F2 activity signature was used to divide the TCGA breast cohort into low/high E2F2 activity groups. To create these predictions we used Ad-GFP and Ad-E2F2 infected cell lines and uses binary regression analysis to compare the TCGA data to the known adenovirus infected controls to get an activity score between 0 (low E2F2 activity) and 1 (high E2F2 activity) as previously described<sup>39,45,46</sup>. Consistent with the prediction from the mouse mammary tumor data, human breast cancer with low E2F2 activity contained significantly more copy number variants than those with high E2F2 activity ( $p < 0.05$ ) (Fig. 4A). However, we did not identify this in any particular PAM50 subtype (Fig. S2A,B) as determined by a t-test.

Furthermore, tumors with low predicted E2F2 activity were observed to have enrichment for genomic instability in gene set enrichment analysis such as genes involved with the response to UV induced damage (Fig. 4B).

In order to determine if E2F2 preferentially regulated a specific repair pathway, we utilized single sample gene set enrichment analysis (ssGSEA)<sup>47</sup>. The resulting scores were low/high normalized to return an activity score between 0 and 1 for the four major repair pathways. Low E2F2 activity resulted in significantly lower activity in each repair pathway including: Base Excision Repair, Nucleotide Excision Repair, Homologous End Joining, and Non-Homologous End joining (Fig. S3). Unsupervised hierarchical clustering of this data revealed that regardless of the pathway, low E2F2 activity was associated with low ssGSEA pathway scores (Fig. 4C).

Alterations in DNA repair pathways are manifested in response to therapy. Accordingly, we sought to test whether E2F2 levels impacted therapeutic response using the CancerRx dataset. After predicting the E2F2 activity level across all breast cancer datasets we identified differentially lethal compounds between E2F2 low cell lines



and E2F2 high cell lines (Table S2). This revealed a number of interesting candidate compounds. For example, tumors with low E2F2 activity responded poorly to cell cycle inhibiting compounds such as Cisplatin. However, they responded well to PIK3 targeted therapy such as PI-103. In addition, we observed that tumors with high E2F2 activity were sensitive to cell cycle compounds and resistant to other forms of therapy. This is in contrast to E2F1 in which activity was not correlated to sensitivity with any compound (Table S2). E2F1 activity is correlated to resistant to an EGFR inhibitor, HG-5-88-01, and MAPK inhibitor, ZG-10.

With a role for E2F2 in repair, we examined response to repair targeted therapy, including PARP inhibitors. Surprisingly we identified that high E2F2 activity was associated with a significantly higher response to common PARP inhibitors including Talazoparib (Fig. 5A), Olaparib (Fig. 5B), and Rucaparib (Fig. 5C) as identified by a significant decrease in the IC50 of each compound on each cell line. Importantly, this correlation was independent of the BRCA1/2 status of the cell line. To test if this correlation was a function of any E2F or specific to E2F2, we performed a similar analysis for E2F1. In this analysis, we predicted E2F1 activity through the use of a gene activity signature and identified the IC50 of each cell line. Importantly, in this analysis we saw no difference in the IC50 or PARP targeted therapy in low or high E2F1 activity (Fig. S4). This indicates the identified differences in response to PARP targeted therapy are specific for E2F2.

## Discussion

In this manuscript we perform a re-analysis of publicly available data to uncover new roles of E2F2. In this manuscript we use a number of different cell types and models to identify conserved function of E2F2. Through this we identified a number of E2F2 target associated genes are involved with repair through the integration of E2F2 expression data and ChIP-Chip data. This was also investigated within the E2F2 null MMTV-null model which showed a large increase in the number of copy number alterations. Interestingly, despite their instability, these tumors showed an enrichment for an increased repair process in their transcriptome. This may be due to previously identified compensation from the E2F family members<sup>48,49</sup> and merits further investigation. In total, we describe a role of E2F2 in repair and maintenance of genome integrity in both MMTV-Neu mouse model mammary tumors as well as in human breast cancer patients.

An important note is that this study was not controlled for cell cycle. Within each model it can be expected that there were a variety of cell cycle states present. It is known that E2F2 is more active in some states rather than other, so some roles of E2F2 may have been muted by the diversity of cell states. This merits further investigation.

Despite this heterogeneity, we have identified a consistent role of E2F2 in DNA stability across a combination of *in silico*, *in vitro* and *in vivo* models. However, due to the high-throughput nature of these datasets they have a high propensity to generate false negative or positive data. Throughout, we have utilized stringent statistical parameters to minimize these, but this should be considered when understanding this data. As with any high-throughput study, this data should be used as a hypothesis generating tool, that requires additional confirmation and mechanistic follow-up. We propose that E2F2 controls key members of many different repair pathways including HEJ, NHEJ, BER, and NER. This involvement is associated with a genomically unstable tumors with the loss of E2F2 activity in both the mouse model and the human disease.

Furthermore, this finding has potential clinical application. We have identified E2F2 activity as a biomarker for response to cell cycle inhibition therapy. E2F2 activity, as determined by a gene expression signature, correlates with PARP inhibition therapy. Interestingly, although tumors with lowly active E2F2 are significantly more unstable than E2F2 tumors they are resistant to PARP inhibitor therapy. We have identified that this effect is independent of BRCA1/2 status and is correlated with E2F2 levels. This indicates that E2F2 has a role in PARP response and loss of E2F2 may phenocopy or directly trigger other known causes of PARP inhibitor resistance. This is unintuitive and mechanistically this would be important to study. Potentially, due to the fact that E2F2 low tumors are highly unstable they have adapted to better tolerate DNA damage. As a result, they are more able to tolerate PARPi induced damage than an E2F2 high cell line.

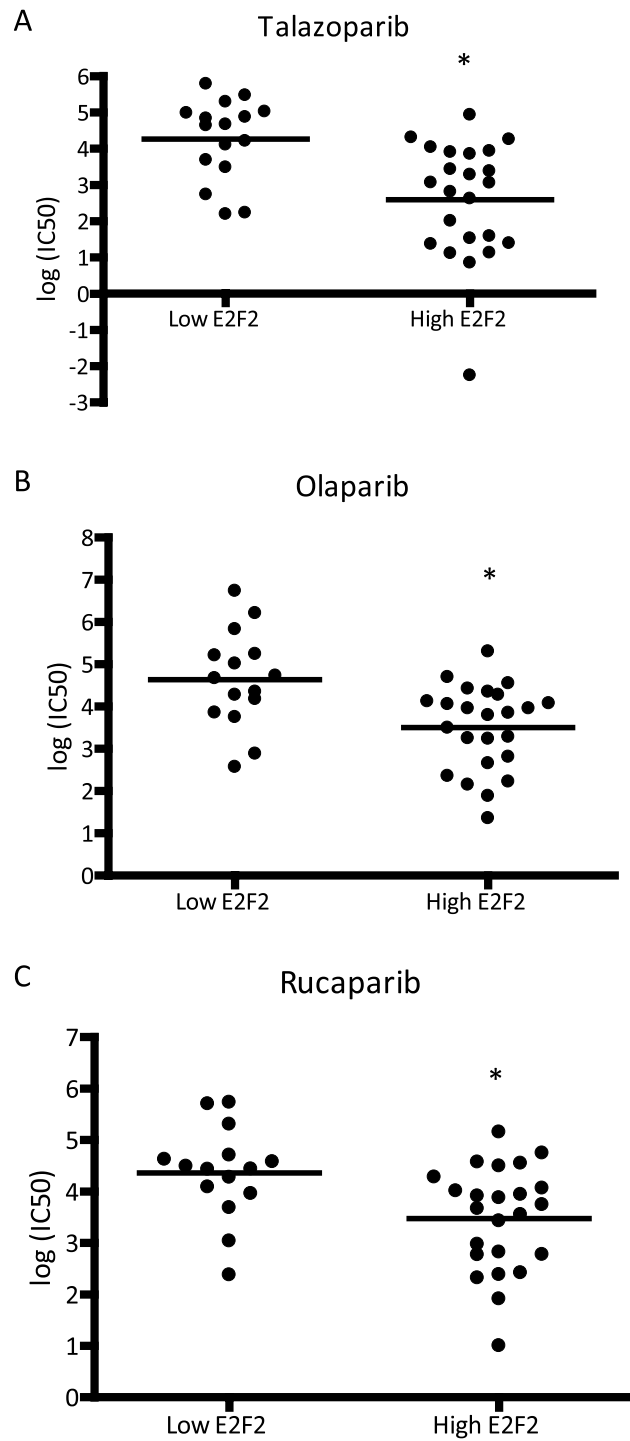
This manuscript serves as an informatic proof of concept study for an expanded role of the E2Fs in DNA repair in breast cancer. The data presented here, combined with the previously established literature, show key roles of E2F1 and E2F2 as contributing factors in the genomic instability seen in breast cancer. Further research needs to be completed of the other E2F family members to identify if this role is unique to E2F1 and E2F2 or if other E2Fs may play similar roles. Indeed, while certain E2Fs have specific roles in mammary development<sup>49,50</sup> and tumor biology<sup>3,4,6</sup>, there is significant overlap and compensation amongst the E2Fs<sup>48,49</sup>. Taken together, this data shows the key role that the E2Fs play in cancer progression and heterogeneity. As central drivers of repair and due to the impact they have on the ability of a tumor to respond to key therapies, there must be more research to understand the clinical application of E2F status and the way that it should shape patient care.

## Materials and methods

**Datasets used.** For the E2F2 overexpression data, Affymetrix cDNA microarray profiled gene expression data was downloaded from previously published data<sup>39</sup>. This dataset consisted of adGFP and adE2F2 infected mouse embryonic fibroblasts. The data was processed using Mas5 normalization through Affymetrix expression console.

For E2F2 binding analysis we utilized ChIP-Chip for data for E2F2 T lymphocytes isolated from 4-week-old C57B16:129SV mice<sup>41</sup>. This manuscript identified E2F2 bound promoters through the use of a gene promoter (− 2000 to + 500 base pairs from the transcriptional start site) and processed through the use of the NimbleScan program.

For the E2F2 knockout data in the various mouse backgrounds we downloaded GSE24594<sup>4</sup> GSE104397<sup>6,51</sup> and GSE42533<sup>3</sup> for the MMTV-Myc, MMTV-PyMT, and MMTV-Neu data accordingly. Raw data was downloaded



**Figure 5.** Low E2F2 activity is associated with resistance to PARP inhibitors across breast cancer cell lines. Breast cancer cell lines divided in lowly active E2F2 and highly active E2F2 show a consistent resistance to PARP targeted therapies associated with low E2F2 activity ( $p < 0.05$ ) for Talazoparib (A), Olaparib (B), and Rucaparib (C).

from ncbi gene expression omnibus and processed through the use of Affymetrix expression console and RMA normalization.

**Overrepresentation analysis.** All overrepresentation analysis experiments were performed through the use of GATHER<sup>52</sup> and PANTHER bioinformatic analysis to identify overrepresented gene ontology groups. Significant groups were noted filtered by a p value of less than 0.05 and a Bayes factor greater than 3.



**ssGSEA and clustering analysis.** ssGSEA was performed on the Broad genepattern software with the designated genesets for each repair pathway downloaded from msig-db. These scores were normalized between 0 (low) and 1 (high). Samples were clustered through the use of Morpheus (<https://software.broadinstitute.org/morpheus/>). Clustering was performed using Complete linkage unsupervised hierarchical clustering. Heatmaps were visualized using MATLAB.

**ACE analysis.** ACE analysis was performed as previously described to infer copy number alterations from gene expression data<sup>42,53</sup>. The analysis was performed with default parameters and used a significance cutoff of  $q < 0.05$ .

**E2F2 activity levels.** E2F2 activity was assayed using a gene expression signature as previously described<sup>45,46,50</sup>. Briefly this method identifies differentially expressed genes between Ad-GFP and Ad-E2F2 infected cell lines and uses binary regression analysis to compare unknown samples and known controls from each group to get a score between 0 (low E2F2 activity) and 1 (high E2F2 activity).

**Copy number analysis of TCGA cohort.** For the human patient analysis, we utilized the TCGA breast cancer cohort<sup>54</sup>. Patients were classified as E2F2 high or low through the gene expression signature approach described above. The gene copy number changes was identified as the number of genes which had a deep deletion or high level amplification as identified by a GISTIC score of  $-2$  or  $2$ . p-value was determined by t-test between low and high E2F2 activity cohorts.

**Survival analysis.** For the overall survival data, we utilized the KMplot.com dataset<sup>43,44</sup>. Patients were identified into the E2F2 high (top 33%) or low (rest of samples) by E2F2 expression levels and classified by PAM50 status. P-value was calculated by log rank test.

**E2F2 drug sensitivity data.** Breast cancer gene expression data was downloaded from the cancer cell line encyclopedia<sup>55</sup>. From this data E2F2 activity was predicted as described above. For drug sensitivity data, we downloaded the small molecule sensitivity dataset from CancerRx.org. Breast cancer cell lines were divided into high and low E2F2 activity groups and significantly different compounds between the two groups were identified by significantly different IC50's as identified by a student's T-test.

**Informed consent.** This articles does not contain any studies with human participants performed by any of the authors.

## Data availability

All datasets used in this publication can be accessed on Gene Expression Omnibus through their appropriate GSE number as noted in the manuscript. This article is present on a university repository website and can be accessed on [https://urldefense.com/v3/https://www.biorxiv.org/content/10.1101/777870v1\\_1;!!HXCxUKc!klXu58cpO7ltAWssBJNGruwrIBog3joS5t3Ul-X4DHqhj9G3MApaifU2FW4UQs](https://urldefense.com/v3/https://www.biorxiv.org/content/10.1101/777870v1_1;!!HXCxUKc!klXu58cpO7ltAWssBJNGruwrIBog3joS5t3Ul-X4DHqhj9G3MApaifU2FW4UQs).

Received: 13 February 2020; Accepted: 28 September 2020

Published online: 21 October 2020

## References

- Bastien, R. R. L. *et al.* PAM50 breast cancer subtyping by RT-qPCR and concordance with standard clinical molecular markers. *BMC Med. Genomics* **5**, 1 (2012).
- Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).
- Andrechek, E. R. HER2/Neu tumorigenesis and metastasis is regulated by E2F activator transcription factors. *Oncogene* **34**, 217–225 (2015).
- Fujiwara, K., Yuwanita, I., Hollern, D. P. & Andrechek, E. R. Prediction and genetic demonstration of a role for activator E2Fs in myc-induced tumors. *Cancer Res* **71**, 1924–1932 (2011).
- Ho, G. H., Calvano, J. E., Bisogna, M. & Van Zee, K. J. Expression of E2F-1 and E2F-4 is reduced in primary and metastatic breast carcinomas. *Breast Cancer Res. Treat.* **69**, 115–122 (2001).
- Hollern, D. P., Honeysett, J., Cardiff, R. D. & Andrechek, E. R. The E2F transcription factors regulate tumor development and metastasis in a mouse model of metastatic breast cancer. *Mol. Cell Biol.* **34**, 3229–3243 (2014).
- Worku, D. *et al.* Evidence of a tumour suppressive function of E2F1 gene in human breast cancer. *Anticancer Res.* **28**, 2135–2139 (2008).
- Nevins, J. R. The Rb/E2F pathway and cancer. *Hum. Mol. Genet.* **10**, 699–703 (2001).
- Trimarchi, J. M. & Lees, J. A. Sibling rivalry in the E2F family. *Nat. Rev. Mol. Cell Biol.* **3**, 11–20 (2002).
- Van Den Heuvel, S. & Dyson, N. J. Conserved functions of the pRB and E2F families. *Nat. Rev. Mol. Cell Biol.* **9**, 713–724 (2008).
- Chong, J. L. *et al.* E2f1-3 switch from activators in progenitor cells to repressors in differentiating cells. *Nature* **462**, 930–934 (2009).
- Freedman, J. A., Chang, J. T., Jakoi, L. & Nevins, J. R. A combinatorial mechanism for determining the specificity of E2F activation and repression. *Oncogene* **28**, 2873–2881 (2009).
- Leone, G., DeGregori, J., Sears, R., Jakoi, L. & Nevins, J. R. Myc and Ras collaborate in inducing accumulation of active cyclin E/Cdk2 and E2F. *Nature* **387**, 422–426 (1997).
- Dyson, N. The regulation of E2F by pRB-family proteins. *Genes Dev.* **12**, 2245–2262 (1998).
- Pierce, A. M., Schneider-Broussard, R., Philhower, J. L. & Johnson, D. G. Differential activities of E2F family members: unique functions in regulating transcription. *Mol. Carcinog.* **22**, 190–198 (1998).
- Ivey-Hoyle, M. *et al.* Cloning and characterization of E2F-2, a novel protein with the biochemical properties of transcription factor E2F. *Mol. Cell Biol.* **13**, 7802–7812 (1993).

17. Rabinovich, A., Jin, V. X., Rabinovich, R., Xu, X. & Farnham, P. J. E2F in vivo binding specificity: comparison of consensus versus nonconsensus binding sites. *Genome Res.* **18**, 1763–1777 (2008).
18. Zheng, N., Fraenkel, E., Pabo, C. O. & Pavletich, N. P. Structural basis of DNA recognition by the heterodimeric cell cycle transcription factor E2F-DP. *Genes Dev.* **13**, 666–674 (1999).
19. Saavedra, H. I. *et al.* Specificity of E2F1, E2F2, and E2F3 in mediating phenotypes induced by loss of Rb. *Cell Growth Diff.* **13**, 215–225 (2002).
20. Schlisio, S., Halperin, T., Vidal, M. & Nevins, J. R. Interaction of YY1 with E2Fs, mediated by RYBP, provides a mechanism for specificity of E2F function. *EMBO J.* **21**, 5775–5786 (2002).
21. Wang, S., Wang, L., Wu, C., Sun, S. & Pan, J. H. E2F2 directly regulates the STAT1 and PI3K/AKT/NF-kappaB pathways to exacerbate the inflammatory phenotype in rheumatoid arthritis synovial fibroblasts and mouse embryonic fibroblasts. *Arthritis Res. Ther.* **20**, 225 (2018).
22. Sears, R. C. & Nevins, J. R. Signaling networks that link cell proliferation and cell fate. *J. Biol. Chem.* **277**, 11617–11620 (2002).
23. Lee, R. J. *et al.* Cyclin D1 is required for transformation by activated Neu and is induced through an E2F-dependent signaling pathway. *Mol. Cell Biol.* **20**, 672–683 (2000).
24. Chen, H. Z., Tsai, S. Y. & Leone, G. Emerging roles of E2Fs in cancer: an exit from cell cycle control. *Nat. Rev. Cancer* **9**, 785–797 (2009).
25. Weijts, B. *et al.* Atypical E2Fs inhibit tumor angiogenesis. *Oncogene* **37**, 271–276 (2018).
26. Fajas, L. Metabolic control in cancer cells. *Ann. Endocrinol.* **74**, 71–73 (2013).
27. Phillips, A. C. & Vousden, K. H. E2F-1 induced apoptosis. *Apoptosis* **6**, 173–182 (2001).
28. Hallstrom, T. C., Mori, S. & Nevins, J. R. An E2F1-dependent gene expression program that determines the balance between proliferation and cell death. *Cancer Cell* **13**, 11–22 (2008).
29. Rennhack, J. & Andrechek, E. Conserved E2F mediated metastasis in mouse models of breast cancer and HER2 positive patients. *Oncoscience* **2**, 867–871 (2015).
30. Yuwanita, I., Barnes, D., Monterey, M. D., O'Reilly, S. & Andrechek, E. R. Increased metastasis with loss of E2F2 in Myc-driven tumors. *Oncotarget* **6**, 38210–38224 (2015).
31. Wu, L. *et al.* Selective roles of E2Fs for ErbB2- and Myc-mediated mammary tumorigenesis. *Oncogene* **34**, 119–128 (2015).
32. Biswas, A. K. & Johnson, D. G. Transcriptional and nontranscriptional functions of E2F1 in response to DNA damage. *Cancer Res* **72**, 13–17 (2012).
33. Lin, W. C., Lin, F. T. & Nevins, J. R. Selective induction of E2F1 in response to DNA damage, mediated by ATM-dependent phosphorylation. *Genes Dev.* **15**, 1833–1844 (2001).
34. Liu, K., Lin, F. T., Ruppert, J. M. & Lin, W. C. Regulation of E2F1 by BRCT domain-containing protein TopBP1. *Mol. Cell Biol.* **23**, 3287–3304 (2003).
35. Guo, R. *et al.* E2F1 localizes to sites of UV-induced DNA damage to enhance nucleotide excision repair. *J. Biol. Chem.* **285**, 19308–19315 (2010).
36. Chen, J. *et al.* E2F1 promotes the recruitment of DNA repair factors to sites of DNA double-strand breaks. *Cell Cycle* **10**, 1287–1294 (2011).
37. Castillo, D. S. *et al.* E2F1 and E2F2 induction in response to DNA damage preserves genomic stability in neuronal cells. *Cell Cycle* **14**, 1300–1314 (2015).
38. Lee, M. Y., Moreno, C. S. & Saavedra, H. I. E2F activators signal and maintain centrosome amplification in breast cancer cells. *Mol. Cell Biol.* **34**, 2581–2599 (2014).
39. Huang, E. *et al.* Gene expression phenotypic models that predict the activity of oncogenic pathways. *Nat. Genet.* **34**, 226–230 (2003).
40. Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. USA* **98**, 5116–5121 (2001).
41. Laresgoiti, U. *et al.* E2F2 and CREB cooperatively regulate transcriptional activity of cell cycle genes. *Nucleic Acids Res.* **41**, 10185–10198 (2013).
42. Hu, G., Wei, Y. & Kang, Y. The multifaceted role of MTDH/AEG-1 in cancer progression. *Clin. Cancer Res.* **15**, 5615–5620 (2009).
43. Gyorffy, B. *et al.* An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res. Treat.* **123**, 725–731 (2010).
44. Gyorffy, B., Surowiak, P., Budczies, J. & Lanczky, A. Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. *PLoS ONE* **8**, e82241 (2013).
45. Bild, A. H. *et al.* Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* **439**, 353–357 (2006).
46. Gatzka, M. L. *et al.* A pathway-based classification of human breast cancer. *Proc. Natl. Acad. Sci. USA* **107**, 6994–6999 (2010).
47. Barbie, D. A. *et al.* Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108–112 (2009).
48. Kong, L. J., Chang, J. T., Bild, A. H. & Nevins, J. R. Compensation and specificity of function within the E2F family. *Oncogene* **26**, 321–327 (2007).
49. To, B. & Andrechek, E. R. Transcription factor compensation during mammary gland development in E2F knockout mice. *PLoS ONE* **13**, e0194937 (2018).
50. Andrechek, E. R., Mori, S., Rempel, R. E., Chang, J. T. & Nevins, J. R. Patterns of cell signaling pathway activation that characterize mammary development. *Development* **135**, 2403–2413 (2008).
51. Hollern, D. P. *et al.* E2F1 drives breast cancer metastasis by regulating the target gene FGF13 and altering cell migration. *Sci. Rep.* **9**, 10718 (2019).
52. Chang, J. T. & Nevins, J. R. GATHER: a systems approach to interpreting genomic signatures. *Bioinformatics* **22**, 2926–2933 (2006).
53. Rennhack, J., To, B., Wermuth, H. & Andrechek, E. R. Mouse models of breast cancer share amplification and deletion events with human breast cancer. *J. Mammary Gland Biol. Neoplasia* **22**, 71–84 (2017).
54. Network, T. C. G. A. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
55. Barretina, J. *et al.* The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).

## Acknowledgements

We would like to acknowledge member of the Andrechek lab for their helpful comments and critical reading of the manuscript.

## Author contributions

J.R. and E.A. conceived of experiments, J.R. performed experiments and both J.R. and E.A. wrote the manuscript.

## Funding

This work was supported with NIH R01CA160514 and Worldwide Cancer Research WCR-14-1153 to E.R.A as well as NIH 1K00CA212221 to J.P.R.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-74877-1>.

**Correspondence** and requests for materials should be addressed to E.R.A.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020