Nuclear PKM2 promotes pre-mRNA processing by binding folded RNA G-quadruplexes

Dimitrios Anastasakis
National Institutes of Health https://orcid.org/0000-0003-2306-275X
Maria Apostolidi
Yale University
Jesse Rinehart
Yale University
Markus Hafner (markus.hafner@nih.gov)
National Institutes of Health / NIAMS https://orcid.org/0000-0002-4336-6518

Article

Keywords:

Posted Date: April 6th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1495021/v1

Abstract

Nuclear localization of the pyruvate kinase M2 isoform (PKM2) is widely observed in cancer and correlates with its aggressiveness, but its contribution to gene regulation and pathogenesis remains unclear. We find that PKM2 is a non-canonical RNA binding protein that specifically recognizes RNA G-quadruplex (rG4) structures and binds precursor mRNAs (pre-mRNAs) in the nucleus of human cells. Translocation of PKM2 to the nucleus increases expression of transcripts with rG4 elements close to splice sites and promotes RNA biogenesis by competing with repressive rG4 binding proteins, such as HNRNPF. PKM2 target mRNAs are also upregulated in a cancer model undergoing epithelial-to-mesenchymal transition (EMT), and consistently drug-induced exclusion of PKM2 from the nucleus results in downregulated rG4-mRNAs and reduced migration and invasion of triple-negative breast cancer cells. We show a gene-regulatory role for rG4 formation on pre-mRNA and propose that rG4s allow for collective co-transcriptional regulation of related mRNAs by competing RBPs.

Background

Pyruvate kinases are enzymes that catalyze the final step of glycolysis by transferring a phosphate from phosphoenolpyruvate to ADP to generate pyruvate and ATP. In mammals, pyruvate kinases are encoded by two genes: *PKLR* expressed in liver and red blood cells, and *PKM* expressed in all other tissues[1]. *PKM* produces two isozymes, PKM1 and 2, that differ by 22 amino acids[2]. Proliferating cells and tumor cells typically express PKM2, which is crucial to support their characteristic shift in energy metabolism towards aerobic glycolysis, also called Warburg effect[3-5]. The alternative splicing event producing the two PKM isoforms by inclusion of either exon 9 in PKM1 or exon 10 in PKM2 is tightly regulated. In highly proliferative cells, a set of HNRNP proteins upregulated by MYC repressively bind to sequences flanking exon 9 to favor the formation of PKM2 at the expense of PKM1[6], leading to the observed high expression levels of PKM2 in tumors[4,7].

Accumulating evidence shows that non-metabolic activities of PKM2 in tumors promote aerobic glycolysis, cell proliferation, migration, and survival[4,7-10]. Some of the PKM2 non-metabolic activities may be linked to the nuclear translocation of PKM2[11,12], which is widely observed in cancers and correlates with their aggressiveness[13]. Nuclear translocation of PKM2 is dependent on an ERK1/2-dependent phosphorylation at serine 37 (S37) event not observed in PKM1, which in turn results in structural reorganization of the PKM2 complex to expose a nuclear localization signal included in exon 10 that is recognized by the nucleocytoplasmic transporter Importin a5 (KPNA1). In the nucleus, PKM2 may directly participate in gene regulation by incompletely characterized mechanisms[11,12,14,15].

RNA interactome capture, orthogonal organic phase separation (OOPS), as well as crosslinking and immunoprecipitation (iCLIP) approaches recently revealed that PKM2 can bind RNA *in vivo*[16-19], analogous to a number of other metabolic enzymes[20-22]. We hypothesized that a ‘moonlighting’ RNA

binding activity may underlie non-metabolic functions of PKM2 and therefore explored whether PKM2 controls gene expression by directly binding and regulating RNA in the nucleus.

Here we show that PKM2, but not PKM1 is indeed an RNA binding protein (RBP) that selectively and specifically interacts with RNA G-quadruplexes (rG4) with high affinity[23]. In the nucleus of human cells, PKM2 binds thousands of precursor mRNAs (pre-mRNAs) at potential rG4 sites. Furthermore, nuclear PKM2, but not PKM1, increases the expression of rG4 containing mRNAs, by promoting pre-mRNA processing through competition with HNRNPF that binds G-rich sequence elements but prevents their folding into rG4 structures. PKM2 rG4 binding and posttranscriptional gene regulatory activity may represent a key to understand its well-documented role in promoting tumorigenesis. We find that rG4-containing mRNAs are upregulated at the highly tumorigenic hybrid E-M stage in an *in vitro* breast cancer model undergoing EMT. Consistently drug-induced exclusion of PKM2 from the nucleus results in downregulated rG4-mRNAs and reduced migration and invasion of triple-negative breast cancer cells. Our results suggest that nuclear PKM2 functions by promoting expression of rG4 containing mRNAs and that rG4s form on pre-mRNA and have a gene-regulatory role. We propose that rG4 elements may form a novel cancer-regulon and propose PKM2-rG4 RNP as a drug target for EMT-associated tumorigenicity.

Results

Nuclear PKM2 binds G-rich sequences in precursor mRNAs

We recently demonstrated a correlation of nuclear PKM2 levels with aggressive tumor cell phenotypes in breast cancer by developing a phosphospecific antibody for phosphorylated, nuclear PKM2 (PKM2pS37)[13]. Immunoblotting with this antibody showed that PKM2pS37 was also present in a panel of standard cell lines (Extended Data Fig. 1a). Immunofluorescence analysis in HEK293 showed nuclear PKM2pS37 staining patterns similar to those previously observed for cancer cells and patient tumor samples[13] (Extended Data Fig. 1b) and suggested that the potential gene regulatory function of nuclear PKM2 can be studied in these model cell lines.

To characterize the nuclear function of PKM2, we engineered stable HEK293 cells expressing FLAG/HA-tagged PKM2 (FH-PKM2-HEK293) under control of a doxycycline-inducible promoter (Extended Data Fig. 1b, Extended Data Fig. 1c). We next comprehensively mapped the nuclear and cytoplasmic RNA interactome of PKM2 and characterized its RNA recognition elements (RREs) using a modified 4-thiouridine (4SU) PAR-CLIP approach[24,25]. Fluorescent imaging of the crosslinked and ribonuclease-treated FH-PKM2 immunoprecipitates revealed a single band migrating at ~ 100 kDa in both cytoplasmic and nuclear fractions, corresponding to the FH-PKM2-RNP ligated to a fluorescent 3’adapter oligonucleotide (Fig. 1a). We recovered PKM2-bound RNA fragments from the nucleus and cytoplasm of two biological replicates and transformed them into small RNA cDNA libraries for next-generation sequencing. Using PARalyzer[26], we determined clusters of overlapping reads that harbor characteristic T-to-C conversions, diagnostic of 4SU-crosslinking events at higher frequencies than expected by chance.

The biological replicates showed good correlation (Extended Data Fig. 2a-c), which allowed us to define a set of 4,649 and 719 high-confidence nuclear and cytoplasmic PKM2 binding sites, respectively.

PKM2 binding sites were largely non-overlapping (Fig. 1b) and reflected the cellular compartment of origin. PKM2 mainly bound pre-mRNAs in the nucleus while it bound mainly mature mRNAs in the cytoplasm (Fig. 1c). Binding of nuclear and cytoplasmic PKM2 to mRNAs showed little correlation to transcript abundance (Extended Data Fig. 2d), suggesting sequence- or structure-dependent determinants of PKM2 binding, rather than unspecific interactions. Motif analysis revealed a strong enrichment of G-rich sequences and G-triplet-containing 5-mers for the nuclear PKM2 RNPs, while cytoplasmic PKM2 did not show statistically significant enrichment of any sequence (Fig. 1b,d and Extended Data Fig. 2f). Taken together, our data indicate that in the nucleus, PKM2 bound predominantly precursor mRNA at defined, G-rich sites, and that 2.) in the cytoplasm, PKM2 showed strong unspecific interaction with mRNA (Extended Data Fig. 2g), consistent with a recent report placing PKM2 close to the elongating ribosome[17].

In vitro selection with PKM2 enriches for rG4 ligands

To identify sequence and structural determinants of PKM2 RNA binding in an unbiased manner we performed RNA Bind-n-Seq (RBNS), an *in vitro* selection procedure[28]. Recombinantly expressed and purified human PKM2 (Extended Data Fig. 3a) was incubated with a library of fully randomized 29 nt long oligoribonucleotides followed by quantification of input and RBP-bound RNA using small RNA cDNA sequencing. In RBNS experiments PKM2 showed only mild enrichment (1.2-1.4-fold) for G-rich 5-mers (Fig. 2a), seemingly inconsistent with the strong enrichment of Gs in our PAR-CLIP binding sites of nuclear PKM2. This pointed to additional binding determinants beyond primary sequence and therefore, we investigated whether we could identify structural determinants of PKM2 binding. The long G-stretches and enrichment of G-triplets from PAR-CLIP binding sites were plausible candidates to form a type of stable RNA secondary structure called G-quadruplex (rG4), formed by stacked Hoogsteen-bonded guanine tetrads[23]. We reanalyzed our RBNS data and observed a ~ 15-fold enrichment for sequences matching the canonical criteria for stable rG4 structures formed by three stacked tetrads of guanosines (“three-tetrad” rG4; $((GGG(N_{1-7}))_3GGG)$[29,30] (Fig. 2b). This enrichment of rG4-forming sequences was abrogated when performing RBNS in the presence of Li^+, rather than K^+ (Fig. 2c), indicating that PKM2 interacted with folded rG4 structures that typically require large monovalent cations (K^+ or Na^+) for formation [23,31]. Less stable “two-tetrad” G4 structures $(GG(N_{1-7}))_3GG)$[32] were also enriched, but to a much lesser degree (~ 1.5-fold). Our data establish PKM2 as a non-canonical RBP, binding and possibly regulating rG4s.

PKM2 binds rG4 structures in vitro

We confirmed that PKM2 recognized a panel of folded G4s by electrophoretic mobility shift assays (EMSA). PKM2 bound the RNA and DNA form of the widely used G4-forming sequence found in the promoter region of *MYC* with comparable affinity (K_D ~ 300 nM, Extended Data Fig. 3g) to the G4-melting helicase DHX36 (K_D ~ 100nM, Fig. 2d and Extended Data Fig. 3g)[33,34]. In EMSAs, PKM2 also bound

“three-tetrad” rG4-forming oligoribonucleotides corresponding to nuclear PKM2 PAR-CLIP binding sites found in the telomerase RNA component (TERC)[35] and the 3’UTR of the DNA polymerase delta catalytic subunit (POLD1) with high-affinity (K_D ~150 nM) (Fig. 2e,f and Extended Data Fig. 3g).

Recombinant PKM1 (Extended Data Fig. 3a) did not bind rG4-structures *in vitro* (Fig. 2g), suggesting that PKM2 rG4 binding activity is linked to the 22 amino acids encoded by the alternative exon 10, either directly or indirectly through conformational changes. PKM2 alternative splicing, and thus likely rG4-binding activity, is restricted to vertebrates, and consistently, the closely related *E. coli* homolog pykF (43% identity) did not interact with rG4s (Extended Data Fig. 3b). Moreover, in EMSAs, PKM2 did not recognize linear RNA or the less stable “two-tetrad” rG4 structures (Extended Data Fig. 3c), suggesting that it requires stable and folded “three-tetrad” rG4s for binding. PKM2 did not bind tRNAs indicating that its specificity towards RNA structures is limited to folded rG4s. Recombinant PKM2 phosphorylated at S37 (Extended Data Fig. 3d,e), the PKM2 form predominantly found in the nucleus, retained its high affinity for rG4 (Fig. 2g), further supporting our observation that PKM2 directly binds RNA in the nucleus. Using circular dichroism, we found that PKM2 binding to folded rG4 structures did not melt the structure (Extended Data Fig. 4a-d), in contrast to the rG4 remodeling activity reported for DHX36 or CNBP proteins[34,36]. *In vitro* enzymatic assays showed PKM2 activity in the presence of rG4, with and without the physiological activator fructose-1,6-bisphoposphate, suggesting that PKM2 substrate binding sites are not occluded by RNA binding (Extended Data Fig. 4e). Taken together, our results indicate that PKM2 selectively and specifically interacts with folded rG4s with high nanomolar affinity (Extended Data Fig. 3g) and does not alter their structure.

Depletion of nuclear PKM2 decreases abundance of its *pre-mRNA targets*

Next, we aimed to characterize the potential role of nuclear PKM2 in posttranscriptional gene regulation on a transcriptome-wide level using loss- and gain-of-function analyses. In our HEK293 model cell lines PKM2 is the main pyruvate kinase expressed and thus we opted not to directly knockout or knock down PKM2 to avoid anticipated secondary effects due to changes in metabolic activity. Therefore, we generated stable HEK293 cell lines expressing FH-PKM2 fused to a Nuclear Export Signal (NES) at its C-terminus that abrogated nuclear localization and silenced endogenous PKM2, including the nuclear fraction, with siRNAs targeting the 3’UTR of the *PKM* mRNA (Fig. 3a). Using these cell lines for transcriptome profiling, we found that exclusion of PKM2 from the nucleus led to a decrease in nuclear PKM2 PAR-CLIP target mRNA levels. The magnitude of this effect was dependent on the overall strength of binding (defined by the number of crosslinked reads normalized by transcript abundance) with the expression levels of the top 10% of targets (n = 208) reduced on average by 10% (Fig. 3b). This phenomenon was independent of the cell line used and similar results were obtained in MCF-7 cells after depletion of nuclear PKM2 (Extended Data Fig. 4a,b). We concluded that nuclear PKM2 promotes expression of its targets.

Nuclear PKM2 increases the expression of rG4-containing mRNA

A further increase of PKM2 levels in the nucleus of HEK293 cells by overexpression of PKM2 fused to a Nuclear Localization Signal (NLS) (Fig. 3c) resulted only a negligible increase in expression for PKM2 targets (Extended Data Fig. 5b). This observation suggested that most targets were already saturated at endogenous nuclear PKM2 levels, but also that increasing nuclear PKM2 to levels more typically observed in tumors may result in regulation of additional transcripts not captured by our PAR-CLIP experiment. We predicted based on our *in vitro* interaction studies that transcripts containing the canonical “three-tetrad” rG4 motif would be potentially functional PKM2 targets. Therefore, we scanned the human transcriptome for ($GGG(N_{1-7})_3GGG$) and separated transcripts based on the presence and number of this sequence element in exons, or introns within 200 nt or further removed from the 5’ or 3’ splice site. We found that higher PKM2 levels in the nucleus significantly increased the abundance of rG4 containing mRNAs with the magnitude of the effect depending on the number of predicted rG4s within exons or near the splice sites (Fig. 3d,e). Importantly, no change in abundance was observed for mRNAs with rG4s located in introns further than 200 nt from the splice sites (Fig. 3f).

Nuclear PKM2 promotes co-transcriptional processing of rG4-containing pre-mRNAs

Multiple lines of evidence support the notion that DNA G4 structures forming in gene promoters are regulatory elements controlling transcription[37]. While PKM2 is capable to interact with DNA G4 elements *in vitro* (Fig. 2d), nuclear PKM2 did not affect expression of genes with G4s in their promoter in HEK293 or MCF7 cells (Extended Data Fig. 5b-e). Moreover, nuclear PKM2 did not affect the expression of genes with G4 elements on the template strand of the gene body, i.e. the reverse complement of the RNA transcript, further indicating that rG4 structures, rather than DNA-G4 structures, are the relevant PKM2 targets *in vivo* (Extended Data Fig. 5f).

Our data showed that PKM2 prominently bound pre-mRNA, i.e. nascent transcripts, which led us to test if this interaction affected transcriptional dynamics. To address this, we used ChIP-seq to target the unphosphorylated C-terminal domain (CTD) of the large PolII subunit RBP1, which mainly recovers inactive PolII recruited to gene promotors. Our experiments revealed lower occupancy of PolII at transcription initiation sites of rG4 containing mRNAs upon expression of FH-PKM2-NLS (Fig. 4a). Nevertheless, genes encoding rG4 mRNAs show an increase in occupancy of elongating PolII marked by phosphorylation of the CTD serine 2 (S2) (Fig. 4b), indicating an increased transcriptional output. This is mirrored in an overall increase of RNAseq sequence reads derived from introns of target genes with rG4 (Fig. 4c), a metric which can be used to measure transcriptional activity[38]. We also found an increase in the overall ratio of exonic over intronic reads on rG4 mRNAs, which suggested an increased splicing efficiency and also indicated that PKM2 regulates target mRNAs in a co-transcriptional manner (Fig. 4d).

Considering 1.) the tight coupling of transcriptional elongation rates with pre-mRNA splicing[39–41], 2.) that the transcripts most profoundly affected by PKM2 nuclear localization contained rG4 elements near their splice sites and that 3.) we did not observe systematic splicing changes in our RNAseq data, we hypothesized that PKM2 promoted efficient processing of bound nascent mRNA. Mechanistically, the PKM2 effect on nascent transcripts is likely to be mediated passively by pre-mRNA binding and/or displacement of suppressive rG4 binding proteins, rather than an assembly of a larger effector complex. Interaction and recruitment of effectors would require the evolution of additional binding surfaces for those proteins, which appears unlikely considering its close sequence and structural relationship to all other pyruvate kinase enzymes.

PKM2 competes with HNRNPF to stabilize RNA G-quadruplex formation

To identify other nuclear RBPs that could potentially compete with PKM2 for rG4 binding, we mined the ENCODE database consisting of RBNS data for 78 human RBPs and found that HNRNPF showed robust enrichment for rG4 sequences *in vitro* (Extended Data Fig. 6a). Intriguingly, HNRNPF was previously implicated as a splicing repressor binding rG4 elements *in vivo*[42,43]. To investigate a potential relationship between HNRNPF and nuclear PKM2 we generated stable HEK293 cells expressing an inducible FLAG/HA-tagged HNRNPF and used them for fPARCLIP and RNAseq. These experiments confirmed that HNRNPF binds seuences elements containg GGG triplets (Fig. 5b). Analogous to PKM2 binding site distribution, analysis of HNRNPF crosslinking events showed a clear enrichment within 200 nt from 5’ and 3’ splice sites, leading us to assume that these two RBPs share a functional relationship in the posttranscriptional gene regulation of mRNAs containing G-rich elements (Fig. 5c). Consistent with a possible competition for access to rG4-forming elements, the overexpression of HNRNPF showed the opposite effects of PKM2-NLS and resulted in a significantly reduced expression of HNRNPF or PKM2 targets, as well as all rG4 containing mRNAs (Fig. 5d and Extended Data Fig. 6c,d).

We hypothesized that HNRNPF could compete in two ways with PKM2 at rG4-forming sites, either by having a higher affinity to folded rG4s than PKM2, or by stabilizing the unfolded, linear rG4 sequence through interactions with its three quasi RNA recognition motifs[42,44–46]. To distinguish between these possibilities, we turned again to RBNS using the same random 29-mer oligoribonucleotide pool we used for PKM2 RBNS, pre-folded in the presence of either K^+ promoting rG4 formation or Li^+ promoting the linear form. We calculated the enrichment of sequences containing different numbers of GGG repeats (Fig. 5e,f) and found - in contrast to the PKM2 RBNS - a marked decrease in enrichment of sequences containing three or four sets of GGG-repeats that could fold into rG4 structures when performing the RBNS in the presence of K^+ rather than Li^+. This suggested that HNRNPF prefers binding to the linear from of rG4-forming sequences. RBNS results for PKM2 and HNRNPF were similar independent of whether the proteins were expressed recombinantly in *E. coli* or purified from HEK293 cells, ruling out that the lack of typical posttranslational modifications of proteins expressed in bacteria influenced specificity or selectivity of these RBPs (Fig. 5g and Extended Data Fig. 6e,f). Taken together our results suggest that HNRNPF competes with PKM2 by preventing rG4-structure formation at sites harboring four GGG-repeats

and that the balance between folded and unfolded rG4 elements ultimately influences mRNA processing efficiency[44–46].

rG4-containing mRNAs are upregulated in cells undergoing partial EMT

Previous reports show that nuclear translocation of PKM2 promotes invasive properties of tumor cells[14,15]. Consistently, a Gene Set Enrichment Analysis (GSEA) of transcriptional changes in FH-PKM2-NLS-HEK293 cells showed that nuclear PKM2 significantly affected multiple gene sets connected with tumorigenesis, with epithelial-to-mesenchymal transition (EMT)-related genes most significantly changed (Fig. 6a)[47]. This suggests that cells undergoing EMT would rely on nuclear PKM2 for stabilization of rG4-containing mRNAs. To test this hypothesis, we reanalyzed gene expression data from an established breast cancer model of EMT [48]. In this model, cells can be separated into three phenotypic states: a basic, less tumorigenic state (E), an epithelial/mesenchymal hybrid (E/M) state that is strongly associated tumor invasiveness and metastasis, and a final mesenchymal state (xM) that is less tumorigenic.

We found that the top upregulated mRNAs after nuclear overexpression of PKM2 (FH-PKM2-NLS) were also highly enriched when cells transit from the E to the aggressive E/M state. In contrast these same mRNAs were downregulated by HNRNPF overexpression, highlighting a potential competition with PKM2 and, in line with a previous study[42], suggesting an inhibitory role of HNRNPF on EMT (Fig. 6b). As cells acquired the E/M state, we again saw a dramatic increase in abundance of rG4-containing mRNAs either in their exons or introns within 200 bp of their splice site in E/M compared to E cells (Fig. 6c,d). Notably, as the cells transition from the aggressive E/M state to the less tumorigenic xM state, the rG4-transcriptome decreases in abundance again (Fig. 6e). This suggested a correlation between expression levels of rG4-transcripts and cancer phenotypes. The PKM2-rG4 interaction may allow a rapid modulation of key gene sets marked by rG4 to promote a migratory phenotype, which correlates to transitional phases of tumorigenesis and metastasis. Consistently, cell migration regulatory pathways and specific genes related to cell motility are significantly enriched in a gene ontology (GO) analysis of transcripts that are upregulated upon forced nuclear PKM2 expression with a NLS and contain more than three rG4s (Extended Data Fig. 7). Our previous studies have shown that by targeting nuclear PKM2 in a metastatic model of triple negative breast cancer (TNBC), highly aggressive cells could transition to less aggressive phenotypic states both *in vitro* an *in vivo*[13]. We found that TEPP-46, a small molecule that acts as an isoform specific pyruvate kinase activator, reduced nuclear PKM2 and TNBC aggressive phenotypes. To investigate whether nuclear depletion of PKM2 with TEPP-46 would directly modulate the G4-rich RNAs we treated MDA-MB-231 TNBC cells with TEPP-46 and analyzed their transcriptome. Consistent with our previous observations and new insights, TEPP-46 treated cells showed a reduction in rG4-containing transcripts and decreased migratory and invasive properties (Fig. 6g,h). This confirms that nuclear PKM2 localization plays a critical role in shaping the transcriptome of highly aggressive cancer phenotypes via rG4 interactions.

Discussion

PKM2 as rG4-interacting RBP

Our study shows that, in addition to its well-characterized role in the energy metabolism, PKM2 has an unexpected new role in rG4-mRNA regulation in the nucleus. It establishes PKM2 among the more than 20 other metabolic enzymes known to bind RNA[22]. Overall, the effects of RNA binding to metabolic enzymes have been poorly characterized, but could include 1.) modulation of their enzymatic activity, as seen for GAPDH[49], 2.) feedback control of expression of a defined set of mRNAs encoding components of the same metabolic pathways as shown for iron response protein 1 (IRP1) binding to ferritin mRNAs to regulate iron metabolism[50], or 3.) formation of larger assemblies of the metabolic enzymes to form a 'metabolon' with superior metabolic flux capacity[51], as well as heteromeric complexes, analogous to the gene repressing PRC2 complex[52]. Translocation of PKM2 to the nucleus removes it from its normal role in glycolysis and allows it to posttranscriptionally regulate the expression of hundreds of mRNAs. Our analysis suggests this mechanism is likely to contribute to the large-scale rewiring of gene expression programs during phenotypic transitions and possibly metastasis, but also in other biological contexts such as in development, where nuclear PKM2 is required for T-cell differentiation[53].

While the precise molecular mechanism of how PKM2 regulates gene expression is not clear yet, our data suggest that PKM2 affects the efficiency of co-transcriptional mRNA processing feeding back on transcriptional efficiency itself[54], considering 1.) that it acts in the nucleus, 2.) that intronic and exonic rG4 sites close to splice sites confer regulation, 3.) that intronic rG4 sites found more than 200 nt removed from splice sites are occupied but do not affect RNA levels, 4.) that we find no evidence of systematic splicing pattern changes, and 5.) that importantly, we find increased transcriptional elongation measured by ChIP-seq of S2 phosphorylated PolII CTD without increases in PolII occupancy of promotor regions.

Folded rG4 are gene regulatory elements in pre-mRNA

Our study sheds light on the widely debated function of rG4s and highlights their global gene expression regulatory role. Previous systems-wide studies found that rG4 structures are globally unfolded in the cytoplasm due to the action of a highly redundant machinery of RBPs and helicases that either prevent their formation or unwind them[34,36,55,56]. Knockout of these factors and concomitant rG4 formation in the cytoplasm is typically detrimental to the cell, e.g. by interfering with translation initiation or elongation[36,57], or rG4-containing RNA self-aggregation resulting in the activation of stress response pathways[34]. However, rG4 sequences in the vertebrate transcriptome are deeply conserved[58,59]. Our study suggests that PKM2 allows co-transcriptional folding of rG4 elements leading to transcriptome-wide changes in gene expression. Considering that rG4 containing transcripts including PKM2 targets are enriched among transcripts required for the migratory phenotype in cancer cells, the rG4 elements may be considered as a clear example of a 'regulon'[60,61] that allows for the coordinated regulation of these transcripts. PKM2 may cooperate or compete with other nuclear rG4-binding RBPs for the regulation of

these mRNAs. Consistent with this idea we showed that that overexpression of the rG4 sequence element binding protein HNRNPF reduces abundance of its rG4 containing transcripts. Intriguingly, HNRNPF expression is negatively correlated with an EMT signature[42], the opposite of nuclear PKM2 which promotes the expression of EMT associated genes and more specifically those that are associated with highly aggressiveness of cancer cell phenotypes.

PKM2-rG4 interaction as a therapeutic intervention point

We propose that the PKM2-rG4 interaction may be the key to understanding the tumorigenic activity of nuclear PKM2. Disruption of this interaction could represent a therapeutic vulnerability to target migratory cell phenotypes, which coordinate tumorigenesis and metastasis. PKM2-rG4 interactions could be manipulated in multiple ways, e.g. 1.) by influencing PKM splicing patterns using oligoribonucleotide inhibitors suppressing exon 10 inclusion, as prominently used to correct SMN splicing[62], 2.) prevention of PKM2 nuclear translocation, e.g stabilization of the cytoplasmic form[10], or interference with PKM2 phosphorylation[13], or 3.) structure-guided targeting of the, yet to be determined, PKM2-rG4 interaction surface.

Declarations

Acknowledgments

The authors thank the members of the Hafner group and Drs. Michael Banco (NHLBI/NIH) and Eugene Valkov (NCI/NIH) for helpful discussions. The authors thank Dr. Craig Thomas (NCATS/NIH) for providing them with the TEPP-46 compound. We thank Faiza Naz and Dr. Stefania dell'Orso (NIAMS/NIH) for sequencing support, Parthena Konstantinidou (Haase Lab NIDDK/NIH) for image acquisition assistance and Patroula Nathanailidou (Shiv Grewal Lab, NCI/NIH) for image analysis. We also thank Svetlana Rogulina, member of the Rinehart lab, for technical support and overall help and Dr. Joerg Nikolaus of the West Campus Imaging Core (Yale) for training on the use and maintenance of equipment.

Funding

Intramural Research Program of the National Institute for Arthritis and Musculoskeletal and Skin Diseases (D.A, M.H), National Institutes of Health (NIH) / National Cancer Institute (NCI) U54 Grant #CA209992 (M.A, J.R)

Contributions

D.A, M.A, J.R, and M.H. conceived study and designed experiments. D.A, M.A. performed experiments and analyzed data. D.A, M.A, J.R, and M.H. wrote the manuscript. D.A, M.A, J.R, and M.H. performed final edits.

Competing interests

The authors declare no competing interests.

Data availability

All data used in the analysis are available in from the Gene Expression Omnibus (GEO) under the accession number GSE179817.

References

1. Imamura, K. & Tanaka, T. Multimolecular forms of pyruvate kinase from rat and other mammalian tissues. I. Electrophoretic studies. J Biochem **71**, 1043–1051, doi:10.1093/oxfordjournals.jbchem.a129852 (1972).
2. Vander Heiden, M. G., Cantley, L. C. & Thompson, C. B. Understanding the Warburg effect: the metabolic requirements of cell proliferation. Science **324**, 1029–1033, doi:10.1126/science.1160809 (2009).
3. Warburg, O. On the origin of cancer cells. Science **123**, 309–314, doi:10.1126/science.123.3191.309 (1956).
4. Christofk, H. R. *et al.* The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. Nature **452**, 230–233, doi:10.1038/nature06734 (2008).
5. Mazurek, S., Boschek, C. B., Hugo, F. & Eigenbrodt, E. Pyruvate kinase type M2 and its role in tumor growth and spreading. Semin Cancer Biol **15**, 300–308, doi:10.1016/j.semcancer.2005.04.009 (2005).
6. David, C. J., Chen, M., Assanah, M., Canoll, P. & Manley, J. L. HnRNP proteins controlled by c-Myc deregulate pyruvate kinase mRNA splicing in cancer. Nature **463**, 364–368, doi:10.1038/nature08697 (2010).
7. Dayton, T. L., Jacks, T. & Vander Heiden, M. G. PKM2, cancer metabolism, and the road ahead. EMBO Rep **17**, 1721–1730, doi:10.15252/embr.201643300 (2016).
8. Wang, C. *et al.* PKM2 promotes cell migration and inhibits autophagy by mediating PI3K/AKT activation and contributes to the malignant development of gastric cancer. Scientific Reports **7**, 2886, doi:10.1038/s41598-017-03031-1 (2017).
9. Li, C., Zhao, Z., Zhou, Z. & Liu, R. PKM2 Promotes Cell Survival and Invasion Under Metabolic Stress by Enhancing Warburg Effect in Pancreatic Ductal Adenocarcinoma. Digestive Diseases and Sciences **61**, 767–773, doi:10.1007/s10620-015-3931-2 (2016).

10. Anastasiou, D. *et al.* Pyruvate kinase M2 activators promote tetramer formation and suppress tumorigenesis. Nat Chem Biol **8**, 839–847, doi:10.1038/nchembio.1060 (2012).
11. Yang, W. & Lu, Z. Nuclear PKM2 regulates the Warburg effect. Cell Cycle **12**, 3154–3158, doi:10.4161/cc.26182 (2013).
12. Yang, W. *et al.* Nuclear PKM2 regulates β-catenin transactivation upon EGFR activation. Nature **480**, 118–122, doi:10.1038/nature10598 (2011).
13. Apostolidi, M. *et al.* Targeting Pyruvate Kinase M2 phosphorylation reverses aggressive cancer phenotypes. Cancer Res, doi:10.1158/0008-5472.Can-20-4190 (2021).
14. Tanaka, F., Yoshimoto, S., Okamura, K., Ikebe, T. & Hashimoto, S. Nuclear PKM2 promotes the progression of oral squamous cell carcinoma by inducing EMT and post-translationally repressing TGIF2. *Oncotarget* **9**, 33745–33761, doi:10.18632/oncotarget.25850 (2018).
15. Hamabe, A. *et al.* Role of pyruvate kinase M2 in transcriptional regulation leading to epithelial-mesenchymal transition. Proc Natl Acad Sci U S A **111**, 15526–15531, doi:10.1073/pnas.1407717111 (2014).
16. Queiroz, R. M. L. *et al.* Comprehensive identification of RNA-protein interactions in any organism using orthogonal organic phase separation (OOPS). Nat Biotechnol **37**, 169–178, doi:10.1038/s41587-018-0001-2 (2019).
17. Simsek, D. *et al.* The Mammalian Ribo-interactome Reveals Ribosome Functional Diversity and Heterogeneity. Cell **169**, 1051–1065
.e1018
, doi:10.1016/j.cell.2017.05.022 (2017).
18. Baltz, A. G. *et al.* The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol Cell **46**, 674–690, doi:10.1016/j.molcel.2012.05.021 (2012).
19. Castello, A. *et al.* Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. Cell **149**, 1393–1406, doi:https://doi.org/10.1016/j.cell.2012.04.031 (2012).
20. Perez-Perri, J. I. *et al.* Discovery of RNA-binding proteins and characterization of their dynamic responses by enhanced RNA interactome capture. Nat Commun **9**, 4408, doi:10.1038/s41467-018-06557-8 (2018).
21. Hentze, M. W., Castello, A., Schwarzl, T. & Preiss, T. A brave new world of RNA-binding proteins. Nat Rev Mol Cell Biol **19**, 327–341, doi:10.1038/nrm.2017.130 (2018).
22. Castello, A., Hentze, M. W. & Preiss, T. Metabolic Enzymes Enjoying New Partnerships as RNA-Binding Proteins. Trends Endocrinol Metab **26**, 746–757, doi:10.1016/j.tem.2015.09.012 (2015).
23. Sen, D. & Gilbert, W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. Nature **334**, 364–366, doi:10.1038/334364a0 (1988).
24. Hafner, M. *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell **141**, 129–141, doi:10.1016/j.cell.2010.03.009 (2010).

25. Anastasakis, D. G. *et al.* A non-radioactive, improved PAR-CLIP and small RNA cDNA library preparation protocol. Nucleic Acids Res, doi:10.1093/nar/gkab011 (2021).
26. Corcoran, D. L. *et al.* PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. Genome Biol **12**, R79, doi:10.1186/gb-2011-12-8-r79 (2011).
27. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell **38**, 576–589, doi:10.1016/j.molcel.2010.05.004 (2010).
28. Lambert, N. *et al.* RNA Bind-n-Seq: quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. Mol Cell **54**, 887–900, doi:10.1016/j.molcel.2014.04.016 (2014).
29. Huppert, J. L. & Balasubramanian, S. Prevalence of quadruplexes in the human genome. Nucleic Acids Res **33**, 2908–2916, doi:10.1093/nar/gki609 (2005).
30. Todd, A. K., Johnston, M. & Neidle, S. Highly prevalent putative quadruplex sequence motifs in human DNA. Nucleic Acids Res **33**, 2901–2907, doi:10.1093/nar/gki553 (2005).
31. Bhattacharyya, D., Mirihana Arachchilage, G. & Basu, S. Metal Cations in G-Quadruplex Folding and Stability. Front Chem **4**, 38, doi:10.3389/fchem.2016.00038 (2016).
32. Shao, X. *et al.* RNA G-Quadruplex Structures Mediate Gene Regulation in Bacteria. mBio **11**, e02926-02919, doi:10.1128/mBio.02926-19 (2020).
33. Chen, M. C. *et al.* Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36. Nature **558**, 465–469, doi:10.1038/s41586-018-0209-9 (2018).
34. Sauer, M. *et al.* DHX36 prevents the accumulation of translationally inactive mRNAs with G4-structures in untranslated regions. Nature Communications **10**, 2421, doi:10.1038/s41467-019-10432-5 (2019).
35. Lattmann, S., Stadler, M. B., Vaughn, J. P., Akman, S. A. & Nagamine, Y. The DEAH-box RNA helicase RHAU binds an intramolecular RNA G-quadruplex in TERC and associates with telomerase holoenzyme. Nucleic Acids Research **39**, 9390–9404, doi:10.1093/nar/gkr630 (2011).
36. Benhalevy, D. *et al.* The Human CCHC-type Zinc Finger Nucleic Acid-Binding Protein Binds G-Rich Elements in Target mRNA Coding Sequences and Promotes Translation. Cell Rep **18**, 2979–2990, doi:10.1016/j.celrep.2017.02.080 (2017).
37. Huppert, J. L. & Balasubramanian, S. G-quadruplexes in promoters throughout the human genome. Nucleic acids research **35**, 406–413, doi:10.1093/nar/gkl1057 (2007).
38. Gaidatzis, D., Burger, L., Florescu, M. & Stadler, M. B. Analysis of intronic and exonic reads in RNA-seq data characterizes transcriptional and post-transcriptional regulation. Nat Biotechnol **33**, 722–729, doi:10.1038/nbt.3269 (2015).
39. Fong, Y. W. & Zhou, Q. Stimulatory effect of splicing factors on transcriptional elongation. Nature **414**, 929–933, doi:10.1038/414929a (2001).
40. Nojima, T. *et al.* Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. Cell **161**, 526–540, doi:10.1016/j.cell.2015.03.027 (2015).

41. Nojima, T. *et al.* RNA Polymerase II Phosphorylated on CTD Serine 5 Interacts with the Spliceosome during Co-transcriptional Splicing. *Mol Cell* **72**, 369–379 e364, doi:10.1016/j.molcel.2018.09.004 (2018).
42. Huang, H., Zhang, J., Harvey, S. E., Hu, X. & Cheng, C. RNA G-quadruplex secondary structure promotes alternative splicing via the RNA-binding protein hnRNPF. Genes Dev **31**, 2296–2309, doi:10.1101/gad.305862.117 (2017).
43. Herviou, P. *et al.* hnRNP H/F drive RNA G-quadruplex-mediated translation linked to genomic instability and therapy resistance in glioblastoma. Nat Commun **11**, 2661, doi:10.1038/s41467-020-16168-x (2020).
44. Dominguez, C., Fisette, J.-F., Chabot, B. & Allain, F. H. T. Structural basis of G-tract recognition and encaging by hnRNP F quasi-RRMs. Nature Structural & Molecular Biology **17**, 853–861, doi:10.1038/nsmb.1814 (2010).
45. Dominguez, C. & Allain, F. H.-T. NMR structure of the three quasi RNA recognition motifs (qRRMs) of human hnRNP F and interaction studies with Bcl-x G-tract RNA: a novel mode of RNA recognition. Nucleic Acids Research **34**, 3634–3645, doi:10.1093/nar/gkl488 (2006).
46. Samatanga, B., Dominguez, C., Jelesarov, I. & Allain, F. H. The high kinetic stability of a G-quadruplex limits hnRNP F qRRM3 binding to G-tract RNA. Nucleic Acids Res **41**, 2505–2516, doi:10.1093/nar/gks1289 (2013).
47. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102**, 15545–15550, doi:10.1073/pnas.0506580102 (2005).
48. Kröger, C. *et al.* Acquisition of a hybrid E/M state is essential for tumorigenicity of basal breast cancer cells. Proc Natl Acad Sci U S A **116**, 7353–7362, doi:10.1073/pnas.1812876116 (2019).
49. Nagy, E. & Rigby, W. F. Glyceraldehyde-3-phosphate dehydrogenase selectively binds AU-rich RNA in the NAD(+)-binding region (Rossmann fold). J Biol Chem **270**, 2755–2763, doi:10.1074/jbc.270.6.2755 (1995).
50. Hentze, M. W. *et al.* Identification of the iron-responsive element for the translational regulation of human ferritin mRNA. Science **238**, 1570–1573, doi:10.1126/science.3685996 (1987).
51. Robinson, J. B., Jr. & Srere, P. A. Organization of Krebs tricarboxylic acid cycle enzymes in mitochondria. J Biol Chem **260**, 10800–10805 (1985).
52. Long, Y. *et al.* RNA is essential for PRC2 chromatin occupancy and function in human pluripotent stem cells. Nat Genet **52**, 931–938, doi:10.1038/s41588-020-0662-x (2020).
53. Damasceno, L. E. A. *et al.* PKM2 promotes Th17 cell differentiation and autoimmune inflammation by fine-tuning STAT3 activation. J Exp Med **217**, doi:10.1084/jem.20190613 (2020).
54. Jonkers, I. & Lis, J. T. Getting up to speed with transcription elongation by RNA polymerase II. Nat Rev Mol Cell Biol **16**, 167–177, doi:10.1038/nrm3953 (2015).
55. Zhang, Y. *et al.* G-quadruplex structures trigger RNA phase separation. Nucleic Acids Research **47**, 11746–11754, doi:10.1093/nar/gkz978 (2019).

56. Guo, J. U. & Bartel, D. P. RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. Science **353**, aaf5371, doi:10.1126/science.aaf5371 (2016).
57. Wolfe, A. L. *et al.* RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. Nature **513**, 65–70, doi:10.1038/nature13485 (2014).
58. Lee, D. S. M., Ghanem, L. R. & Barash, Y. Integrative analysis reveals RNA G-quadruplexes in UTRs are selectively constrained and enriched for functional associations. Nature Communications **11**, 527, doi:10.1038/s41467-020-14404-y (2020).
59. Zhang, Y. *et al.* G-quadruplex structures trigger RNA phase separation. Nucleic Acids Res **47**, 11746–11754, doi:10.1093/nar/gkz978 (2019).
60. Keene, J. D. RNA regulons: coordination of post-transcriptional events. Nat Rev Genet **8**, 533–543, doi:10.1038/nrg2111 (2007).
61. Mansfield, K. D. & Keene, J. D. The ribonome: a dominant force in co-ordinating gene expression. Biol Cell **101**, 169–181, doi:10.1042/BC20080055 (2009).
62. Hua, Y., Vickers, T. A., Okunola, H. L., Bennett, C. F. & Krainer, A. R. Antisense masking of an hnRNP A1/A2 intronic splicing silencer corrects SMN2 splicing in transgenic mice. Am J Hum Genet **82**, 834–848, doi:10.1016/j.ajhg.2008.01.014 (2008).
63. Landthaler, M. *et al.* Molecular characterization of human Argonaute-containing ribonucleoprotein complexes and their bound target mRNAs. RNA **14**, 2580–2596, doi:10.1261/rna.1351608 (2008).
64. Lajoie, M. J. *et al.* Genomically recoded organisms expand biological functions. Science **342**, 357–360, doi:10.1126/science.1241459 (2013).
65. Pirman, N. L. *et al.* A flexible codon in genomically recoded Escherichia coli permits programmable protein phosphorylation. Nat Commun **6**, 8130, doi:10.1038/ncomms9130 (2015).
66. Gassaway, B. M. *et al.* Distinct Hepatic PKA and CDK Signaling Pathways Control Activity-Independent Pyruvate Kinase Phosphorylation and Hepatic Glucose Production. *Cell Rep* **29**, 3394–3404 e3399, doi:10.1016/j.celrep.2019.11.009 (2019).
67. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. Bioinformatics **29**, 15–21, doi:10.1093/bioinformatics/bts635 (2013).
68. Trapnell, C. *et al.* Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol **31**, 46–53, doi:10.1038/nbt.2450 (2013).
69. Ge, S. X., Jung, D. & Yao, R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics **36**, 2628–2629, doi:10.1093/bioinformatics/btz931 (2020).

Methods

Cell lines and plasmids

HEK293 T-Rex Flp-In cells (Thermo Fisher Scientific, Cat# R78007) were cultured in DMEM medium (Gibco) supplemented with 10% (v/v) fetal bovine serum (FBS), 100 µg/ml zeocin, and 10 µg/ml

blasticidin (Gibco). MCF-7 cells were obtained from ATCC (Cat# HTB-22) and cultured in the recommended EMEM medium (ATCC) supplemented with 10% (v/v) FBS and 0.01 mg/ml bovine insulin (Sigma). MDA-MB-231 cells were obtained from ATCC (Cat# HTB-26) and cultured in DMEM-high glucose (Gibco) supplemented with 10% (v/v) FBS. pFRT/TO-FLAG/HA-PKM2 and pFRT/TO-FLAG/HA-PKM2-NLS and -NES plasmids were generated as described previously [63]. Plasmids were transfected in cells together with pOG44 plasmid expressing Flp-recombinase (Invitrogen) using Lipofectamine 2000 transfection reagent (Invitrogen) according to manufacturer's instructions to create stable HEK293 T-Rex Flp-In cells that conditionally express (doxycycline inducible) either wild-type Flag-HA-tagged PKM2 or Flag-HA-tagged PKM2-NES/-NLS fusions. Stable cell lines were maintained under selection using 10 µg/ml blasticidin and 50 µg/ml hygromycin B (Gibco).

Knockdown of PKM2 in MCF-7 cell lines was performed by lentiviral transduction. pLKO.1 (Addgene #84530) and 3rd generation lentiviral vectors were obtained from Addgene (#12251, #12253, #12259). shRNAs targeting the PKM 3' UTR and control scramble shRNA were designed and cloned into pLKO.1 plasmid. Lentiviral and expression plasmids were co-transfected (1:1 ratio) into HEK293T/17 cells (ATCC) using Lipofectamine 3000 transfection reagent (Invitrogen) according to manufacturer's instructions. The viral containing supernatant were collected 24 and 48 hours after transfection, filtered, and used to infect MCF-7 cells in the presence of 10 µg/ml polybrene. Infected cells were cultured for 9 days with 2 µg/ml puromycin (Gibco). Silencing of endogenous PKM in cells expressing PKM2-NES was performed using siRNA targeting 3'UTR of PKM (IDT hs.Ri.PKM.13.3) or scramble control (IDT, 51-01-14-03) using Lipofectamine™ RNAiMAX Transfection Reagent.

shPKM DNA oligomers:

5'-CCGGCAACGCTTGTAGAACTCACTCCTCGAGGAGTGAGTTCTACAAGCGTTGTTTTTG-3'

5'-AATTCAAAAACAACGCTTGTAGAACTCACTCCTCGAGGAGTGAGTTCTACAAGCGTTG-3'

Scramble-shRNA DNA oligomers

5'-CCGGCCTAAGGTTAAGTCGCCCTCGCTCGAGCGAGGGCGACTTAACCTTAGGTTTTTG-3'

5'-AATTCAAAAACCTAAGGTTAAGTCGCCCTCGCTCGAGCGAGGGCGACTTAACCTTAGG-3'

PAR-CLIP

Fluorescent Photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (fPAR-CLIP) for nuclear and cytoplasmic PKM2 was performed as described in [25] with the following modifications. Briefly, 10 × 15 cm plates at 70% confluency cells conditionally expressing the wild-type PKM2 were induced with doxycycline for 24 hours followed by 100 µM 4-thiouridine treatment for 12 hours and UV crosslinked at 365 nm (300 mJ/cm^2). The cytoplasm was extracted by resuspending cells in a modified hypotonic lysis buffer (10 mM Tris-HCl pH 7.5, 50 mM NaCl, 3 mM $MgCl_2$, 0.1% NP-40, 10% glycerol) for 2 min. The nuclear pellet was washed twice with the same buffer and lysed in NP-40 buffer,

followed by sonication. Proteins were immunoprecipitated using the anti-Flag M2 magnetic beads (Sigma). After FLAG-IP, the RNPs were dephosphorylated, phosphorylated and a 5' fluorescent DNA adapter modified with a 5' IRDye® 800CW and 4 degenerate nucleotides that serve as unique molecular identifiers at the 3' end (IR800-GTTCAGAGTTCTACAGTCCGACGATCrNrNrNrN). After SDS-PAGE analysis the fluorescent bands corresponding to the RNP-adapter were isolated, shredded, and treated with proteinase K as described in [25]. 3' adapter ligation was then performed on the recovered RNA footprint. Reverse transcription, PCR amplification, and library purification were performed as described in [25]. HNRNPF fPAR-CLIP was performed as described in [25] follow supplementary protocol for nuclear RNPs(Total nuclear fraction).

Recombinant protein expression and purification

To express recombinant PKM1 and PKM2 proteins, pET-28a-hPKM1 (Addgene #44241) and pET-28a-hPKM2 (Addgene #44242) were transformed into *E. coli* BL21(DE3) cells. 0.4 l of LB cultures were induced with 0.5 mM IPTG at an OD600 of 0.5 and incubated for additional 5-7 hours at RT. Cell pellets were resuspended in 8 ml lysis buffer (300 mM NaCl 50mM Tris-HCl pH 7.5 and 1% NP40, 20 mM imidazole, 1 mM DTT, and protease inhibitor cocktail (cOmplete, Roche). 0.5 ml of HisPur™ Ni-NTA Magnetic Beads (Thermo Fisher Scientific), equilibrated in the same buffer, were incubated with the lysate for 1 hour at 4 ºC under rotation and washed 3 times with 10 ml wash buffer (600 mM NaCl, 50mM Tris-HCl pH 7.5, 1% NP40, 30 mM Imidazole). Proteins were eluted with 1.5 ml elution buffer (300 mM NaCl, 50 mM Tris-HCl pH=7.5, 1% NP40 and 180 mM imidazole). Purified proteins were dialyzed into storage buffer (50 mM Tris-HCl pH 7.5, 300 mM NaCl) using Slide-A-Lyzer™ Dialysis Cassette, 10K MWCO, 3 ml (Thermo Fisher Scientific, Cat# 66380) and concentrated using Protein Concentrators PES, 10K MWCO, 0.5 ml (Thermo Fisher Scientific, Cat# 88513).

Phosphoprotein synthesis and purification

Plasmids and strains. DNA sequences encoding the full-length N-term 6xHis-tagged Human PKM2 including a S37STOP(TAG) mutation was assembled into pCRT7/NT-TOPO bacterial expression plasmid (Addgene #73446) at NdeI/SacI cloning sites respectively. pCRT7/NT-TOPO (Addgene #68306) was transformed into genomically recoded *E. coli* (C321.ΔA) cells along with SepOTSλ plasmid for phosphoprotein production (Addgene #68292) [64,65].

Expression and purification. Expression and purification of the full-length PKM2pS37 protein was performed as described before with modifications [66]. In brief, expression was induced with 0.2% arabinose at an OD600 of 0.8-0.9 for 20 hours at 30 $^{\circ}$C. Bacterial cells were lysed in 20 mM Tris-HCl pH 7.2, 500 mM NaCl, 1 mM Tris (2-carboxyethyl) phosphine (TCEP), 50 mM NaF, 1 mM $NaVO_4$ buffer supplemented with 1 mg/ml lysozyme (Sigma), and protease inhibitor (cOmplete, Roche). Recombinant protein was purified on a HisTrap HP column (GE Healthcare) using continuous imidazole gradient (10-500 mM), followed by size exclusion chromatography on a Superdex 200 10/300 GL (GE Healthcare) with 20 mM Tris-HCl pH 7.2, 100 mM NaCl buffer containing 0.5 mM DTT and 5% glycerol. Purified

proteins were buffer exchanged into 50 mM Tris-HCl pH 7.4, 150 mM NaCl, and 20% glycerol for storage at -20 °C.

Immunofluorescence analysis

7 to 8 x 10^4 cells were seeded in ibiTreat chamber slides (Ibidi). After 20-24 hours, cells were fixed with 4% PFA for 15 min at room temperature, washed three times with 1x DPBS, permeabilized with 0.1% Triton X-100 for 15 min and blocked with 10% goat serum (constituted in 0.1% Triton X-100) at room temperature. Cells were stained overnight at 4 °C with primary antibodies against PKM2pS37 (rabbit polyclonal, custom-made) and total PKM2 (Cell Signaling Technology, Cat# 4053) a 1:200 and 1:100 dilution in blocking solution respectively, or FLAG antibody (Sigma-Aldrich, Cat# F3165) (1:200 dilution) followed by incubation with secondary antibodies (1:300 dilution) conjugated with Alexa Fluor 488 (Thermo Fisher Scientific, Cat# A-11034) or Alexa Fluor 594 (Thermo Fisher Scientific, Cat# A-11032) and Hoechst 33342 (Invitrogen) labeling. All samples were imaged by confocal microscopy on a Leica Laser Scanning SP8 Microscope (20x or 63x oil objective) at room temperature using the same settings. Images were analyzed using ImageJ software (Fiji, RRID: SCR_002285). Fluorescence intensity was determined using manual masking and identical brightness and contrast parameters.

Electromobility Shift Assay (EMSA)

For Electromobility Shift Assays (EMSA) 6xHis-tagged proteins were incubated in increasing concentrations with 15 nM of Cy-5 fluorescent oligonucleotide in binding buffer (25 mM Tris-HCl pH 7.5, 100 mM KCl, 3 mM $MgCl_2$, 0.01% tween, 1 mg/ml BSA, 1 mM DTT) adjusted with water to a final volume of 40 µl. After incubation for 30 min at room temperature 10 µl of 50% glycerol was added and reactions were separated on a 0.6 % agarose gel (running buffer 1x TBE, 25 mM KCl) at 80 V for 90 min at room temperature.

Bind-n-Seq Assay

Bind-n-Seq analysis was performed as described previously with a few modifications [28]. 5 µg of 6xHis-tagged PKM2 protein was bound to 25 µl Ni-Charged MagBeads (GenScript), washed, and incubated with 1 µM synthetic, completely randomized 29-nt long oligoribonucleotides (IDT) in binding buffer (25 mM Tris-HCl pH 7.5, 100 mM KCl, 3 mM $MgCl_2$, 0.01% Tween, 1 mg/ml BSA, 1 mM DTT) for 30 min at room temperature. Beads were washed once with binding buffer and RNA and proteins eluted with 10 mM Tris pH 7.0, 1 mM EDTA, 1% SDS. Eluted RNA was purified using the Oligo Clean & Concentrator kit according to manufacturer's instructions (Zymo Research) and a 3' adapter was ligated using T4 RNA Ligase 2, truncated (NEB) following manufacturer's instructions. The reaction was cleaned and a 5' adapter was ligated using T4 RNA Ligase (Ambion, Thermo Fisher Scientific), followed by reverse transcription using SuperScript IV Reverse Transcriptase and RT primer (GCCTTGGCACCCGAGAATTCCA) and PCR amplified using Platinum™ Taq DNA Polymerase and Illumina RNA PCR primers.

RNA Sequencing

RNA from cells was isolated using the Direct-zol RNA Miniprep Kit (Zymo Research, Cat# R2050) according to the manufacturer's instructions. 1 µg total RNA was Ribosomal RNA depleted using the NEBNext® rRNA Depletion Kit and cDNA libraries were prepared using the NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina® (NEB). cDNA libraries were sequenced on the Illumina HiSeq 3000 HiSeq X or NovaSeq 6000 platform. Reads were aligned to human genome version hg38 using STAR (star/2.7.2b) [67]. Cufflinks was used for differential expression [68]. All gene set files were obtained from GSEA website (www.broadinstitute.org/gsea). GSEA was performed using the GSEA v4.1.0 software. Gene ontology enrichment analysis was performed using the ShinyGO v0.66 tool [69].

ChIP-Seq

ChIP-seq was performed following abcam protocols. 2x 15cm plates of HEK293 T-Rex FH-PKM2-NLS (Treated with Doxycycline or DMSO for 72 hours) were cross-linked with 0.75% (Alfa Aesar 43368) for 10 min followed by 5 min treatment with 125 mM Glycine. Cells were washed and collected with PBS and lysed with 0.5 ml of lysis buffer (50 mM HEPES-KOH pH7.5, 140 mM NaCl, 1 mM EDTA pH8, 1% Triton X-100, 0.1% Sodium Deoxycholate, 0.1% SDS and protease inhibitor cocktail (cOmplete, Roche)). After sonication using a bath sonicator (Bioruptor, diagenode) for 14 min (medium amplitude, 30 s on 30 s off) at 4 ºC. 0.25 ml of lysate was diluted in RIPA buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 2 mM EDTA pH 8, 1% NP-40, 0.5% Sodium Deoxycholate, 0.1% SDS, protease inhibitor cocktail) to a volume of 2 ml and IP was performed using 5 µg of anti-RNA polymerase II CTD repeat YSPTSPS antibody (Abcam, ab26721) or Anti-RNA polymerase II CTD repeat YSPTSPS (phospho S2) (Abcam, ab5095) for 1 hour at 4 $^{\circ}$C. 30 µl of Protein G Dynabeads were equilibrated in the same buffer and added to the samples for 16 hours at 4 $^{\circ}$C under rotation. Beads were washed once with the following buffers. Low Salt Wash Buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.0, 150 mM NaCl), High Salt Wash Buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.0, 500 mM NaCl), LiCl Wash Buffer (0.25 M LiCl, 1% NP-40, 1% Sodium Deoxycholate, 1 mM EDTA, 10 mM Tris-HCl pH 8.0). Elution was performed with 120 µl elution buffer at 30 $^{\circ}$C for 15 min with elution buffer (1% SDS, 100mM $NaHCO_3$). 4.8 µL of 5 M NaCl and 2 µL RNase A/T1 (Thermo EN0551) were added to the eluent and incubated while shaking at 65°C overnight. 2 µL proteinase K (20 mg/mL) and incubate while shaking at 60°C for 1 h. Phenol/chloroform extraction was performed and the sample was cleaned and concentrated using the DNA Clean & Concentrator-5 kit Zymo Research, D013). Library for illumine sequencing was prepared using the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina.

Matrigel invasion assays

Invasion assays were carried out as described previously [13]

Kinase activity assays

Pyruvate kinase reactions was performed under the following conditions: 30 mM Tris pH 7.4, 10 mM $MgCl_2$, 0.6 mM PEP, 1.5 mM ADP (final concentrations) with or without the presence of F1,6BP (70 uM).

Reactions were assembled to final volume of 100 ul with the addition of Kinase-Glo Reagent (Promega) according to manufacturer's instructions. The reactions were then initiated by adding the enzyme alone or in complex with equal amount (22 nM) of folded rG4 oligonucleotide (sequence derived from SLC7A5 3'UTR G4). ATP production was followed by measuring luminescence over time on a SpectraMax i3x plate reader (Molecular Devices). Measurements were converted to enzymatic units (umol/min) and specific activities were calculated by normalizing with protein mg.

Figures

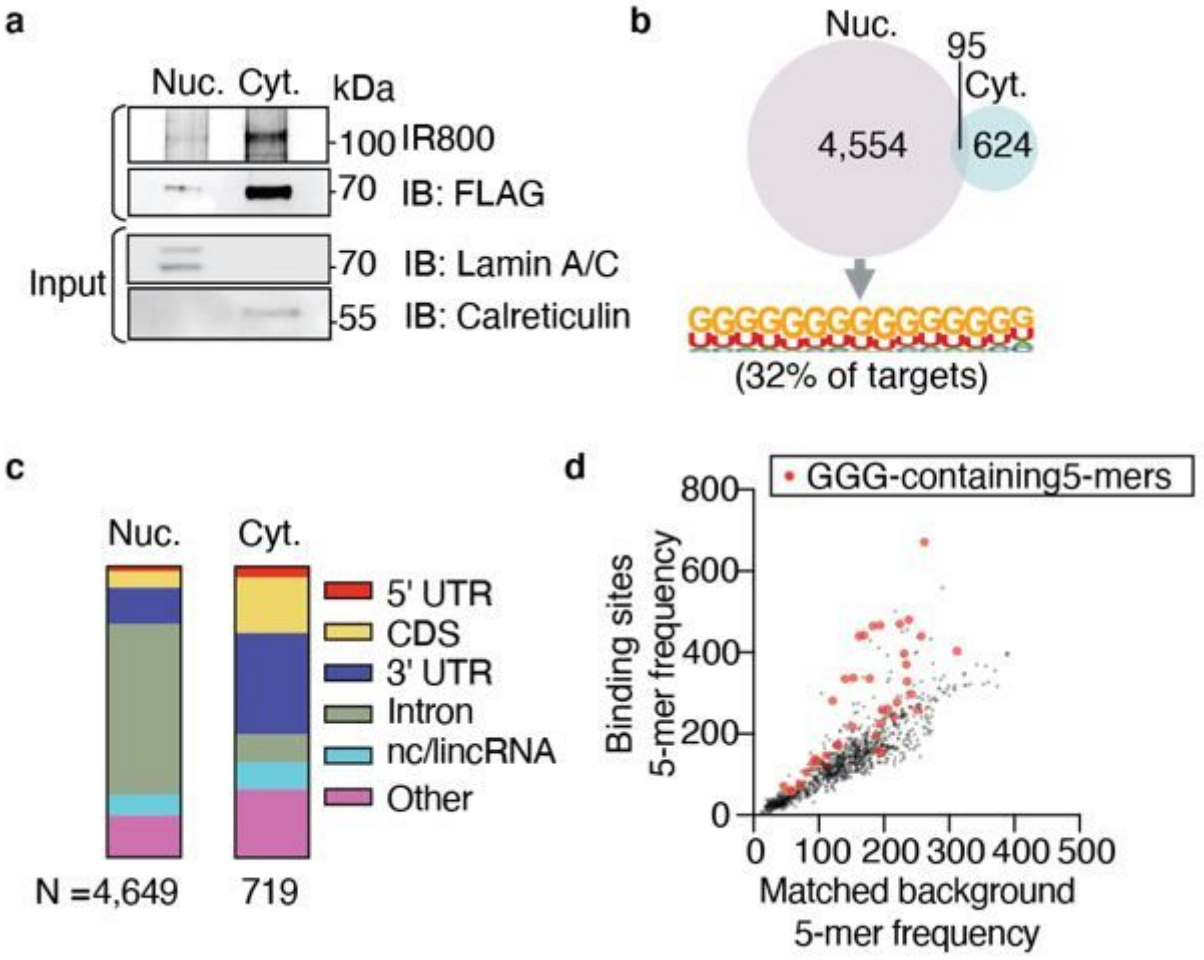

Figure 1

Nuclear PKM2 binds G-rich sequences in precursor mRNAs | a, IR800 fluorescent image of crosslinked, ribonuclease-treated and fluorescent adapter-ligated FH-PKM2-RNPs from nucleus (Nuc) and cytoplasm (Cyt) of FH-PKM2-HEK293 cells separated by SDS-PAGE. FLAG-immunoblot as a loading control. Biochemical fractionation of FH-PKM2-HEK293 cells is controlled by immunoblots for Calreticulin (cytoplasm) and LAMINA/C (nucleus). **b**, Venn diagram of nuclear and cytoplasmic PKM2 binding site overlap. (Lower panel) Highest probability sequence motif shared in nuclear FH-PKM2 PAR-CLIP binding sites generated by HOMER [27]. **c**, Distribution of PKM2 binding sites across different RNA annotation

categories in the cytoplasm or the nucleus. **d**, Scatter plot of the frequency of all possible 5-mers in PAR-CLIP binding sites versus matched background sequences.

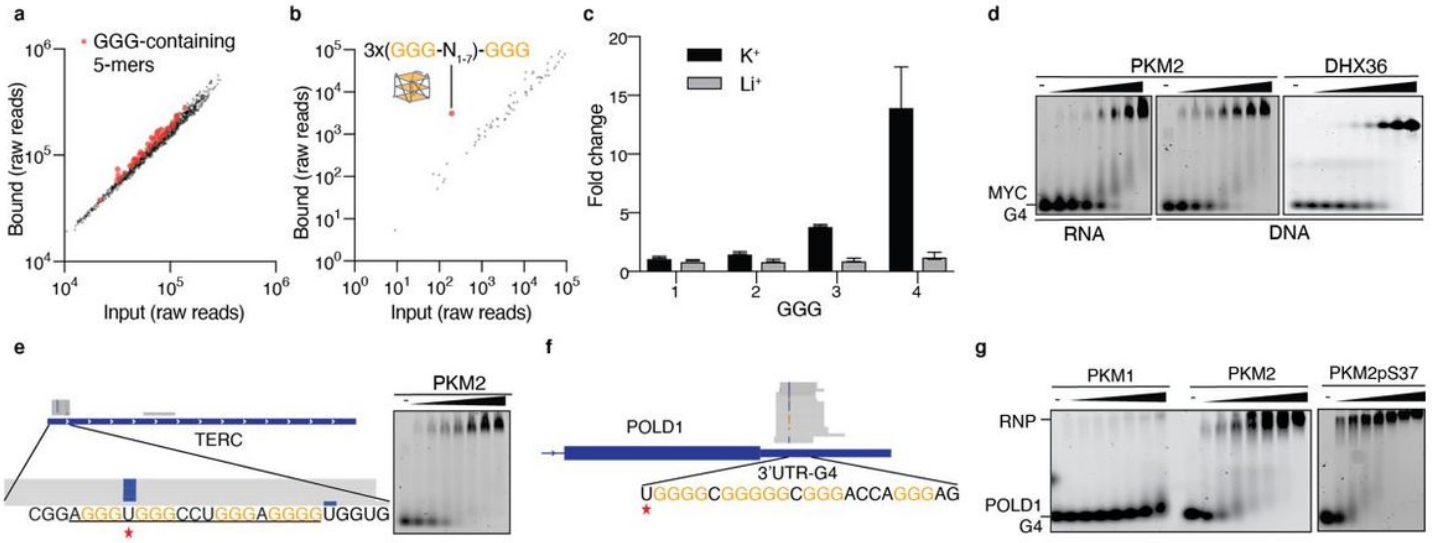

Figure 2

PKM2 binds folded rG4s *in vitro* | a, Scatter plot (left panel) of 5-mers enriched in RBNS using recombinant, His-tagged (HIS) PKM2 and a randomized 29-nt oligoribonucleotide library. **b**, Scatter plot showing the relative abundance of four repeats of any nucleotide triplet separated by any 1-7 nucleotides ($XYZ(N_{1-7})_3XYZ$) in input and PKM2-bound RBNS libraries. **c**, Enrichment (bound over input) of sequences containing 1-4 G-triplets in HIS-PKM2 RBNS performed in the presence of K^+ or Li^+. **d**, EMSA using 31.25-2,000 nM recombinant, HIS-PKM2 or 3.9-500 nM DHX36 with 7.5 nM of the G4 forming RNA or DNA 16-mer derived from the *MYC* promoter. **e**, Alignment of nuclear FH-PKM2 PAR-CLIP reads with the TERC rG4. The asterisk (*) indicates the main crosslinking site. (Left panel) EMSA using PKM2 rG4 sequence derived from TERC (component concentration as in (D)). **f**, Alignment of nuclear FH-PKM2 PAR-CLIP reads around rG4-forming sequence in the 3'UTR of POLD1 mRNA. **g**, EMSA using recombinant His-tagged PKM1, PKM2, or PKM2pS37 with oligoribonucleotide corresponding to the POLD1 rG4 (component concentration as in (**d**).

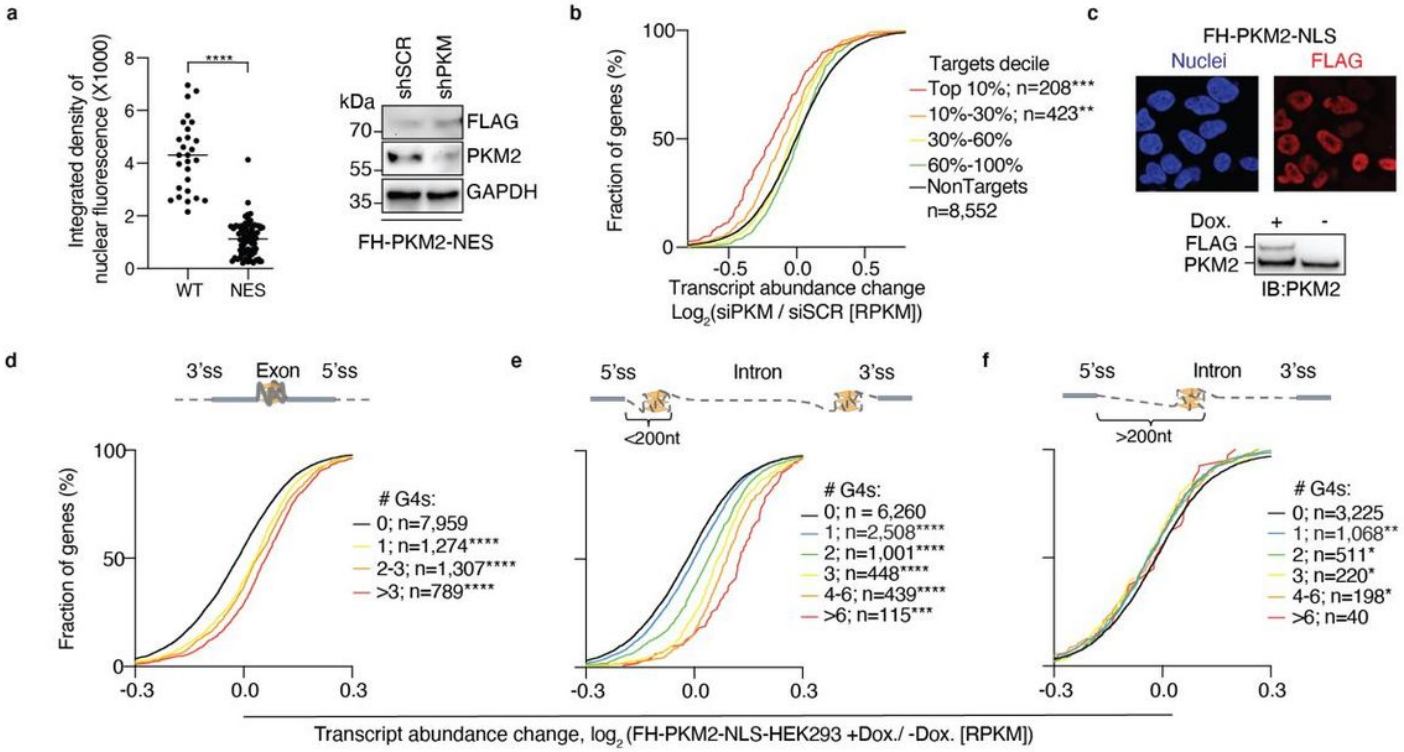

Figure 3

PKM2 regulates expression of rG4-containing mRNAs | a, (Left panel) Quantification of nuclear FLAG signal in cells expressing FH-PKM2 and FH-PKM2-NES. (Right panel) Immunoblot confirming silencing of endogenous PKM and expression of FH-PKM2-NES. **b**, mRNA expression changes after silencing of endogenous PKM2 in cells expressing FH-PKM2-NES. The empirical cumulative distribution function (CDF) of nuclear PKM2 targets binned by normalized crosslinked reads (colored lines) compared to expressed non-targets (black line). **c**, (Upper panel) Immunofluorescence analysis showing nuclear localization of FH-PKM2-NLS; Nuclei, Hoechst 33342 (blue); FLAG, Alexa Fluor 594 (red); Scale bar, 20 μm. (Lower panel) Immunoblot confirming expression of FH-PKM2-NLS after doxycycline induction. **d**, mRNA expression changes upon expression of FH-PKM2-NLS. Transcripts were binned according to the number of canonical rG4 motifs ($GGG(N_{1-7})_3$-GGG) in their exons. **e,f,** Same as (d), except transcripts were binned according to presence of intronic rG4 motifs within (e) or greater than 200 nt (f) of the 5' or 3' splice site. *P ≤ 0.05, **P ≤ 0.01, ***P ≤ 0.001, ****P ≤ 0.0001.

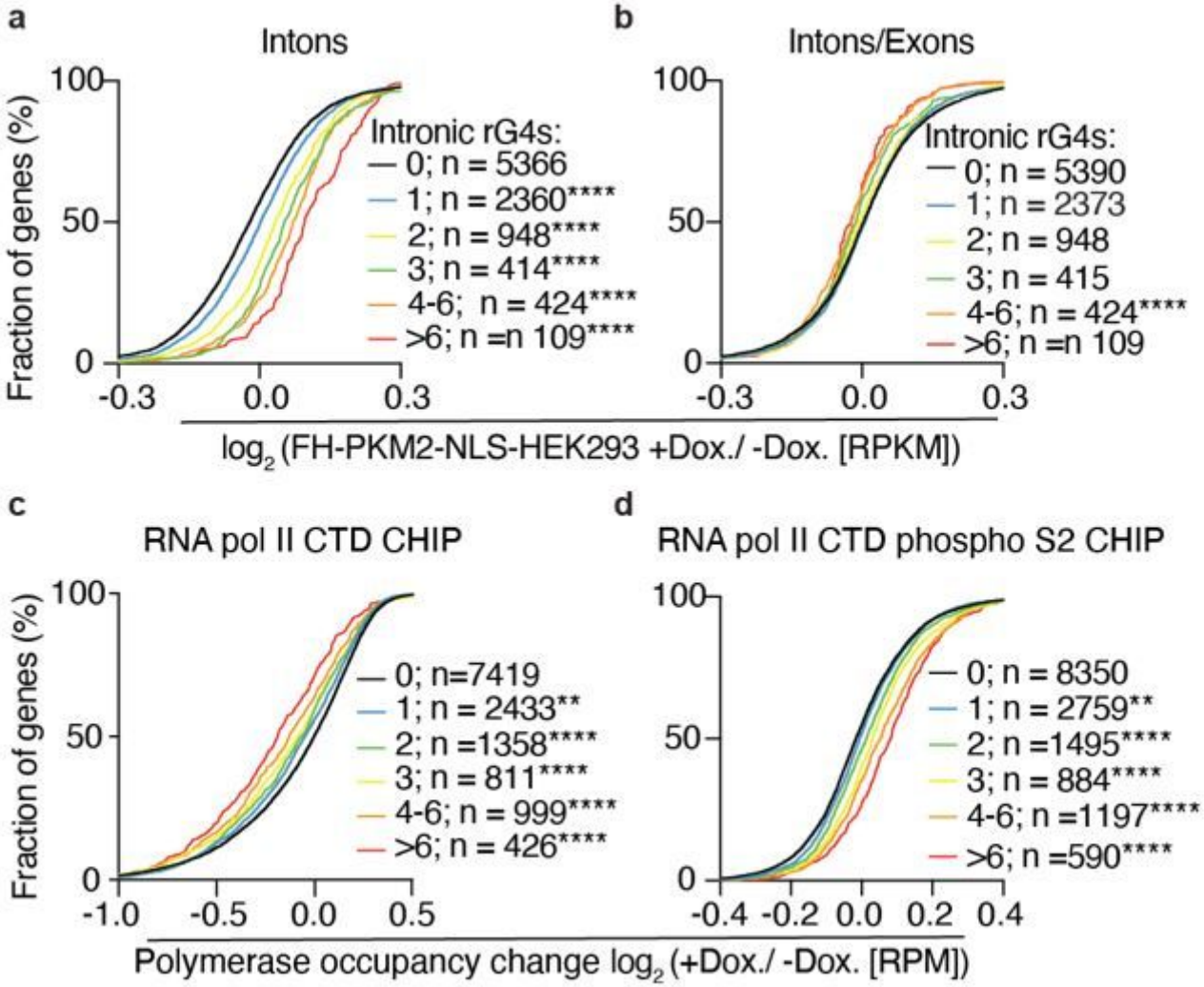

Figure 4

PKM2 co-transcriptionally regulates gene expression | a,b, Intron (a) or intron/exon (b) abundance changes upon expression of FH-PKM2-NLS. Transcripts were binned according to the number of canonical rG4 motifs ($GGG(N_{1-7})_3GGG$) in their exons or within 200 nt of the 5' or 3' splice site. **c,d,** Polymerase occupancy changes of Unphosphorylated PolII near transcription start site (a) or elongating PolII (phospho S2 CTD repeat YSPTSPS) along the gene body (b). Transcripts were binned according to the number of canonical rG4 motifs ($GGG(N_{1-7})_3GGG$) in their exons or within 200 nt of the 5' or 3' splice site. *P ≤ 0.05, **P ≤ 0.01, ***P ≤ 0.001, ****P ≤ 0.0001; Mann-Whitney U test (MWU).

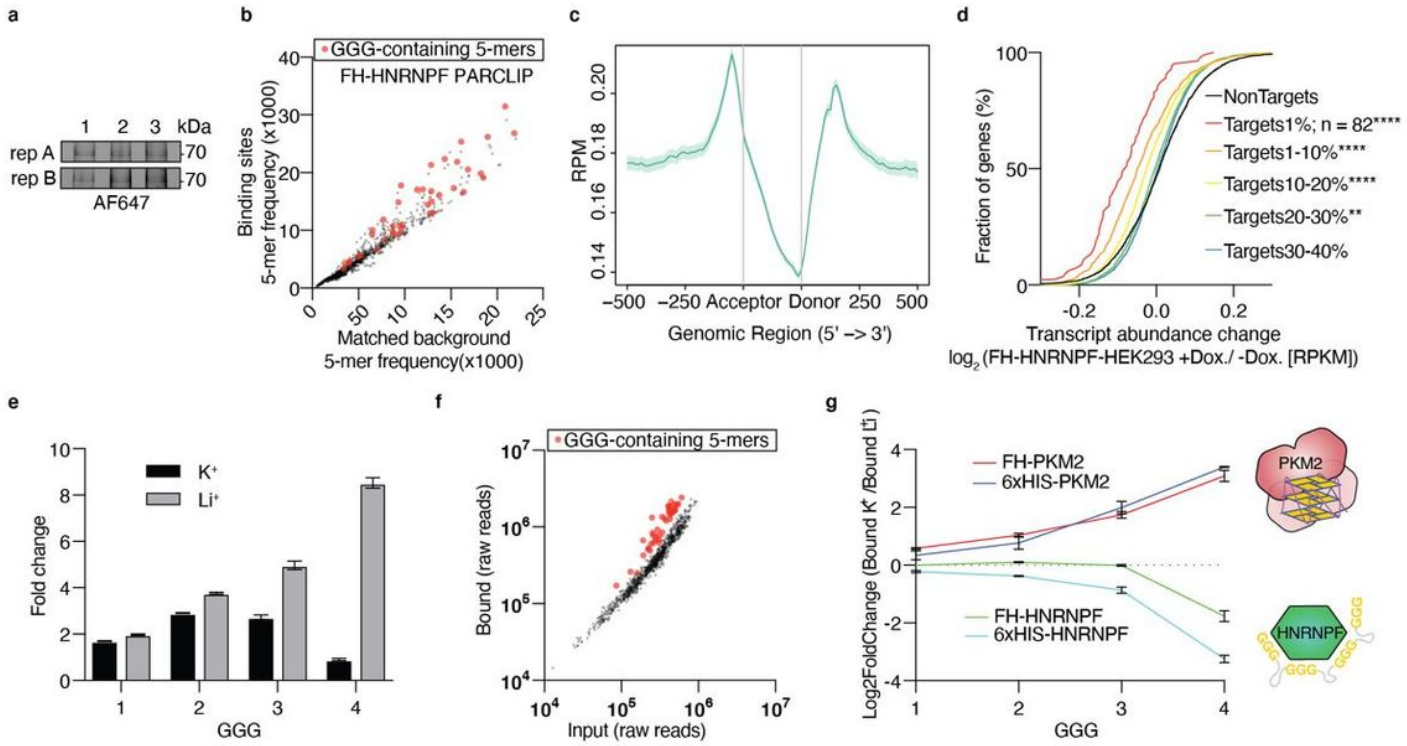

Figure 5

PKM2 competes with HNRNPF to stabilize rG4 structures | a, Fluorescent image of crosslinked, ribonuclease-treated and fluorescent adapter-ligated FH-HNRNPF-RNPs separated by SDS-PAGE. Lane 1: 0.1 µg/ml RNaseA plus 0.25 U/ml RNaseT1. Lane 2: 1 µg/ml RNaseA plus 2.5 U/ml RNaseT1. Lane 3: 50 U/ml RNaseT1. **b,** Scatter plot of the frequency of all possible 5-mers in HNRNPF PAR-CLIP binding sites versus matched background sequences **c**, Distribution of crosslinked HNRNPF footprints around 5' and 3' splice sites. **d**, mRNA expression changes upon expression of FH-HNRNPF. The empirical cumulative distribution function (CDF) of HNRNPF targets binned by normalized crosslinked reads (colored lines) compared to expressed non-targets (black line). *P ≤ 0.05, **P ≤ 0.01, ***P ≤ 0.001, ****P ≤ 0.0001. **e**, Enrichment (bound over input) of sequences containing 1-4 G-triplets in HIS-HNRNPF RBNS performed in the presence of K^+ or Li^+. **f**, Scatter plot of 5-mers enriched in RBNS using recombinant HIS-HNRNPF and a randomized 29-nt oligoribonucleotide library folded and bound in the presence of Li^+. **g**, PKM2 prefers binding folded rG4s, while HNRNPF binds the linear sequences. Log_2RBNS(K^+/Li^+) of sequences containing 1-4 GGG from RBNS of purified PKM2 and HNRNPF recombinantly expressed in HEK cells (FH tagged) or *E.coli* (HIS tagged).

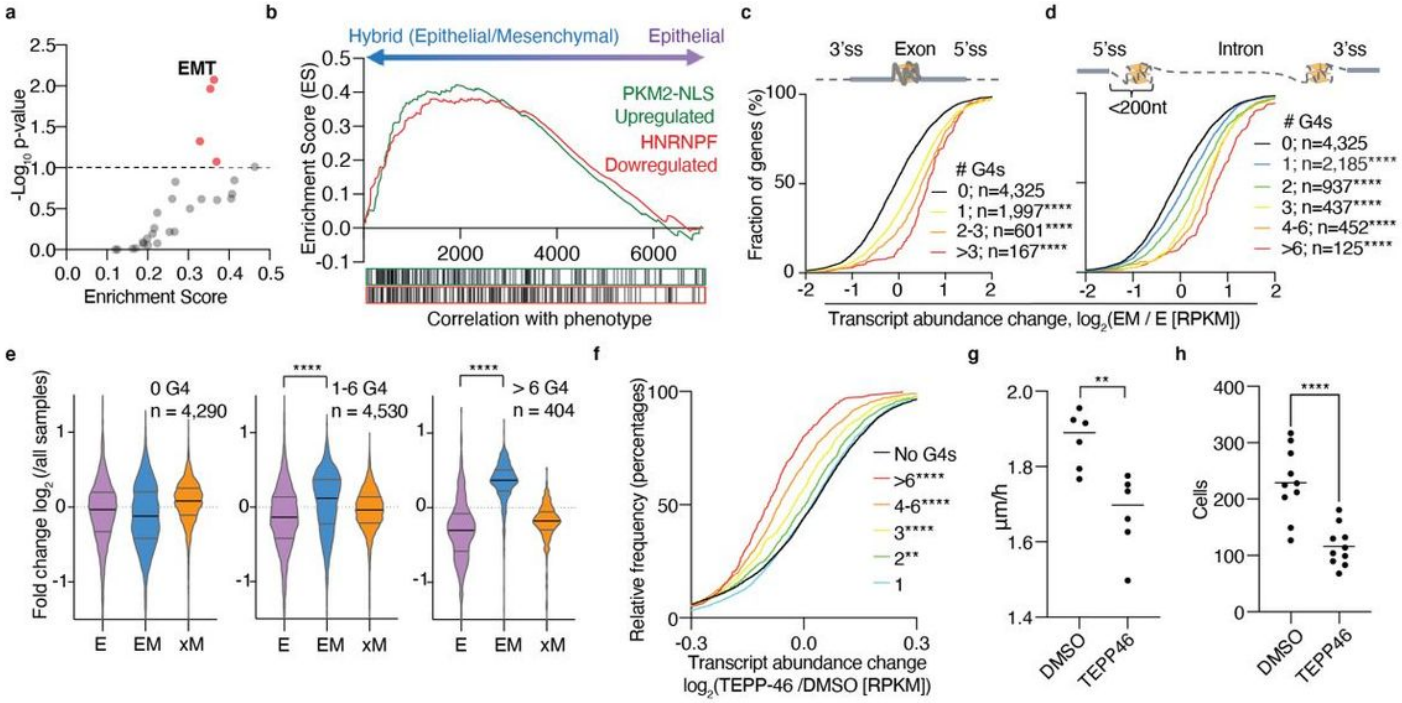

Figure 6

Nuclear PKM2 upregulates rG4-containing pre-mRNAs to promote an aggressive cancer cell phenotype | **a**, Scatter plot of hallmark gene sets enriched among genes upregulated by FH-PKM2-NLS expression (-$\log_{10}$ p-value versus Enrichment Score). **b**, Gene Set Enrichment Analysis of transcripts with a >25% increase upon expression of FH-PKM2-NLS in HEK293 compared to gene expression changes in HMLER cells as they transition from a less tumorigenic E-state to the highly aggressive hybrid E/M state. **c**, mRNA expression changes comparing cells in the aggressive E/M-state with the basic E-state. Transcripts were binned according to the number of canonical rG4 motifs ($GGG(N_{1-7})_3$-GGG) in their exons or **d**, in their introns within 200 nt of the splicing sites. **e**, Violin plots of the relative expression of transcripts with 0 (left panel), 1-6(center panel) or more than 6(right panel) rG4 motifs in their exons or introns near a splicing site (combined) for E, EM and xM cells. **f**, mRNA expression changes upon treatment of MDA-MB-231 cells with 5 µM TEPP-46. Transcripts were binned according to the number of canonical rG4 motifs ($GGG(N_{1-7})_3$-GGG) in their exons or within 200 nt of the 5' or 3' splice site. (**g**) Dot plot of cell front velocity in wound healing assay (**h**) Dot plot of the proportion of invading cells per tested area of Matrigel-coated transwell filters in invasion assays . *P ≤ 0.05, **P ≤ 0.01, ***P ≤ 0.001, ****P ≤ 0.0001

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ExtendedData.docx