

Mitochondrial DNA Copy Number Variation Across Human Cancers

Ed Reznik^{1,*}, Martin L. Miller², Yasin Senbabaoglu¹, Nadeem Riaz³, William Lee^{1,3}, Chris Sander¹

¹ Computational Biology Center, Sloan Kettering Institute, New York NY

² Cancer Research UK, Cambridge Institute, Cambridge CB2 0RE, UK

² Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York NY

* E-mail: reznike@mskcc.org

Abstract

In cancer, mitochondrial dysfunction, through mutations, deletions, and changes in copy number of mitochondrial DNA (mtDNA), contributes to the malignant transformation and progress of tumors. Here, we report the first large-scale survey of mtDNA copy number variation across 21 distinct solid tumor types, examining over 13,000 tissue samples profiled with next-generation sequencing methods. We find a tendency for cancers, especially of the bladder and kidney, to be significantly depleted of mtDNA, relative to matched normal tissue. We show that mtDNA copy number is correlated to the expression of mitochondrially-localized metabolic pathways, suggesting that mtDNA copy number variation reflect gross changes in mitochondrial metabolic activity. Finally, we identify a subset of tumor-type-specific somatic alterations, including IDH1 and NF1 mutations in gliomas, whose incidence is strongly correlated to mtDNA copy number. Our findings suggest that modulation of mtDNA copy number may play a role in the pathology of cancer.

1 Introduction

Human cells contain many copies of the 16 kilobase mitochondrial genome, which encodes thirteen essential components of the mitochondrial electron transport chain. Alterations of mitochondrial DNA (mtDNA), via inactivating genetic mutations or depletion of the number of copies of mtDNA in a cell, impair mitochondrial respiration and contribute to pathologies as diverse as encephelopathies and neuropathies [9], and the process of aging [1, 11]. In cancer, a number of studies have examined the role of mtDNA mutations in carcinogenesis [13, 16, 19, 41]. However, the contribution of changes in the gross number of mtDNA molecules in a tumor (*i.e.* the "mtDNA copy number") to tumor development and progression has not been adequately investigated.

In contrast to the fixed (diploid) copy number of the nuclear genome, many copies of mtDNA exist within each cell, and these levels can fluctuate based on energetic demand. Because mitochondria undergo a constant process of fusion and fission, it is difficult to meaningfully determine the number of mtDNA molecules per mitochondrion. Instead, studies have focused on measuring mtDNA copy number per cell, with estimates that vary between a few hundred and over one hundred thousand copies, depending on the tissue under examination [40]. Furthermore, because mtDNA serves as a template for the transcription of essential electron transport chain complexes, the quantity of mtDNA in a cell is a surrogate marker for the cell's capacity to conduct oxidative phosphorylation. For instance, a recent study estimated that energy-intensive tissues such as cardiac and skeletal muscle contained between 4000 and 6000 thousand copies of mtDNA per cell, while liver, kidney, and lung tissues averaged between 500 and 2000 copies [8].

Mitochondrial dysfunction plays several distinct roles in cancer [19, 34, 41]. First, the normal functions of mitochondria (*e.g.* respiration) may be subverted to support the growth of the tumor. A canonical example of this is the observation that many tumors suppress mitochondrial respiration in favor of increased uptake of glucose and secretion of lactate ("the Warburg effect"), a phenomenon which has found clinical utility for imaging of tumors using FDG-PET [39]. Second, mitochondria are susceptible to mutations in nuclear- and mitochondrially-encoded genes, and a subset of tumors are known to be caused by mutations of the mitochondrial enzymes FH, SDH, and IDH [18]. Furthermore, mtDNA dysfunction affecting the electron transport chain can lead to generation of excess reactive oxygen species (ROS), contributing to tumor cell metastasis [15].

To date, no comprehensive analysis of mtDNA copy number changes in tumors has been completed, despite a large literature of isolated reports [46]. Large-scale studies of mtDNA in cancer have instead focused on the analysis of mutations and heteroplasmy, largely ignoring the contribution of mtDNA copy number variation to the development and progression of tumors. Here, we use whole-genome and whole-exome sequencing data to examine changes in mtDNA copy number across a large panel of cancer types profiled by The Cancer Genome Atlas (TCGA) consortium. Using the resulting mtDNA copy number estimates, we ask fundamental questions about mtDNA and cancer. We investigate whether evidence of the Warburg effect can be found in patterns of mtDNA accumulation or depletion, which reflect a tumor’s capacity to respire. We examine the connection between gene expression levels and mtDNA copy number, and identify a subset of mitochondrially-localized metabolic pathways exhibiting a high degree of co-variation with mtDNA levels. Finally, we ask whether gross variations of mtDNA copy number are linked to the incidence of somatic alterations (including mutations and copy number alterations) across cancer types. Altogether, our results shed light on the contribution of aberrant mitochondrial function, through changes in mtDNA content, to cancer.

[Figure 1 about here.]

2 Calculation of mtDNA Abundance

To estimate the copy number of mtDNA in a tumor sample, we implemented a computationally efficient and fast approach based on comparing the number of sequencing reads aligning to (1) the mitochondrial (MT) genome and (2) the nuclear genome. Comparable approaches have been used to estimate somatic copy number alterations within the nuclear genome in cancer (for a review, see [47]). The approach assumes that regions of the genome of equal ploidy should be sequenced to comparable depth. In a normal human cell, the autosomal nuclear genome is at a fixed (diploid) copy number. Thus, by calculating the ratio of reads aligning to the mitochondrial and nuclear genomes, respectively, it is possible to estimate mtDNA ploidy relative to a diploid standard. This approach to assaying mtDNA copy number has been proposed and implemented by others in prior work [8, 12, 32].

To estimate mtDNA copy number, we calculated the ratio of (1) the number of sequencing reads mapping to the MT genome (r_m) to (2) the number of reads mapping to the nuclear genome (r_n) (Equation 1). Because tumor cells can exhibit large-scale amplifications and deletions, and be infiltrated by stromal and immune cells, we applied a ploidy/purity correction (“ R ”), described in detail in the Methods. This calculation yields the relative mtDNA copy number m . Assuming two samples have been processed in identical manners, then the sample with a higher value of m contains more copies of mtDNA [8, 12]. Because the approach relies simply on counting the number of reads aligning to each chromosome and does not calculate statistics at the individual-exon level, the running time was fast (on the order of a few minutes per sequencing sample) and the calculations easily parallelizable across the thousands of samples interrogated.

$$m = \frac{r_m}{r_n} \times R \quad (1)$$

We applied this method to whole exome sequencing (WXS) and whole genome sequencing (WGS) data from 21 distinct TCGA studies (Figure 2, see Methods for further details on data collection). To validate estimates of mtDNA copy number, we compared estimates from samples submitted to both WXS and WGS. Although mitochondrial reads are abundant in both WGS and WXS, the two sequencing methods capture mtDNA at different efficiencies: exome sequencing involves the targeted enrichment of exonic regions prior to sequencing and does not target mtDNA [31], while WGS sequences cellular DNA in an unbiased manner. If our approach to estimating mtDNA copy number is accurate, then we expect that

the two sequencing platforms should offer comparable estimates of mtDNA copy number across a panel of samples, *i.e.* samples with high mtDNA copy number in WGS should have similarly high mtDNA copy number in WXS. We compared mtDNA copy number estimates in 1174 samples across 9 tumor types profiled by both WXS and WGS, controlling for sequencing center due to previously reported biases in exome enrichment across different centers [16]. We confirmed that across all combinations of cancer types and sequencing centers, WXS and WGS offer significantly correlated estimates of mtDNA copy number (Table 1 and in the SI Figure S1).

[Figure 2 about here.]

3 Gross Changes in mtDNA Content are Evident in Many Cancers

Do tumors harbor different numbers of copies of mtDNA than normal tissue? This question is particularly intriguing because many cancers divert metabolic flux away from mitochondrial respiration and towards glycolysis in a phenomenon termed the Warburg effect. We investigated if tumor samples showed a significant depletion in mtDNA content, relative to matched normal tissues. To do so, for each pair of tumor/adjacent-normal samples collected from the same patient (1264 patients in total), we calculated the ratio r between the mtDNA copy number m in each of the samples:

$$r = \log_2 \left(\frac{m_T}{m_N} \right) \quad (2)$$

where m_T and m_N are the mtDNA abundances in tumor and normal tissues, respectively. We then used non-parametric Wilcoxon signed rank tests to assess whether each cancer type was significantly enriched for tumor samples with higher or lower mtDNA than matched normal tissue. The analysis was restricted to 14 cancer types for which we had at least 10 matched tumor/normal pairs. To ensure a meaningful comparison, we only used adjacent-normal tissue (and not blood) for the analysis. We elected to focus on analyzing whole-exome sequencing data, for which we had the largest number of samples. A complete list of all results is available in Supplementary Table 2.

Strikingly, eight of the fourteen tumor types analyzed showed a statistically significant (Bonferonni-corrected Mann-Whitney p-value < 0.05) decrease in mtDNA abundance in tumor samples (Figure 3A). These “mtDNA-depleted” tumor types included bladder, breast, esophageal, head and neck squamous, kidney (both clear-cell and papillary subtypes), liver, and stomach cancers. Despite a tendency towards mtDNA depletion, all tumor types contained at least one sample with higher mtDNA content than adjacent normal tissue. Nevertheless, the depletion effect was exceptionally strong in several tumor types: over 90% of bladder tumor samples and 88% of clear-cell kidney tumor samples contained less mtDNA than their normal tissue counterparts. In contrast, a single tumor type, lung adenocarcinoma, showed statistically significant mtDNA accumulation. In cases where sufficient numbers of both WGS and WXS data were available (bladder, breast, head and neck, clear-cell kidney, and lung adenocarcinomas), we observed a consistent effect across samples processed by both platforms (Figure S2).

[Figure 3 about here.]

To further relate changes in mtDNA content to clinical progression of disease, we used Cox regression to determine if tumor mtDNA copy number was a significant predictor of patient survival (Figure 4). In total, 5 cancer types showed statistically significant association between patient survival and mtDNA content. In three cancer types (ACC, p-value 0.017; KICH, p-value 0.058; and LGG, p-value 0.007), high tumor mtDNA content was associated with better survival. The opposite trend, of poor survival in

patients with high tumor mtDNA, was observed in SKCM (p-value 0.038), and UCEC (p-value 0.004). The finding regarding KICH is particularly intriguing given the central role mitochondrial dysfunction has been proposed to play in the disease [7]. That mtDNA copy number correlates with better or worse survival, depending on cancer type, suggests that other confounding factors strongly tied to survival, such as the presence of somatic mutations, may influence mtDNA levels. In a later section, we will investigate this hypothesis.

[Figure 4 about here.]

4 mtDNA is Correlated to the Expression of Mitochondrial Metabolic Genes

Proteins encoded in mtDNA localize exclusively to the mitochondrial electron transport chain, and fluctuations in mtDNA copy number are well-known to influence the level of transcription of these genes. It has also been observed that complete depletion of mtDNA in cell lines by exposure to ethidium bromide affects a number of additional signaling pathways [2]. Thus, we were compelled to ask if changes in mtDNA content narrowly influenced changes in the expression of oxidative phosphorylation genes, or if they were more broadly connected to the other functions of mitochondria.

Our approach to this question was to search for gene sets whose transcriptional signatures were highly correlated to mtDNA copy number. To do so, we calculated the non-parametric Spearman correlation between each gene and mtDNA copy number, and then used the mean-rank gene set test implemented in limma [20] to identify gene sets which were significantly enriched for highly correlated genes. The approach was applied in an unbiased manner to all 674 Reactome gene sets in the Canonical Pathways group from the MSigDB database [24].

In general, each tissue exhibited specific gene sets which were strongly correlated to mtDNA copy number levels. However, when aggregating across all cancer types, mitochondrially-localized metabolic pathways showed the highest correlation with mtDNA abundance (Figure 5). The recurrent positive correlation between expression of mitochondrial genes and mtDNA copy number across many tumor types served as a second, independent validation that estimates of mtDNA copy number reflected *in vivo* mtDNA ploidy.

In line with expectation, we found that the “Respiratory Electron Transport” gene set was the most correlated to mtDNA copy number (1st out of 674 gene sets, Supplementary Table 3, Worksheet Fig5Data). Among the remaining top positively correlated gene sets, all were metabolism-related, and included the TCA cycle, mitochondrial beta oxidation of fatty acids, and branched chain amino acid (BCAA) catabolism. BCAAs (valine, leucine, isoleucine) are essential amino acids whose catabolism depends on the activity of an enzyme complex, branched chain *alpha*-keto acid dehydrogenase, located in the inner mitochondrial membrane [14], and prior work has demonstrated that dietary supplementation of BCAA’s promotes mitochondrial biogenesis [6, 37]. Furthermore, a recent study has shown that elevated plasma levels of BCAAs are found 2 to 5 years before a cohort of patients developed pancreatic ductal adenocarcinoma [26].

A number of gene sets also showed recurrent negative correlation to mtDNA copy number, including those related to mRNA splicing, and S-phase (2nd and 5th most anticorrelated out of 674 gene sets, Supplementary Table 3). Several of these gene sets are associated with known non-metabolic functions of mitochondria in the cell. In particular, the replication of mitochondria and mtDNA is intimately linked to the cell cycle [3], and the nucleotide precursors to mtDNA are in part produced *de novo*, via a pathway that is only active during the S phase of the cell cycle [35]. Interestingly, we did not observe a strong correlation or anti-correlation between mtDNA copy number and glycolysis, suggesting that the transcriptional regulation of respiration and glycolysis may be independently coordinated.

A subset of tumor types did not show strong positive correlation between mtDNA copy number and expression of mitochondrial metabolic genes. In some cases, this was the result of an apparently dominant correlation with another pathway. For example, in UCEC, there is an exceptionally strong correlation between expression of cell cycle genes and mtDNA copy number. Perhaps most interestingly, in both prostate tumors and adjacent normal tissue, the expression of mitochondrial respiratory genes was the most strongly anti-correlated gene set (see SI Table 3). We speculate that this effect may be associated with the unique mitochondrial metabolism of prostate epithelia, which secrete large amounts of citrate generated in the mitochondria, rather than oxidizing it further and using the resulting NADH in the respiratory electron transport chain [4, 5].

[Figure 5 about here.]

5 Association with Mutations and Copy Number Alterations

The landscape of genetic events driving tumors is diverse, and the presence and activity of these genetic lesions is now being used as the basis for the staging of tumors, design of clinical trials, and development of new treatments [30]. We sought to understand whether mtDNA abundance was associated with the incidence of particular mutations/copy number alterations (CNAs) in patient samples. To do so, we evaluated whether patients harboring a particular genetic lesion showed statistically significant increases or decreases in tumor mtDNA abundance. We restricted analysis to whole-exome sequencing data, using tumor types which were sequenced at a single sequencing center, and which were not under embargo by the TCGA as of March 2015. All results for the analysis are reported in Figure 6 and Supplementary Tables 4 and 5.

The most apparent result of our analysis was the association of a large number of CNAs in endometrial carcinomas (UCEC) with increased mtDNA abundance. Recent work by the TCGA proposed a subtype stratification of endometrial carcinomas based on mutation and CNA frequency [17]. Among these subtypes is a serous-like “copy-number-high” subtype with large numbers of somatic CNAs and extremely poor survival. We obtained the UCEC subtype classifications and confirmed that serous-like endometrial carcinomas exhibited substantially higher mtDNA copy number than all other subtypes (Mann-Whitney p-value 3×10^{-10} , Figure 6), explaining the large number of associations we observed. A number of genes are also known to be enriched or depleted for mutations in the serous-like subtype, and these mutations also showed statistically significant association with mtDNA abundance (enriched: *TP53*, BH-corrected p-value 4×10^{-5} , depleted: *PTEN*, BH-corrected p-value 4×10^{-5}). To our knowledge, this is the first report of an association between the serous-like subtype of endometrial carcinoma and mtDNA copy number.

[Figure 6 about here.]

After filtering out associations in UCEC, we were left with a small number of statistically significant mutations and CNAs associated with mtDNA abundance. Among these, the strongest signal arose from increased tumor mtDNA content in *IDH1*-mutant low grade gliomas (Figure 6, BH-corrected p-value 5×10^{-4}). Both *IDH1* and *IDH2* activating mutations induce production of the so-called “onco-metabolite” 2-hydroxyglutarate, which competitively inhibits α -ketoglutarate-dependent histone demethylases and 5-methylcytosine hydroxylases, inducing a hypermethylation phenotype [36, 45]. Surprisingly, *IDH2* mutations showed no statistically significant change in mtDNA abundance, suggesting that the effect is specific to the cytosolic isoform of *IDH*. Notably, mutations in *NF1*, showed a significant decrease in mtDNA abundance (BH-corrected p-value 0.03). These results echo a complementary finding by Navis and colleagues [28], who reported that a mutant *IDH1 R132H* oligodendroglioma xenograft model displayed high densities of mitochondria and increased levels of mitochondrial metabolic activity.

They proposed that an increase in mitochondrial mass would increase activity of mitochondrial *IDH2* and compensate for loss of activity introduced by mutant *IDH1*.

Finally, prompted by a recent report implicating mutations in mtDNA itself with the pathology of kidney chromophobe carcinomas (KICH) [7], we investigated the connection between mtDNA copy number and mtDNA mutations in KICH. Using somatic mtDNA mutation calls provided in [7], we examined whether mtDNA-mutated samples were likely to harbor more or less mtDNA copies than unmutated samples. Surprisingly, we found that samples harboring mtDNA indels contained much higher quantities of mtDNA than unmutated samples (Mann-Whitney U-test p-value 0.002, S3). The same effect was not found when examining only single nucleotide variants. These results suggest that the presence of inactivating mtDNA mutations may induce increased mtDNA replication, perhaps as a response to inadequate mitochondrial energy production.

6 Discussion

The function of the human mitochondrial genome is to encode 13 proteins critical to mitochondrial respiration. In this study, we have investigated the variation of mtDNA copy number levels across many tumor types, arriving at several intriguing observations. Across more than half of the tumor types we studied, we found evidence for depletion of mtDNA, relative to normal tissues. Orthogonal measurements of transcription levels in these samples linked this variation to downregulation of mitochondrially-localized metabolic pathways, including respiration, the TCA cycle, fatty-acid oxidation, and branched-chain amino acid metabolism. Finally, we used evidence of somatic alterations in tumor samples to identify a set of mutations and copy number variants which are significantly correlated to mtDNA levels.

Our findings of gross changes in mtDNA content in tumors echo a number of prior but isolated observations, largely based on quantitative PCR measurements and with substantially smaller sample sizes, of mtDNA copy number changes in cancers (see review by [46] for a complete discussion). These include a decrease in mtDNA in breast cancer [10, 25], liver cancer [22], and clear-cell kidney cancers [27, 29]. While the majority of our observations agree with prior work (when comparing to [46]), some of our results are in contradiction to prior studies. The discordance between findings seems in part due to inadequate sample sizes, and incomplete or unavailable matched normal tissue. For example, in contrast to [25] and [43], we find no clear increase or decrease in mtDNA content in thyroid or endometrial carcinomas, respectively. However, [25] profiled 20 paired thyroid tumors, versus 116 paired thyroid tumors in this report; and [43] utilized unpaired samples of tumor and normal endometrial tissue [43], versus 41 paired samples here.

Mitochondrial DNA copy number itself reflects the cell's capacity to conduct oxidative phosphorylation, and its depletion in many cancer types mirrors the long-standing observation that tumors reduce respiratory activity in favor of glycolytic flux (*i.e.* the Warburg effect). An interesting prospect for future study will be to identify the source of depletion of mtDNA in the cancer types like bladder and clear-cell renal cell carcinoma, where nearly all samples are depleted of mtDNA. A number of factors controlling mitochondrial biogenesis are known to be either activated (*e.g.* HIF1 α , PGC1 α [21]) or somatically mutated in cancers, (*e.g.* NRF2 and its regulator KEAP1 [33]) in cancers, with potential implications for mtDNA copy number.

While mtDNA depletion or accumulation may typify certain cancer types, we further identified that subsets of patients, characterized by the presence of particular somatic mutations/copy number alterations, were themselves enriched/depleted of mtDNA. The presence of activating IDH1 mutations (in low grade gliomas) or a large number of copy number alterations (in serous-like endometrial carcinomas) strongly predisposed patients to high tumor mtDNA content. If these tumors (and others with increased mtDNA content) harbor an increased dependence on mitochondrial metabolism to proliferate, this may offer a novel route for targeted therapy.

7 Methods

7.1 Data Acquisition

Whole exome sequencing (WXS) and whole genome sequencing (WGS) BAM files for 21 distinct TCGA studies were obtained from the TCGA CGHub repository (Figure 2) [44]. In total, we analyzed 13411 samples. We restricted our analyses to sequence data aligned to GRCH37 using the mitochondrial Cambridge Reference Sequence (CRS). We focused only on primary tumor, adjacent normal tissue, and normal blood samples (“01”, “11”, and “10” in the sample type field of the TCGA barcode). We further restricted our analyses to samples which were not whole-genome amplified prior to sequencing (*i.e.* we only used samples containing ‘D’ in the analyte field of the TCGA barcode), because such amplification could potentially bias the relative abundances of mitochondrial and nuclear DNA in the sample. Samtools [23] was used to extract a BAM file containing properly paired reads with minimal mapping quality aligning to the mitochondrial genome (using -F 256 -q 30 flags). The number of properly paired reads aligning to the nuclear genome was obtained by using the samtools flagstat command, and subtracting the number of properly paired reads aligning to the mitochondrial genome.

7.2 Purity and Ploidy Calculation and Correction

Affymetrix SNP6 arrays for tumor and normal samples were acquired for 21 cancer types from the TCGA. Arrays for each individual cancer type were processed together, quantile-normalized and median polished with Affymetrix power tools using the birdseed algorithm to obtain allele-specific intensities. PennCNV [42] was used to generate log R ratio and B-allele frequencies for each tumor. ASCAT [38] was used to generate allele-specific copy number and estimate tumor ploidy and purity using matched arrays from tumor and normal tissue.

In order to estimate mtDNA copy number, we compare the number of reads aligning to the mitochondrial genome to the number of reads aligning to a genome of known ploidy. For samples of normal tissue, we assume this known ploidy (R_{Normal}) is equal to 2. For tumor tissue which may be infiltrated by stromal/immune cells and copy-number altered, we need to estimate the “effective ploidy” of the sample. We define this effective ploidy to be equal to

$$R_{\text{Tumor}} = \text{Purity} \times \text{Ploidy} + (1 - \text{Purity}) \times 2 \quad (3)$$

7.3 Survival Analysis

Survival analysis was performed with univariate Cox proportional hazards regression models where the independent and dependent variables were the log10-transformed mtDNA copy number and the overall survival respectively. The p-values for the significance of mtDNA copy number as a predictor of survival were obtained from Wald tests.

7.4 Gene Set Analysis

RSEM normalized RNA-Seq gene expression data was downloaded from the Broad Firehose, using the most recent data as of November 4, 2014. Data were filtered to remove genes with average read count less than 16. We calculated the nonparametric (Spearman) correlation (and associated p-value) between the expression of each gene and the copy number of mtDNA in the corresponding sample. To remove putatively spurious correlations, we identified all genes with a p-value less than 0.05, and set the correlation coefficient for those genes equal to zero. We then use a rank-based Mann-Whitney U-test (implemented in the geneSetTest function in the limma [20] package) to test whether particular gene sets showed an enrichment for over- or under-expressed genes, relative to the distribution of correlations across all genes.

Thus, a single p-value is calculated for both alternative hypotheses (p_1^u, p_1^d). All p-values were adjusted using the Benjamini-Hochberg procedure. The method described above was applied to every Reactome pathway in the MSigDB canonical pathways gene set.

The analysis was run separately for tumor and normal tissues, and for samples processed by different sequencing center. We applied our gene set analysis pipeline to all studies for which we had at least 20 samples of RNA-Seq data (in order to retain sufficient statistical power). Analyses were run for each combination of tumor type, tissue, and sequencing center, and ensuing results were then aggregated across all studies. All results from the analyses are provided in the Supplementary Table 3.

7.5 Mutation and Copy Number Alteration Analysis

For each study, Gistic2 and MutSigCV results were downloaded from the Broad Firehose (most recent data as of Nov 14, 2014). From Gistic, we retained all arm-level and focal alterations with q-value less than 0.1. For mutations, we obtained the MAF file from the output of MutSig. For each gene, we calculated the number of patients in which this gene exhibited a nonsynonymous, coding mutation (*i.e.*, missense, non-sense, frameshift, in-frame insertion/deletions, and splice-site mutations), excluding those with greater than 600 non-synonymous coding mutations). We then retained any genes which were mutated in greater than 4 percent of patients. We used only WXS data, and did not analyze tumor types which were sequenced at more than one sequencing center. Non-parametric Mann-Whitney U-test were used to evaluate whether tumors bearing a particular somatic alteration contained significantly higher/lower amounts of mtDNA in tumor samples. After testing all associations, p-values obtained from the U-tests were corrected using the Benjamini-Hochberg procedure.

References

1. Robert S Balaban, Shino Nemoto, and Toren Finkel. Mitochondria, oxidants, and aging. *Cell*, 120(4):483–95, February 2005.
2. Navdeep S. Chandel and Paul T. Schumacker. Cells depleted of mitochondrial DNA ($\rho 0$) yield insight into physiological mechanisms. *FEBS Letters*, 454(3):173–176, July 1999.
3. Laurent Chatre and Miria Ricchetti. Prevalent coordination of mitochondrial DNA transcription and initiation of replication with the cell cycle. *Nucleic acids research*, 41(5):3068–78, March 2013.
4. L C Costello, P Feng, B Milon, M Tan, and R B Franklin. Role of zinc in the pathogenesis and treatment of prostate cancer: critical issues to resolve. *Prostate cancer and prostatic diseases*, 7(2):111–7, January 2004.
5. L C Costello, Y Liu, R B Franklin, and M C Kennedy. Zinc inhibition of mitochondrial aconitase and its importance in citrate metabolism of prostate epithelial cells. *The Journal of biological chemistry*, 272(46):28875–81, November 1997.
6. Giuseppe D’Antona, Maurizio Ragni, Annalisa Cardile, Laura Tedesco, Marta Dossena, Flavia Bruttini, Francesca Caliaro, Giovanni Corsetti, Roberto Bottinelli, Michele O Carruba, Alessandra Valerio, and Enzo Nisoli. Branched-chain amino acid supplementation promotes survival and supports cardiac and skeletal muscle mitochondrial biogenesis in middle-aged mice. *Cell metabolism*, 12(4):362–72, October 2010.
7. Caleb F Davis, Christopher J Ricketts, Min Wang, Lixing Yang, Andrew D Cherniack, Hui Shen, Christian Buhay, Hyojin Kang, Sang Cheol Kim, Catherine C Fahey, Kathryn E Hacker, Gyan Bhanot, Dmitry A Gordenin, Andy Chu, Preethi H Gunaratne, Michael Biehl, Sahil Seth, Benny A

- Kaiparettu, Christopher A Bristow, Lawrence A Donehower, Eric M Wallen, Angela B Smith, Satish K Tickoo, Pheroze Tamboli, Victor Reuter, Laura S Schmidt, James J Hsieh, Toni K Choueiri, A Ari Hakimi, Lynda Chin, Matthew Meyerson, Raju Kucherlapati, Woong-Yang Park, A Gordon Robertson, Peter W Laird, Elizabeth P Henske, David J Kwiatkowski, Peter J Park, Margaret Morgan, Brian Shuch, Donna Muzny, David A Wheeler, W Marston Linehan, Richard A Gibbs, W Kimryn Rathmell, and Chad J Creighton. The somatic genomic landscape of chromophobe renal cell carcinoma. *Cancer cell*, 26(3):319–30, September 2014.
8. Anna Maria D’Erchia, Anna Atlante, Gemma Gadaleta, Giulio Pavesi, Matteo Chiara, Caterina De Virgilio, Caterina Manzari, Francesca Mastropasqua, Gian Marco Prazzoli, Ernesto Picardi, Carmela Gissi, David Horner, Aurelio Reyes, Elisabetta Sbisà, Apollonia Tullo, and Graziano Pesole. Tissue-specific mtDNA abundance from exome data and its correlation with mitochondrial transcription, mass and respiratory activity. *Mitochondrion*, 20C:13–21, November 2014.
 9. Ayman W El-Hattab and Fernando Scaglia. Mitochondrial DNA depletion syndromes: review and updates of genetic basis, manifestations, and therapeutic options. *Neurotherapeutics : the journal of the American Society for Experimental NeuroTherapeutics*, 10(2):186–98, April 2013.
 10. Alex Xiu-Cheng Fan, Ramin Radpour, Mahdi Montazer Haghighi, Corina Kohler, Peng Xia, Sinuhe Hahn, Wolfgang Holzgreve, and Xiao Yan Zhong. Mitochondrial DNA content in paired normal and cancerous breast tissue samples from patients with breast cancer. *Journal of cancer research and clinical oncology*, 135(8):983–9, August 2009.
 11. T Finkel and N J Holbrook. Oxidants, oxidative stress and the biology of ageing. *Nature*, 408(6809):239–47, November 2000.
 12. Yan Guo, Jiang Li, Chung-I Li, Yu Shyr, and David C Samuels. MitoSeek: extracting mitochondria information and performing high-throughput mitochondria sequencing analysis. *Bioinformatics (Oxford, England)*, 29(9):1210–1, May 2013.
 13. Yiping He, Jian Wu, Devin C Dressman, Christine Iacobuzio-Donahue, Sanford D Markowitz, Victor E Velculescu, Luis A Diaz, Kenneth W Kinzler, Bert Vogelstein, and Nickolas Papadopoulos. Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. *Nature*, 464(7288):610–4, March 2010.
 14. S M Hutson, D Fenstermacher, and C Mahar. Role of mitochondrial transamination in branched chain amino acid metabolism. *The Journal of biological chemistry*, 263(8):3618–25, March 1988.
 15. Kaori Ishikawa, Keizo Takenaga, Miho Akimoto, Nobuko Koshikawa, Aya Yamaguchi, Hirotake Imanishi, Kazuto Nakada, Yoshio Honma, and Jun-Ichi Hayashi. ROS-generating mitochondrial DNA mutations can regulate tumor cell metastasis. *Science (New York, N.Y.)*, 320(5876):661–4, May 2008.
 16. Young Seok Ju, Ludmil B Alexandrov, Moritz Gerstung, Inigo Martincorena, Serena Nik-Zainal, Manasa Ramakrishna, Helen R Davies, Elli Papaemmanuil, Gunes Gundem, Adam Shlien, Niccolo Bolli, Sam Behjati, Patrick S Tarpey, Jyoti Nangalia, Charles E Massie, Adam P Butler, Jon W Teague, George S Vassiliou, Anthony R Green, Ming-Qing Du, Ashwin Unnikrishnan, John E Pimanda, Bin Tean Teh, Nikhil Munshi, Mel Greaves, Paresh Vyas, Adel K El-Naggar, Tom Santarius, V Peter Collins, Richard Grundy, Jack A Taylor, D Neil Hayes, David Malkin, Christopher S Foster, Anne Y Warren, Hayley C Whitaker, Daniel Brewer, Rosalind Eeles, Colin Cooper, David Neal, Tapio Visakorpi, William B Isaacs, G Steven Bova, Adrienne M Flanagan, P Andrew Futreal, Andy G Lynch, Patrick F Chinnery, Ultan McDermott, Michael R Stratton, and Peter J Campbell. Origins and functional consequences of somatic mitochondrial DNA mutations in human cancer. *eLife*, 3:e02935, October 2014.

17. Cyriac Kandoth, Nikolaus Schultz, Andrew D Cherniack, Rehan Akbani, Yuexin Liu, Hui Shen, A Gordon Robertson, Itai Pashtan, Ronglai Shen, Christopher C Benz, Christina Yau, Peter W Laird, Li Ding, Wei Zhang, Gordon B Mills, Raju Kucherlapati, Elaine R Mardis, and Douglas A Levine. Integrated genomic characterization of endometrial carcinoma. *Nature*, 497(7447):67–73, May 2013.
18. A King, M A Selak, and E Gottlieb. Succinate dehydrogenase and fumarate hydratase: linking mitochondrial dysfunction and cancer. *Oncogene*, 25(34):4675–82, August 2006.
19. Tatianna C Larman, Steven R DePalma, Angela G Hadjipanayis, Alexei Protopopov, Jianhua Zhang, Stacey B Gabriel, Lynda Chin, Christine E Seidman, Raju Kucherlapati, and J G Seidman. Spectrum of somatic mitochondrial mutations in five cancers. *Proceedings of the National Academy of Sciences of the United States of America*, 109(35):14087–91, August 2012.
20. Charity W Law, Yunshun Chen, Wei Shi, and Gordon K Smyth. Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome biology*, 15(2):R29, February 2014.
21. Valerie S LeBleu, Joyce T O’Connell, Karina N Gonzalez Herrera, Harriet Wikman, Klaus Pantel, Marcia C Haigis, Fernanda Machado de Carvalho, Aline Damascena, Ludmilla Thome Domingos Chinen, Rafael M Rocha, John M Asara, and Raghu Kalluri. PGC-1 α mediates mitochondrial biogenesis and oxidative phosphorylation in cancer cells to promote metastasis. *Nature cell biology*, 16(10):992–1003, September 2014.
22. Hsin-Chen Lee, Shu-Hui Li, Jin-Ching Lin, Cheng-Chung Wu, Dah-Cherg Yeh, and Yau-Huei Wei. Somatic mutations in the D-loop and decrease in the copy number of mitochondrial DNA in human hepatocellular carcinoma. *Mutation research*, 547(1-2):71–8, March 2004.
23. Heng Li, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, and Richard Durbin. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, 25(16):2078–9, August 2009.
24. Arthur Liberzon, Aravind Subramanian, Reid Pinchback, Helga Thorvaldsdóttir, Pablo Tamayo, and Jill P Mesirov. Molecular signatures database (MSigDB) 3.0. *Bioinformatics (Oxford, England)*, 27(12):1739–40, June 2011.
25. Elizabeth Mambo, Aditi Chatterjee, Mingzhao Xing, Giovanni Tallini, Bryan R Haugen, Sai-Ching J Yeung, Saraswati Sukumar, and David Sidransky. Tumor-specific changes in mtDNA content in human cancer. *International journal of cancer. Journal international du cancer*, 116(6):920–4, October 2005.
26. Jared R Mayers, Chen Wu, Clary B Clish, Peter Kraft, Margaret E Torrence, Brian P Fiske, Chen Yuan, Ying Bao, Mary K Townsend, Shelley S Tworoger, Shawn M Davidson, Thales Papagianakopoulos, Annan Yang, Talya L Dayton, Shuji Ogino, Meir J Stampfer, Edward L Giovannucci, Zhi Rong Qian, Douglas A Rubinson, Jing Ma, Howard D Sesso, John M Gaziano, Barbara B Cochrane, Simin Liu, Jean Wactawski-Wende, JoAnn E Manson, Michael N Pollak, Alec C Kimmelman, Amanda Souza, Kerry Pierce, Thomas J Wang, Robert E Gerszten, Charles S Fuchs, Matthew G Vander Heiden, and Brian M Wolpin. Elevation of circulating branched-chain amino acids is an early event in human pancreatic adenocarcinoma development. *Nature Medicine*, advance on, September 2014.
27. David Meierhofer, Johannes A Mayr, Ulrike Foetschl, Alexandra Berger, Klaus Fink, Nikolaus Schmeller, Gerhard W Hacker, Cornelia Hauser-Kronberger, Barbara Kofler, and Wolfgang Sperl. Decrease of mitochondrial DNA content and energy metabolism in renal cell carcinoma. *Carcinogenesis*, 25(6):1005–10, June 2004.

28. Anna C Navis, Simone P Niclou, Fred Fack, Daniel Stieber, Sanne van Lith, Kiek Verrijp, Alan Wright, Jonathan Stauber, Bastiaan Tops, Irene Otte-Holler, Ron A Wevers, Arno van Rooij, Stefan Pusch, Andreas von Deimling, Wikky Tigchelaar, Cornelis J F van Noorden, Pieter Wesseling, and William P J Leenders. Increased mitochondrial activity in a novel IDH1-R132H mutant human oligodendroglioma xenograft model: in situ detection of 2-HG and α -KG. *Acta neuropathologica communications*, 1(1):18, January 2013.
29. H Nilsson, D Lindgren, A Mandahl Forsberg, H Mulder, H Axelson, and M E Johansson. Primary clear cell renal carcinoma cells display minimal mitochondrial respiratory capacity resulting in pronounced sensitivity to glycolytic inhibition by 3-Bromopyruvate. *Cell death & disease*, 6:e1585, January 2015.
30. Carlota Rubio-Perez, David Tamborero, Michael P. Schroeder, Albert A. Antolín, Jordi Deu-Pons, Christian Perez-Llamas, Jordi Mestres, Abel Gonzalez-Perez, and Nuria Lopez-Bigas. In Silico Prescription of Anticancer Drugs to Cohorts of 28 Tumor Types Reveals Targeting Opportunities. *Cancer Cell*, 27(3):382–396, March 2015.
31. David C Samuels, Leng Han, Jiang Li, Sheng Quangu, Travis A Clark, Yu Shyr, and Yan Guo. Finding the lost treasures in exome sequencing data. *Trends in genetics : TIG*, 29(10):593–9, October 2013.
32. David C Samuels, Chun Li, Bingshan Li, Zhuo Song, Eric Torstenson, Hayley Boyd Clay, Antonis Rokas, Tricia A Thornton-Wells, Jason H Moore, Tia M Hughes, Robert D Hoffman, Jonathan L Haines, Deborah G Murdock, Douglas P Mortlock, and Scott M Williams. Recurrent tissue-specific mtDNA mutations are common in humans. *PLoS genetics*, 9(11):e1003929, November 2013.
33. Richard C Scarpulla. Transcriptional paradigms in mammalian mitochondrial biogenesis and function. *Physiological reviews*, 88(2):611–38, April 2008.
34. Eric A Schon, Salvatore DiMauro, and Michio Hirano. Human mitochondrial DNA: roles of inherited and somatic mutations. *Nature reviews. Genetics*, 13(12):878–90, December 2012.
35. Frederic D Sigoillot, J Andrew Berkowski, Severine M Sigoillot, Damian H Kotsis, and Hedeel I Guy. Cell cycle-dependent regulation of pyrimidine biosynthesis. *The Journal of biological chemistry*, 278(5):3403–9, January 2003.
36. Sevin Turcan, Daniel Rohle, Anuj Goenka, Logan A Walsh, Fang Fang, Emrullah Yilmaz, Carl Campos, Armida W M Fabius, Chao Lu, Patrick S Ward, Craig B Thompson, Andrew Kaufman, Olga Guryanova, Ross Levine, Adriana Heguy, Agnes Viale, Luc G T Morris, Jason T Huse, Ingo K Mellinghoff, and Timothy A Chan. IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature*, 483(7390):479–83, March 2012.
37. Alessandra Valerio, Giuseppe D’Antona, and Enzo Nisoli. Branched-chain amino acids, mitochondrial biogenesis, and healthspan: an evolutionary perspective. *Aging*, 3(5):464–78, May 2011.
38. Peter Van Loo, Silje H Nordgard, Ole Christian Lingjærde, Hege G Russnes, Inga H Rye, Wei Sun, Victor J Weigman, Peter Marynen, Anders Zetterberg, Bjørn Naume, Charles M Perou, Anne-Lise Børresen Dale, and Vessela N Kristensen. Allele-specific copy number analysis of tumors. *Proceedings of the National Academy of Sciences of the United States of America*, 107(39):16910–5, September 2010.
39. Matthew G Vander Heiden, Lewis C Cantley, and Craig B Thompson. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science (New York, N. Y.)*, 324(5930):1029–33, May 2009.

40. Timothy Wai, Asangla Ao, Xiaoyun Zhang, Daniel Cyr, Daniel Dufort, and Eric A Shoubridge. The role of mitochondrial DNA copy number in mammalian fertility. *Biology of reproduction*, 83(1):52–62, July 2010.
41. Douglas C Wallace. Mitochondria and cancer. *Nature reviews. Cancer*, 12(10):685–98, October 2012.
42. Kai Wang, Mingyao Li, Dexter Hadley, Rui Liu, Joseph Glessner, Struan F A Grant, Hakon Hakonarson, and Maja Bucan. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome research*, 17(11):1665–74, November 2007.
43. Yue Wang, Vincent W S Liu, Wei-Cheng Xue, Percy C K Tsang, Annie N Y Cheung, and Hextan Y S Ngan. The increase of mitochondrial DNA content in endometrial adenocarcinoma cells: a quantitative study using laser-captured microdissected tissues. *Gynecologic oncology*, 98(1):104–10, July 2005.
44. Christopher Wilks, Melissa S Cline, Erich Weiler, Mark Diehkans, Brian Craft, Christy Martin, Daniel Murphy, Howdy Pierce, John Black, Donovan Nelson, Brian Litzinger, Thomas Hatton, Lori Maltbie, Michael Ainsworth, Patrick Allen, Linda Rosewood, Elizabeth Mitchell, Bradley Smith, Jim Warner, John Groboske, Haifang Telc, Daniel Wilson, Brian Sanford, Hannes Schmidt, David Haussler, and Daniel Maltbie. The Cancer Genomics Hub (CGHub): overcoming cancer through the power of torrential data. *Database : the journal of biological databases and curation*, 2014, January 2014.
45. Wei Xu, Hui Yang, Ying Liu, Ying Yang, Ping Wang, Se-Hee Kim, Shinsuke Ito, Chen Yang, Pu Wang, Meng-Tao Xiao, Li-xia Liu, Wen-qing Jiang, Jing Liu, Jin-ye Zhang, Bin Wang, Stephen Frye, Yi Zhang, Yan-hui Xu, Qun-ying Lei, Kun-Liang Guan, Shi-min Zhao, and Yue Xiong. Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α -ketoglutarate-dependent dioxygenases. *Cancer cell*, 19(1):17–30, January 2011.
46. Man Yu. Generation, function and diagnostic value of mitochondrial DNA copy number alterations in human cancers. *Life sciences*, 89(3-4):65–71, July 2011.
47. Min Zhao, Qingguo Wang, Quan Wang, Peilin Jia, and Zhongming Zhao. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. *BMC bioinformatics*, 14 Suppl 1(Suppl 11):S1, January 2013.

8 Supplementary Tables

Center	Correlation	N	P-Value	Study
WXS:BI::WGS:HMS-RK	0.49	171	1.7e-11	BLCA
WXS:WUGSC::WGS:WUGSC	0.55	73	5.3e-07	BRCA
WXS:BCM::WGS:BCM	0.87	21	3.5e-07	KIRC
WXS:BI::WGS:BCM	0.48	25	0.015	KIRC
WXS:BI::WGS:BI	1	4	0	KIRC
WXS:BI::WGS:BI	0.71	59	4.5e-10	GBM
WXS:BI::WGS:BCM	0.42	43	0.0049	HNSC
WXS:BI::WGS:BI	0.3	34	0.08	HNSC
WXS:BI::WGS:HMS-RK	0.55	182	9.3e-16	HNSC
WXS:BI::WGS:BI	0.38	32	0.033	LUAD

WXS:BI::WGS:HMS-RK	0.56	162	1.2e-14	LUAD
WXS:BI::WGS:BI	0.89	7	0.0068	OV
WXS:BI::WGS:BI	0.51	69	9.2e-06	THCA
WXS:BI::WGS:HMS-RK	0.64	151	1.3e-18	THCA
WXS:WUGSC::WGS:HMS-RK	0.81	104	1e-25	UCEC
WXS:WUGSC::WGS:WUGSC	0.64	37	2.4e-05	UCEC

Table 1. Correlation of mtDNA estimates between WXS and WGS. N is the number of matched samples. Correlation corresponds to the Spearman correlation coefficient, and P-Value corresponds to the p-value of this correlation coefficient. Center: BI, Broad Insitute; BCM, Baylor College of Medicine; WUGSC, Washington University; HMS: Harvard Medical School.

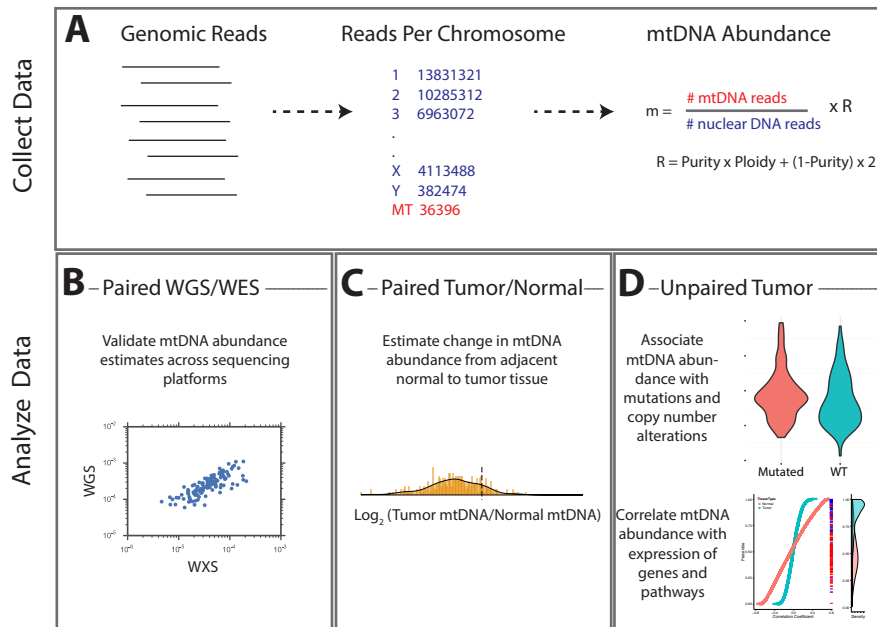


Figure 1. (A) Aligned reads are analyzed to determine the number of primary, properly paired reads aligning to each chromosome. Relative abundance of mitochondrial DNA is calculated as the ratio of mtDNA reads to nuclear DNA reads, and corrected for tumor purity and ploidy. The results of these calculations were employed in three different types of analysis. (B) Comparisons across samples profiled by both whole exome and whole genome sequencing provided validation of mtDNA copy number estimates. (C) Pairs of matched tumor/adjacent-normal samples were compared to uncover patterns of mtDNA accumulation and depletion. (D) Using all data available (including tumor samples lacking matched normal samples), statistical associations between mtDNA copy number and (1) mutation/copy number alterations, and (2) gene expression, were calculated.

FIGURES

15

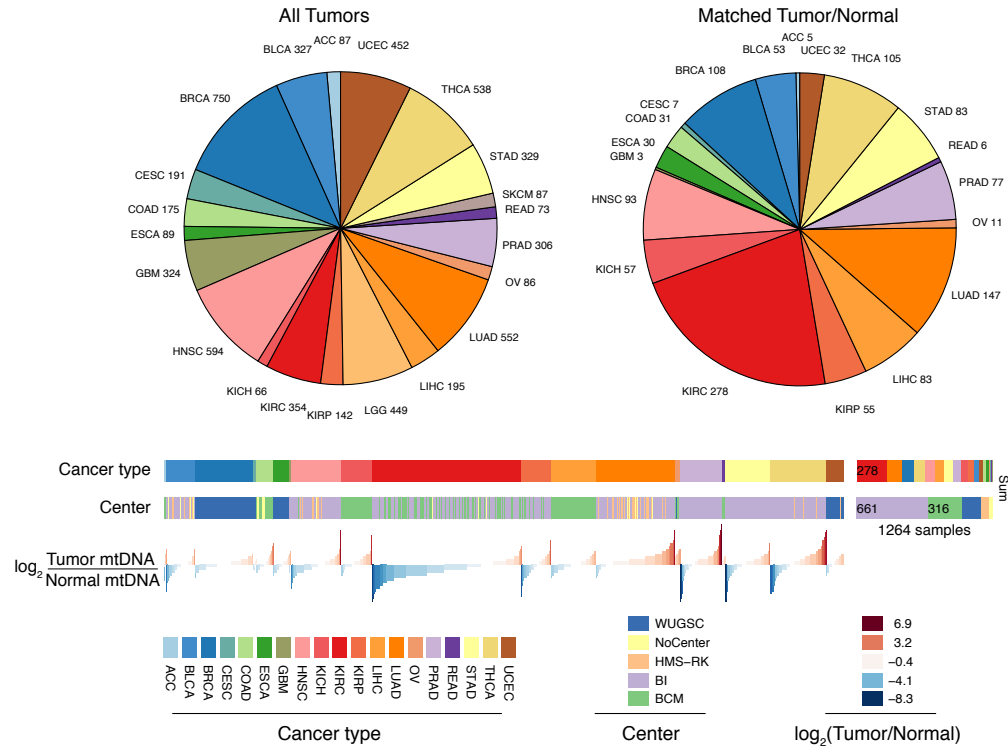


Figure 2. Summary of Data. Whole-exome and whole-genome sequencing data was obtained from 21 TCGA studies. The data was processed at 5 different sequencing centers, each of which was analyzed separately. Over 1400 of our samples are paired instances of tumor/adjacent-normal tissue from the same patient, which are used to quantify changes in mtDNA content across tumors.

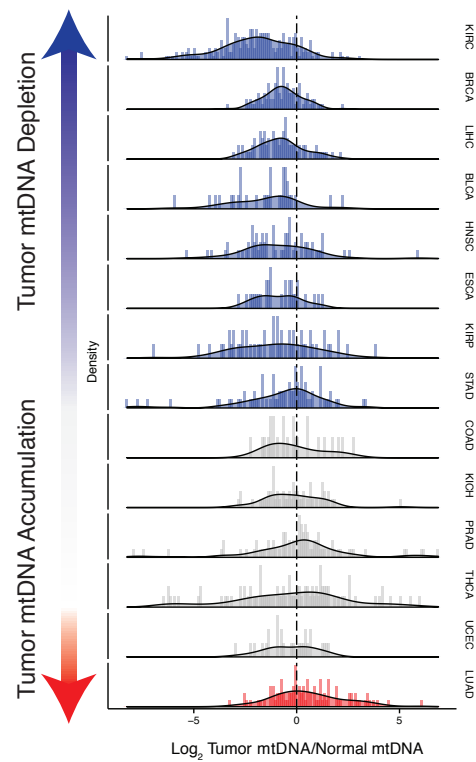


Figure 3. Many tumor types show depletion of mtDNA in tumor samples, relative to adjacent normal tissue. Normalized histograms and density plots illustrate \log_2 ratio of mtDNA content in tumor tissue, to mtDNA content in normal tissue. Tumor types with significant changes in mtDNA copy number changes are indicated in the sidebar with colored circles. Statistical significance of trends is assessed using Wilcoxon rank sign, with $-\log_{10}$ p-values reported in the sidebar by the radius of the circle.

FIGURES

17

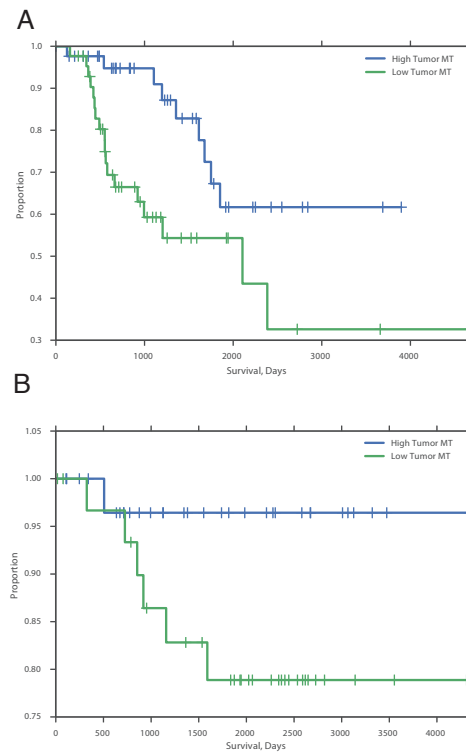


Figure 4. mtDNA content is significantly associated with patient survival in (A) adrenocortical (ACC) and (B) kidney chromophobe carcinoma (KICH). For visualization purposes, patients are partitioned into two groups, based on tumor mtDNA copy number relative to the median mtDNA copy number across all tumor samples in the cancer type. Cox regression identified a significant association between high tumor mtDNA and better survival (ACC, p-value 0.017; KICH, p-value 0.058).

FIGURES

18

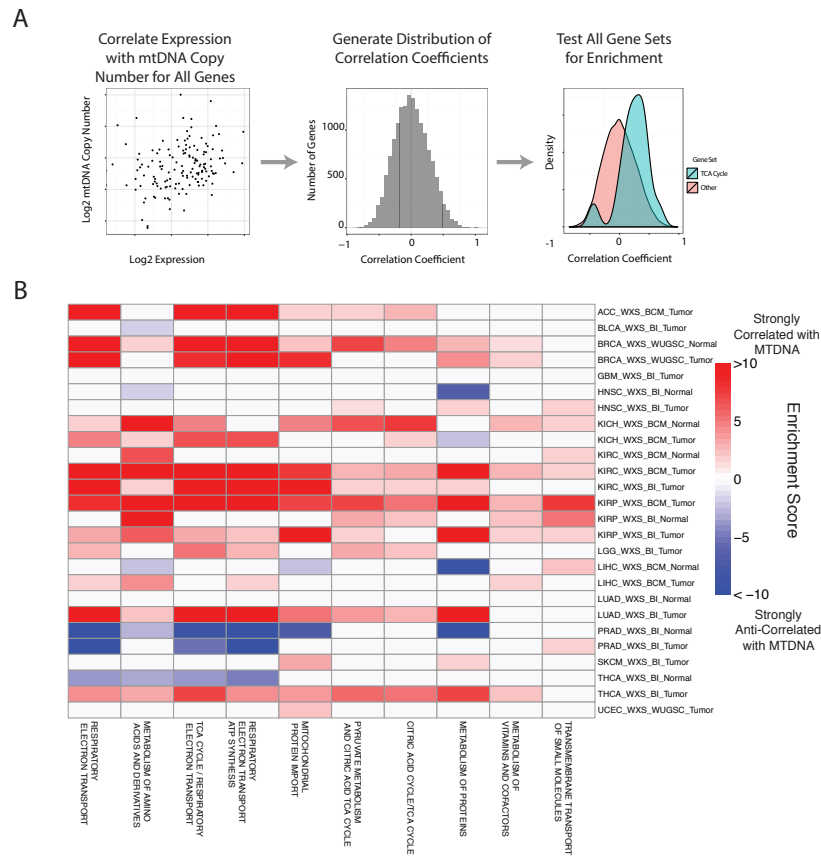


Figure 5. Gene set analysis identifies pathways correlated to mtDNA content. (A) Correlations between each gene and mtDNA content is calculated. Then, gene sets enriched for high/low correlation coefficients are identified. (B) mtDNA copy number is most strongly correlated to metabolic pathways including respiratory electron transport and the TCA cycle, which are localized to the mitochondria.

FIGURES

19

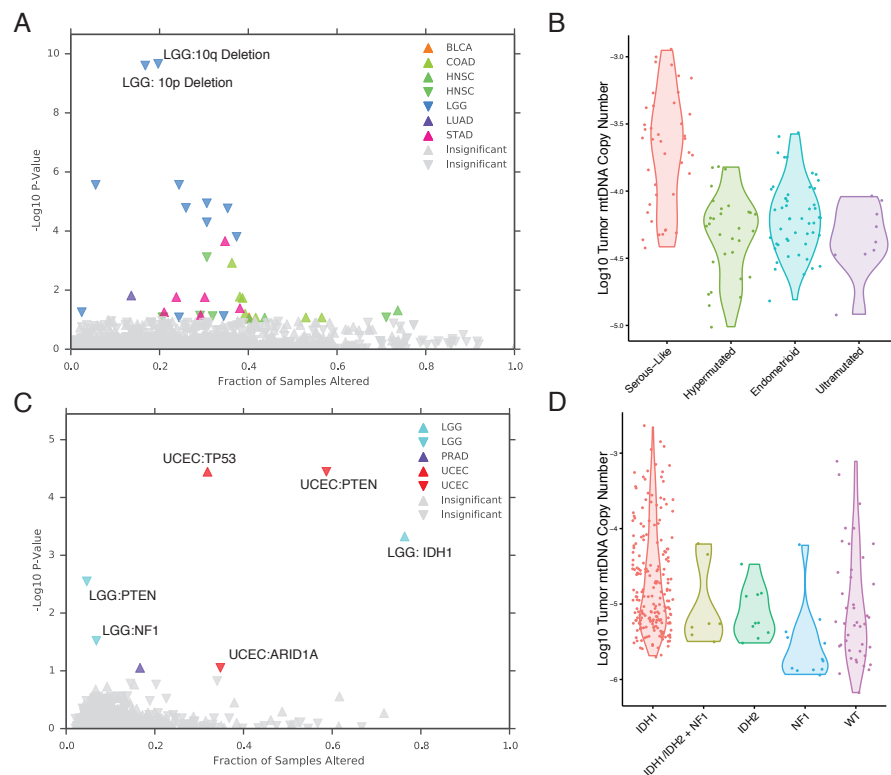


Figure 6. mtDNA content is correlated to the incidence of certain mutations and copy number alterations. Each point corresponds to a single alteration (*e.g.* *P53* mutation). Direction of arrow indicates whether alteration increases or decreases mtDNA content. X-axis in (A) and (B) indicates the fraction of samples in a cancer-type that contained the alteration (*i.e.* $\approx 25\%$ of LGG samples were 10q deleted). (A) 31 out of 1434 CNAs tested were found to be significantly associated with mtDNA content. CNAs from the UCEC cancer type were filtered out because of strong association between mtDNA copy number and the “copy-number high” serous-like subtype in UCEC, shown in (B). (B) The UCEC serous-like subtype displays a marked increase in tumor mtDNA copy number, relative to tumors of other subtypes. (C) Relatively few mutations (7 out of 1766 tested) associate significantly with tumor mtDNA content. The UCEC associations are likely the result of the correlation to the serous-like subtype. (D) *IDH1* and *NF1* mutation status is strongly linked to increased tumor mtDNA copy number in LGG.

FIGURES

20

[Figure S1 about here.]

[Figure S2 about here.]

[Figure S3 about here.]

FIGURES

21

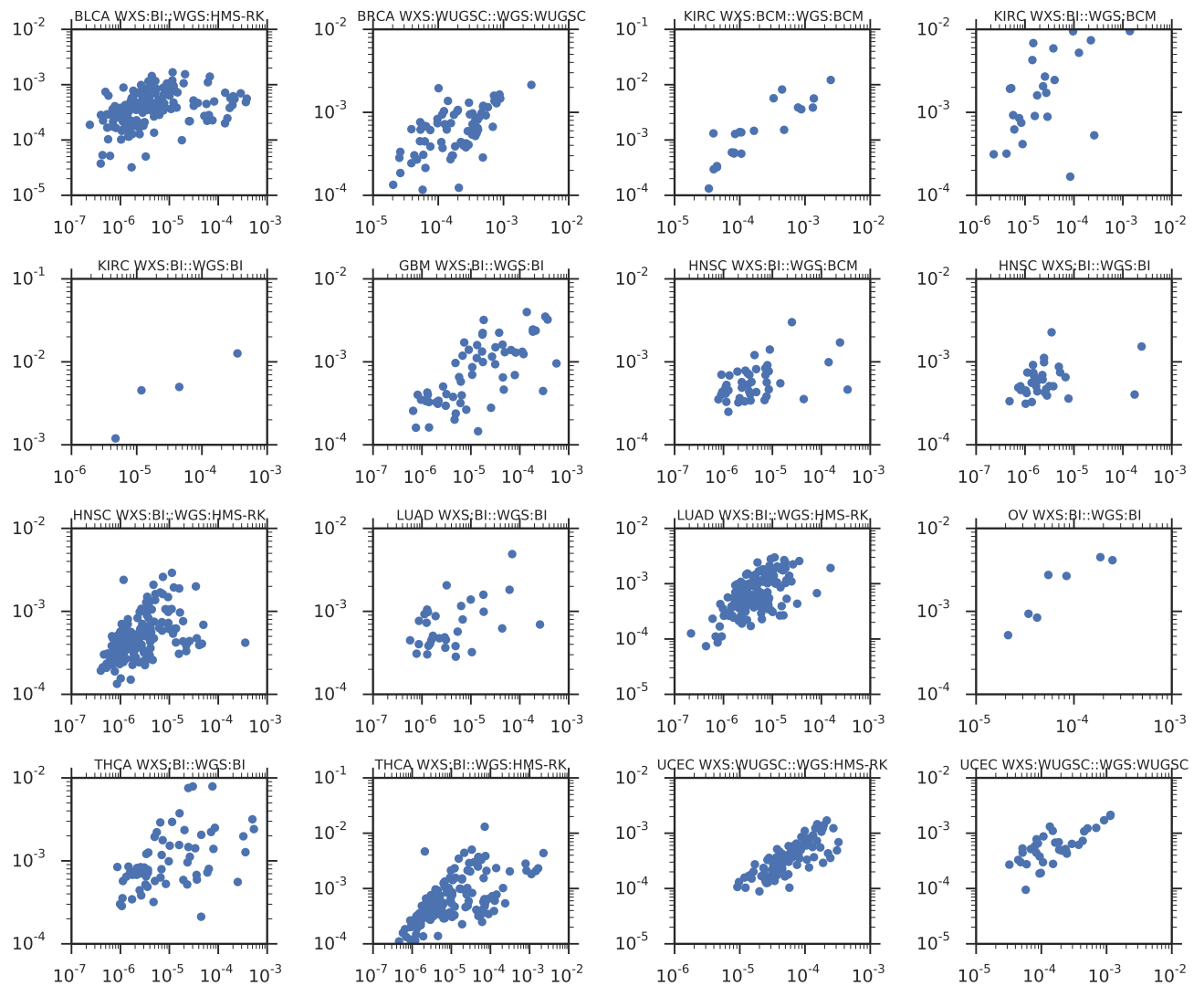


Figure S1. Comparison of mtDNA copy number estimates of samples profiled by both whole genome (WGS) and whole exome (WXS) sequencing. Data are separated by sequencing center, indicated in the title of each subplot. Center: BI, Broad Institute; BCM, Baylor College of Medicine; WUGSC, Washington University; HMS: Harvard Medical School.

FIGURES

22

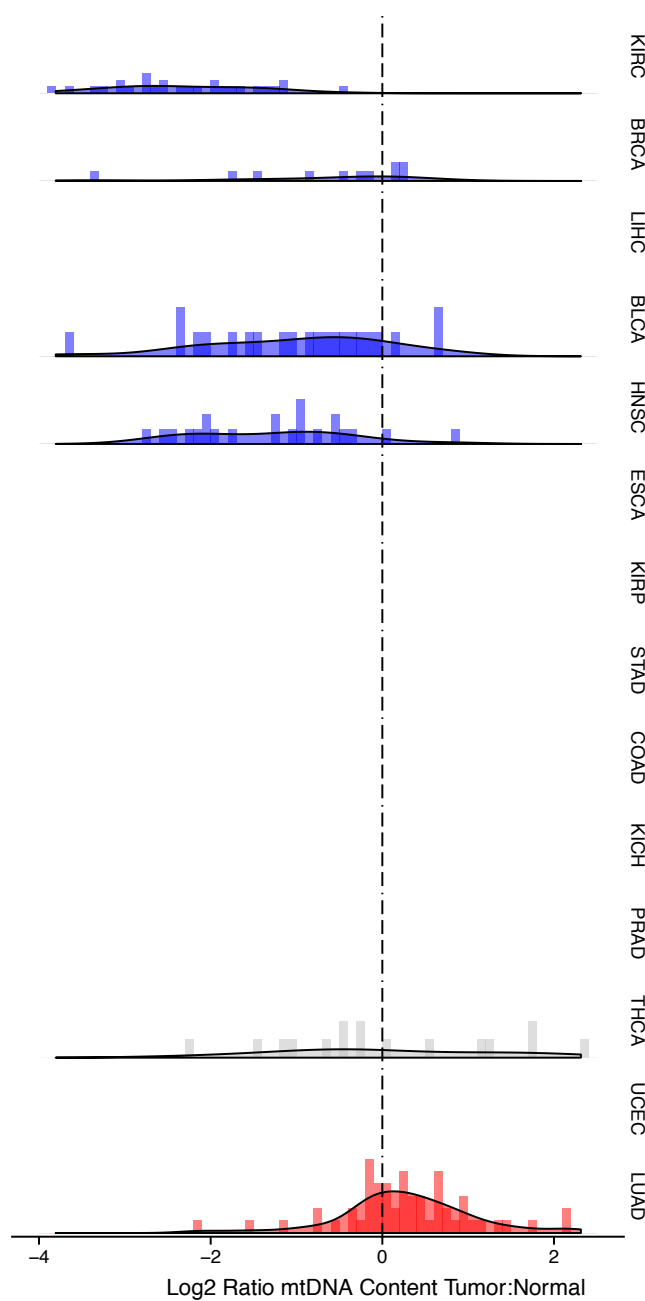


Figure S2. mtDNA tumor:normal copy number ratio using whole genome sequencing (WGS) data. Trends in WGS are in agreement with those in Figure 2 in whole exome sequencing (WXS).

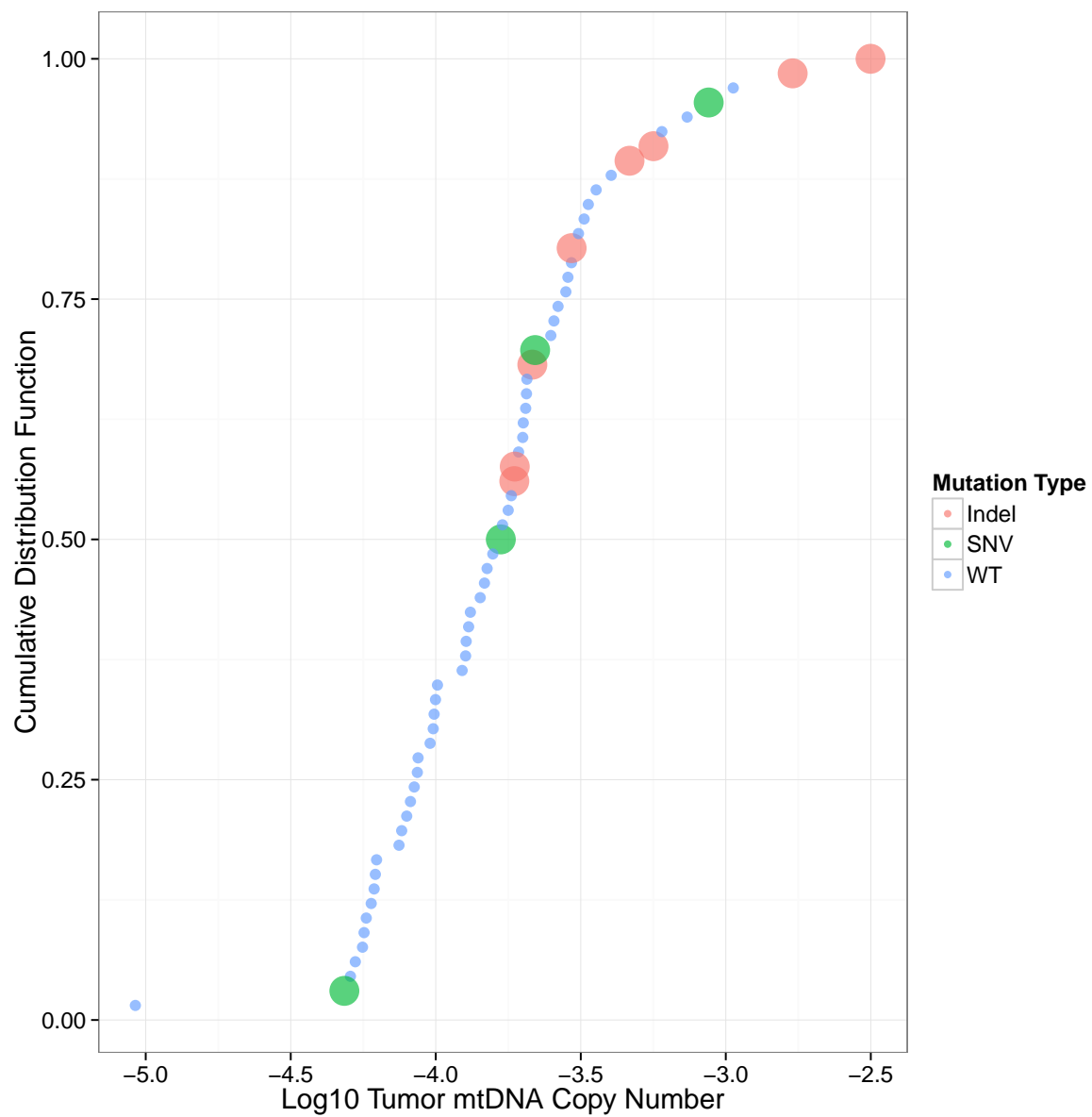


Figure S3. mtDNA copy number levels in kidney chromophobe carcinoma are elevated in samples with truncating mtDNA mutations. Samples are sorted by increasing mtDNA copy number on the X-axis, with percentile indicated on the Y-axis.