Circadian clock circuitry deconvolutes colorectal cancer and lung adenocarcinoma heterogeneity in a dynamic time-related framework

Fabrizio Bianchi (f.bianchi@operapadrepio.it)
Fondazione IRCCS Casa Sollievo della Sofferenza https://orcid.org/0000-0001-7412-8638
Gianluigi Mazzoccoli
IRCCS Casa Sollievo della Sofferenza https://orcid.org/0000-0003-3535-7635
Valentina Melocchi
Institute for Stem-cell Biology, Regenerative Medicine and Innovative Therapies https://orcid.org/0000-0001-9656-3681

Brief Communication

Keywords: Colorectal cancer, Lung cancer, gene expression profiling, molecular subtypes, circadian rhythm, biological clock

Posted Date: December 19th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2297936/v1

Version of Record: A version of this preprint was published at Cancer Gene Therapy on July 21st, 2023.
See the published version at https://doi.org/10.1038/s41417-023-00646-7.

Abstract

Increasing evidence imputes cancer progression and resistance to therapy to intra-tumor molecular heterogeneity set off by cancer cell plasticity. Re-activation of developmental programs strictly linked to epithelial-to-mesenchymal transition and gaining of stem cells properties are crucial in this setting. Many biological processes involved in cancer onset and progression show rhythmic fluctuations driven by the circadian clock circuitry. Novel cancer patient stratification tools taking into account the temporal dimension of these biological processes are definitely needed. Lung cancer and colorectal cancer (CRC) are the leading causes of cancer death worldwide.

Here, we show that the molecular heterogeneity characterizing the two deadliest cancers, colorectal (CRC) and lung adenocarcinoma (LUAD), rather than a merely stochastic event is the readout of specific cancer molecular states which correlate with time-qualified patterns of gene expression. We performed time-course transcriptome analysis of CRC and LUAD cell lines and upon computing circadian genes expression-based correlation matrices we exploited pseudo-time points to infer time-qualified patterns in the transcriptomic analysis of real-world data (RWD) from large cohorts of CRC and LUAD patients. Our temporal classification of CRC and LUAD cohorts was able to effectively render time-specific patterns in cancer phenotype switching determining dynamical distribution of molecular subtypes impacting patient prognosis.

Introduction

Most of the biological processes crucially implicated in carcinogenesis when altered are characterized by rhythmic fluctuations driven by the circadian clock circuitry and in turn, oncogenic processes openly thwart and disrupt the biological clock. The derangement of the molecular clockworks and the disorganization of the circadian timing system contribute to the onset and development of various types of cancer by altering several cancer related pathways [1,2].

Comprehensive genomic profiling of human cancer is unveiling the vast molecular heterogeneity of solid tumors, which nowadays appears to be the major hurdle for the identification of effective therapeutic strategies for neoplastic disease. What we learned so far is that beyond inter-samples heterogeneity each tumor sample is dominated by high intratumor heterogeneity (ITH), which supports the evolution of aggressive cancer cell subpopulations that drive metastatic spreading and drug resistance [3–5]. Furthermore, analyses of human cancer transcriptome unveiled the existence of molecular subtypes, which in part recapitulate such heterogeneity, in that specific tumor subtypes are characterized by peculiar genetic and epigenetic signatures and clinical-pathological characteristics [6,7]. However, this rather static interpretation of tumor molecular subtypes described like distinct entities has been recently challenged by single cell -omics studies performed in colorectal cancer (CRC) that described the co-existence of different CRC molecular subtypes (CMS) in the same tumor tissue [8]. In addition, molecular profiling of circulating tumor DNA in liquid biopsy recently showed a dynamic pattern of cancer molecular features in terms of fluctuation of anti-EGFR mutations, during longitudinal studies of patients treated

with anti-EGFR therapy [9]. Such evidence was not limited to CRC but did also apply to other tumor types [10].

Cancer cell plasticity, which is driven by mechanisms involved in epithelial-to-mesenchymal transition (EMT) and cancer cell stemness, can be at the basis of dynamic molecular heterogeneity as recently suggested by independent studies [8,11,12]. Interestingly, we recently found that loss of expression of the circadian gene TIMELESS triggers an epigenetic remodeling of CRC cells under the control of ZEB1, a canonical EMT master regulator, which dictates the acquirement of metastatic and stem-like phenotypes [13]. Besides, we discovered an aggressive molecular subtype in lung adenocarcinoma characterized by high-cell plasticity and immunoevasion phenotypes [14]. Yet, another circadian gene, namely ARNTL2, was recently found to drive metastatic spreading in lung adenocarcinoma (LUAD) [15]. Taken together, these evidences suggest that IHT, rather than a merely stochastic event, could be the readout of specific cell plasticity states which correlate with time-qualified patterns of gene expression, ultimately determining the hierarchy of cancer molecular subtypes and patient prognosis.

Here, we tackled this issue ordering in a dynamic fashion single-shot tumor sampling through pseudo-time points resulting from cancer cell lines time-series gene expression profiling to infer time-qualified patterns in the transcriptomic analysis of real-world data (RWD) from large cohorts of CRC and LUAD patients. We show, for the first time, that tumor molecular heterogeneity and subtypes do not necessarily follow a rigid topological organization, but rather is endowed with a dynamic time-related framework managed by the circadian clock circuitry.

Materials And Methods

Gene expression data. Raw Affymetrix data for SW480 and SW620 cell lines [16] were downloaded from ArrayExpress database at EMBL-EBI (E-MTAB-5876). Raw data were normalized using the RMA (Robust Multichip Average) methodology. Normalized Agilent Microarray data for PC9 cell lines [17] were downloaded from GEO database (GSE34228). Normalized RNA-seq expression data of TCGA-COAD patients [18] and clinical data were downloaded from cBioPortal [19] likewise normalized RNA-seq expression and clinical data of TCGA-LUAD [6].

TCGA-COAD and TCGA-LUAD pseudo-time classification analysis. CRC and LUAD samples were classified according to pseudo-time points by performing correlation analysis among core-clock network (CCN) genes/TIMELESS/ZEB1 expression profiles in SW480 or PC9 cell lines time-series transcriptome analysis (used as time-qualified ‘reference’) and the CCN genes/TIMELESS/ZEB1 expression profiles in TCGA-COAD or TCGA-LUAD samples. Briefly, Z-score values of the CCN genes/TIMELESS/ZEB1 expression were calculated at each time point of the SW480 and PC9 time-series transcriptome analysis as well as in samples of the TCGA-COAD/TCGA-LUAD cohorts. Next, we calculated Pearson correlation ($\boldsymbol{\rho}$) of CCN genes/TIMELESS/ZEB1 z-score values comparing TCGA-COAD/TCGA-LUAD samples with each time point of the SW480 or PC9 time series. Each TCGA sample was then assigned to a specific

time point according to the highest positive **ρ** value. Lastly, we summarized the expression values of each gene of the 'pseudo-time point' classified TCGA samples using median gene expression in order to obtain a correlation matrix with one expression value for each gene in a specific time point.

Circadian genes identification analysis. To detect transcripts exhibiting rhythmic pattern of expression with circadian (24 ± 4-hour) periodicity, the MetaCycle R package [20] was applied to normalized gene expression data for cell lines and to median normalized gene expression data for TCGA-COAD-time/TCGA-LUAD-time classified patients' samples. The minimum period length was set to 20 hours and the maximum period length to 28 hours. Transcripts with significant rhythms (p-value < 0.05) were selected for further analyses (Table S1). Amplitude fold change (FC) was calculated as previously explained [16]: (AmpFC = (1 + Amprel)/(1 − Amprel)), where Amprel is the relative amplitude reported in MetaCycle output table.

Phase Set Enrichment Analysis. Circadian-related pathways were determined by Phase Set Enrichment Analysis (PSEA) [21] based on the statistically significant circadian transcripts identified by MetaCycle. Nine gene sets were either manually selected from the Molecular Signatures database (MSigDB) gene sets collections [22] or custom-made by literature search. Genes involved in proliferation were annotated by the GO term GO:0008283. Acrophases of all circadian transcripts (rounded to the full hour) were compared to a uniform background distribution and statistically significant circadian-related gene sets were identified (q-value < 0.25).

Colon Molecular Subtype (CMS) classification. CMSclassifier R package [7] was used to classify each time point of the time-course microarray experiments and each TCGA-COAD sample. For TCGA-COAD patients, $\log_2$ normalized gene expression data were used.

Lung Molecular Subtype (C1/nonC1) classification. TCGA-LUAD samples were classified accordingly to our previous publication [14].

Results

We conducted time-series transcriptome analysis of CRC cell lines (SW480 and SW620), derived from matched primary and metastatic (lymph node) lesions of the same patient [16]. Further than the whole transcriptome we focused and investigated the expression profile of circadian genes (N = 15), representative of the core-clock network (CCN) (Figure S1A). Besides, we analyzed TIMELESS (aka TIM), another circadian gene, and ZEB1, an EMT master regulator which we recently showed to be modulated by TIM [13] (Figure S1A). As previously reported [23], all CCN were expressed in both CRC cell lines with peculiar oscillation patterns, but with decreased fluctuation amplitude in SW620 cells (metastatic cell line) compared to SW480 cells (primary cell line) (Figure S1A). Overall, we found 14% (N = 3 666, genes) of the SW480 transcriptome with a circadian pattern of expression profile while 8% (N = 2 126, genes) of SW620 transcriptome (Table S1). In SW480 cell line, we identified the highest peak of gene expression corresponding to 9h after synchronization and 52% of transcripts peaking between 6h and 12h (Figure

S1B, upper panel). In SW620 cell line, we observed a similar trend with the highest peak of gene expression corresponding to 9h after synchronization and 53% of transcripts peaking between 6h and 12h, but with a lower number of circadian transcripts compared to SW480 cell line (Figure S1B). Notwithstanding a lower number of circadian transcripts found in SW620 cell line, the expression amplitudes were comparable to SW480 cell line (Figure S1C). Intriguingly, we observed an anti-phasic expression pattern of TIM and ZEB1, with TIM expression acro-phase around 15h-18h time point, and ZEB1 expression acro-phase around 3h time point (Figure S1A) in both CRC cell lines. As we previously showed, loss of TIM expression induces EMT/stemness by upregulating ZEB1 while TIM upregulation correlates with positive regulation of cell proliferation/DNA repair, all relevant mechanisms for CRC progression [13]. We thus investigated whether the transcriptional pattern of genes enriched in such mechanisms were also featured by circadian rhythmicity. We applied Phase Set Enrichment Analysis (PSEA) [24] and observed an enrichment of pathways involved in tumor cell plasticity and stemness (e.g., Wnt/Notch/Hyppo signaling pathways, and EMT) particularly in the first 12 hours of the time series ($q < 0.25$; Fig. 1A,B), when ZEB1 expression reached the maximum (i.e., 3h; Fig. 1A,B). Contrariwise, we scored the enrichment of signaling pathways controlling cell proliferation, such as G2M checkpoint, and DNA damage repair ($q < 0.25$; Fig. 1A,B), in the second half of the 24-hour time series, when TIM reached its maximal expression (i.e., 15h-18h; Fig. 1A,B).

We then asked how these findings could actually impact the molecular pathology of CRC. To do this, we developed an approach to render the human CRC transcriptome through a more dynamic scenario by remapping 'static' human CRC transcriptome to pseudo time-series. Briefly, we used Pearson product-moment correlation ($\boldsymbol{\rho}$) of z-score values obtained from CCN genes/TIM/ZEB1 expression in time-series microarray analysis of human CRC cell lines to classify real-word data (RWD) of CRC transcriptomes (i.e., TCGA-COAD cohort; N = 436; see methods) according to their highest expression correlation with time-qualified points of CRC cell lines (Fig. 1C; see methods). We selected the SW480 cell lines to be used as reference because the metastatic SW620 cell line showed a stronger dysregulation of the clock, as previously reported [16].

Remarkably, we observed a coherent CCN expression profile in the reclassified CRC samples when compared with CRC cell lines time-course experiment (Fig. 1D). In keeping with this, when we applied PSEA to pseudo time-series of RWD CRC transcriptomes, once more we observed enrichment of WNT signaling pathway which is involved in cancer cell plasticity and stemness acquisition [25] followed (after 12 hours) by mechanisms relevant for cell proliferation (Fig. 1E). Again, the expression of ZEB1 peaked at 3h while TIM expression peaked at 15h, coherently with the type of mechanism enriched (Fig. 1D).

We then asked if the intrinsic molecular subtypes (CMS 1–4) which accurately describe inter-tumor heterogeneity and ITH in CRC might follow a circadian pattern of distribution. Interestingly, the CMS4, corresponding to the mesenchymal subtype (EMT high, TGFβ high, matrix remodeling) was significantly enriched at 3h pseudo-time point along with high-ZEB1 and low-TIM expression (Fig. 1F). Contrariwise, the CMS1 subtype (MSI-high, BRAF mutation, proliferation) was enriched at 9h pseudo-time point (Fig. 1F), along with initial increase of TIM expression (Fig. 1E) followed by the CMS3 (i.e., metabolic

subtype) which was enriched at 18h pseudo-time point (Fig. 1F). Of note, CMS1 distribution appeared to be in anti-phase with CMS4 distribution thus recapitulating TIM/ZEB1 anti-regulation (Fig. 1F). Other genetic characteristics of CRC appeared to follow a circadian patten such as microsatellite status (Fig. 1F, lower panel) and tumor mutational burden (TMB) (Figure S2A), with a trend that recapitulated the acro-phases found in time-series transcriptome analysis of colon cancer cell lines (Figure S1C).

We then investigated whether the molecular features of lung adenocarcinoma (LUAD), another well-characterized tumor type both in terms of molecular subtyping [6,14,26] and cell plasticity [11], follow a circadian pattern as well with the aim to further validate *in silico* our results. As also recently showed by our group, LUAD can be classified in two main molecular subtypes (C1 and nonC1) that are characterized by distinct epigenetic and genetic characteristics, cancer cell phenotypes, and cell plasticity states [14]. First, we interrogated GEO database and found a time-series transcriptome profiling of PC9 LUAD cell line we used as reference to remap static RWD of LUAD transcriptomes to pseudo time-series [17]. Then, we analyzed the TCGA-LUAD dataset (TCGA-LUAD cohort; N = 515; see methods) and reclassified RWD LUAD transcriptomes to realize a proxy time-series similarly to what we performed for CRC samples, in order to recapitulate a 24-hour time-course. Again, we observed a CCN expression profile of the pseudo time-series of LUAD samples coherent with the profile of PC9 cells time-series (Fig. 2A). Importantly, when we applied PSEA we found similar enrichment of signaling pathways as in CRC analysis, with cancer cell plasticity followed by cell proliferation/DNA damage repair mechanisms, once again correlating with TIM/ZEB1 axis regulation (Fig. 2B). Finally, we compared the distribution of C1 (high-cell plasticity status, more aggressive) and nonC1 (low-cell plasticity state, less aggressive) molecular subtypes and observed high correlation with ARNTL2 expression profile ($\boldsymbol{\rho}$=0.84; Fig. 2C). *ARNTL2* is a paralog to *ARNTL* and forms a heterodimer with *CLOCK* which then binds E-box element and induces target genes transcription. Interestingly, ARNTL2 was recently described to drive LUAD metastatic progression by favoring the acquisition of stem cell traits in LUAD cells [15]. In keeping with this, prognosis of LUAD patients segregated between 3-12h pseudo-time interval, where the expression of ARNTL2 is higher (Fig. 2D), significantly associated to worst outcome (Fig. 2E). Surprisingly, when we used the RWD CRC transcriptomes, the prognosis of patients segregated between 3-12h pseudo-time interval significantly associated to better outcome (Fig. 2F) regardless the expression pattern of ARNTL2 gene (Fig. 2D). As a matter of fact, we previously showed that low-expression of TIMELESS unleashes ZEB1 expression, which drives the acquirement of metastatic phenotypes in CRC [13]. Taken together, these findings reveal how the functioning of circadian clock circuit impinges on cancer-relevant pathways which correlate with tumor progression.

Discussion

Recent evidences challenged our actual knowledge that tumor molecular heterogeneity is a stochastic event. In addition, recent studies suggested also that ITH depends, at least in part, on the activation of mechanisms inducing cell plasticity states that influence the molecular architecture of cancer tissue and subtypes composition. Based on previous findings which link the circadian clock circuitry with the

modulation of mechanisms involved in cancer progression and tumor cell plasticity [13,27,28], here we showed that molecular heterogeneity of the two most deadly cancer types (i.e., CRC and LUAD) can be re-shaped as a result of time-ordering in accordance with inherent oscillations of biological processes driven by the molecular clockwork.

A current limit in the comprehension of molecular oncology comes from a rather static perspective of cancer molecular profile due to individual molecular snapshots of tumor samples. Indeed, such molecular profiles do not consider the inherent time-centered variability of almost all biological processes. To circumvent such limitations, we developed an approach which integrates the 'dynamic' dimension extrapolated from time-course analysis of cancer cell lines transcriptomes with 'static' transcriptomes of hundreds of human tumor samples. By Phase Set Enrichment Analysis (PSEA) we found distinct and suggestive time-phasing of cancer-relevant pathways, hinting time-related fluctuation of activity of genetic and epigenetic programs directing phenotypic changes and driving cell plasticity in cancer. Our findings suggested a certain temporal plasticity of cancer molecular subtypes in RWD of both CRC and LUAD, which definitely impacted patient prognosis and could be functional to the adaptation to anti-cancer treatments.

The main purpose of personalized medicine as well as conventional and chrono-modulated chemotherapy in the setting of oncology practice is to include time-oriented in addition to task-oriented procedures in clinical guidelines. The International Agency for Research on Cancer (IARC) classified night shift work-related circadian disruption as a plausible (group 2A) carcinogen for humans in 2007 [29]. We believe that our findings will pave the way toward reliable methodologies to assess the functioning of the circadian clock circuitry for a better ITH interpretation and identification of effective therapeutic strategies.

Declarations

Competing Interests statement: The authors declare no competing interests.

Acknowledgments

This work was supported by the Italian Minister of Health [Ricerca Corrente program 2022-2024 to F.B. and G.M.; Ricerca Finalizzata GR-2016-02363975 and CLEARLY to F.B.], by the Associazione Italiana Ricerca sul Cancro (AIRC) [IG-22827 to F.B.]. The study funders had no role in the design of the study, the collection, analysis, and interpretation of the data, the writing of the manuscript, and the decision to submit the manuscript for publication. All authors gave their consent to publication.

Authors' contributions

Conception and design: GM, FB. Statistical and bioinformatics analyses: VM, FB. Interpretation of data: VM, GM, FB. Writing, review of the manuscript: GM, FB. All authors approved the final version of the manuscript.

References

1. Sulli G, Lam MTY, Panda S. Interplay between Circadian Clock and Cancer: New Frontiers for Cancer Treatment. Trends Cancer 2019; **5**: 475–494.
2. Shafi AA, Knudsen KE. Cancer and the Circadian Clock. Cancer Res 2019; **79**: 3806–3814.
3. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med 2012; **366**: 883–92.
4. Maynard A, McCoach CE, Rotow JK, Harris L, Haderk F, Kerr DL *et al.* Therapy-Induced Evolution of Human Lung Cancer Revealed by Single-Cell RNA Sequencing. Cell 2020; **182**: 1232–1251.e22.
5. de Bruin EC, McGranahan N, Mitter R, Salm M, Wedge DC, Yates L *et al.* Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. Science 2014; **346**: 251–6.
6. Cancer Genome Atlas Research N. Comprehensive molecular profiling of lung adenocarcinoma. Nature 2014; **511**: 543–50.
7. Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Soneson C *et al.* The consensus molecular subtypes of colorectal cancer. Nat Med 2015; **21**: 1350–1356.
8. Marisa L, Blum Y, Taieb J, Ayadi M, Pilati C, Le Malicot K *et al.* Intratumor CMS Heterogeneity Impacts Patient Prognosis in Localized Colon Cancer. Clin Cancer Res 2021; **27**: 4768–4780.
9. Siravegna G, Mussolin B, Buscarino M, Corti G, Cassingena A, Crisafulli G *et al.* Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients. Nat Med 2015; **21**: 795–801.
10. Dawson S-J, Tsui DWY, Murtaza M, Biggs H, Rueda OM, Chin S-F *et al.* Analysis of circulating tumor DNA to monitor metastatic breast cancer. N Engl J Med 2013; **368**: 1199–1209.
11. Marjanovic ND, Hofree M, Chan JE, Canner D, Wu K, Trakala M *et al.* Emergence of a High-Plasticity Cell State during Lung Cancer Evolution. Cancer Cell 2020; **38**: 229–246.e13.
12. Zhang Y, Donaher JL, Das S, Li X, Reinhardt F, Krall JA *et al.* Genome-wide CRISPR screen identifies PRC2 and KMT2D-COMPASS as regulators of distinct EMT trajectories that contribute differentially to metastasis. Nat Cell Biol 2022; **24**: 554–564.
13. Colangelo T, Carbone A, Mazzarelli F, Cuttano R, Dama E, Nittoli T *et al.* Loss of circadian gene Timeless induces EMT and tumor progression in colorectal cancer via Zeb1-dependent mechanism. Cell Death Differ 2022. doi:10.1038/s41418-022-00935-y.
14. Melocchi V, Dama E, Mazzarelli F, Cuttano R, Colangelo T, Di Candia L *et al.* Aggressive early-stage lung adenocarcinoma is characterized by epithelial cell plasticity with acquirement of stem-like traits and immune evasion phenotype. Oncogene 2021; **40**: 4980–4991.
15. Brady JJ, Chuang C-H, Greenside PG, Rogers ZN, Murray CW, Caswell DR *et al.* An Arntl2-Driven Secretome Enables Lung Adenocarcinoma Metastatic Self-Sufficiency. Cancer Cell 2016; **29**: 697–710.

16. El-Athman R, Fuhr L, Relógio A. A Systems-Level Analysis Reveals Circadian Regulation of Splicing in Colorectal Cancer. EBioMedicine 2018; **33**: 68–81.
17. Nakata A, Yoshida R, Yamaguchi R, Yamauchi M, Tamada Y, Fujita A *et al.* Elevated β-catenin pathway as a novel target for patients with resistance to EGF receptor targeting drugs. Sci Rep 2015; **5**: 13076.
18. Cancer Genome Atlas N. Comprehensive molecular characterization of human colon and rectal cancer. Nature 2012; **487**: 330–7.
19. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Science signaling 2013; **6**: pl1.
20. Wu G, Anafi RC, Hughes ME, Kornacker K, Hogenesch JB. MetaCycle: an integrated R package to evaluate periodicity in large scale data. Bioinformatics 2016; **32**: 3351–3353.
21. Corsello SM, Nagari RT, Spangler RD, Rossen J, Kocak M, Bryan JG *et al.* Discovering the anticancer potential of non-oncology drugs by systematic viability profiling. Nat Cancer 2020; **1**: 235–248.
22. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005; **102**: 15545–50.
23. Fuhr L, Abreu M, Carbone A, El-Athman R, Bianchi F, Laukkanen MO *et al.* The interplay between colon cancer cells and tumour-associated stromal cells impacts the biological clock and enhances malignant phenotypes. Cancers 2019; **11**. doi:10.3390/cancers11070988.
24. Zhang R, Podtelezhnikov AA, Hogenesch JB, Anafi RC. Discovering Biology in Periodic Data through Phase Set Enrichment Analysis (PSEA). J Biol Rhythms 2016; **31**: 244–257.
25. Zhan T, Rindtorff N, Boutros M. Wnt signaling in cancer. Oncogene 2017; **36**: 1461–1473.
26. Roh W, Geffen Y, Cha H, Miller M, Anand S, Kim J *et al.* High-Resolution Profiling of Lung Adenocarcinoma Identifies Expression Subtypes with Specific Biomarkers and Clinically Relevant Vulnerabilities. Cancer Res 2022; **82**: 3917–3931.
27. Fuhr L, El-Athman R, Scrima R, Cela O, Carbone A, Knoop H *et al.* The Circadian Clock Regulates Metabolic Phenotype Rewiring Via HKDC1 and Modulates Tumor Progression and Drug Response in Colorectal Cancer. *EBioMedicine* 2018; **33**: 105–121.
28. Parascandolo A, Bonavita R, Astaburuaga R, Sciuto A, Reggio S, Barra E *et al.* Effect of naive and cancer-educated fibroblasts on colon cancer cell circadian growth rhythm. Cell Death Dis 2020; **11**: 289.
29. Stevens RG, Hansen J, Costa G, Haus E, Kauppinen T, Aronson KJ *et al.* Considerations of circadian impact for defining 'shift work' in cancer studies: IARC Working Group Report. Occup Environ Med 2011; **68**: 154–162.

Figures

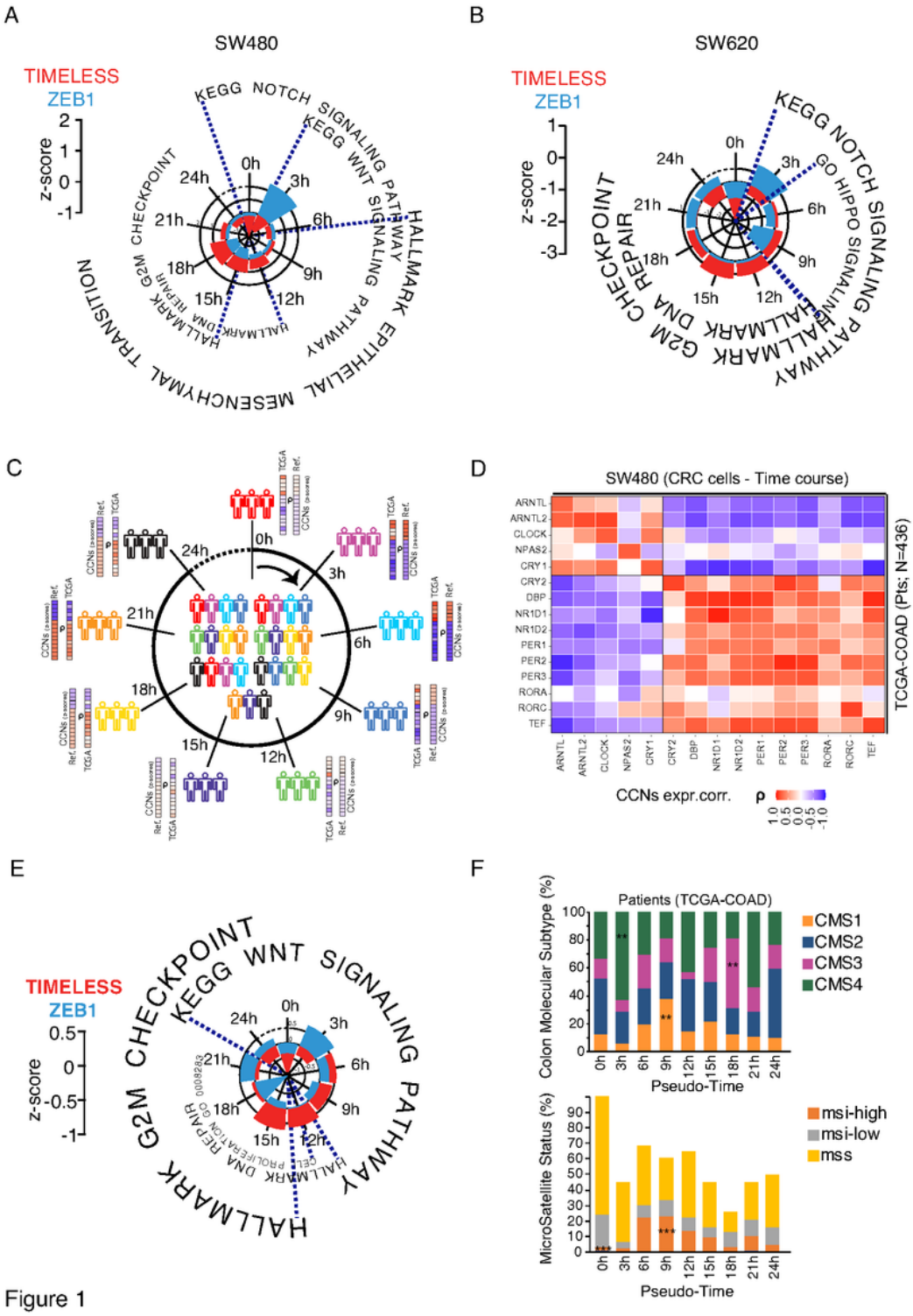

Figure 1

Figure 1

A-B) TIMELESS and ZEB1 expression profile in time-course analysis of SW480 and SW620 cell lines, together with results of the Phase Set Enrichment Analysis (PSEA). Each circle represents the z-scores values of TIMLESS and ZEB1 expression and are as per the legend. The inner the circle the lower the z-score. Color codes are as per the legend.

C) Schematic representation of the approach we developed to remap static human cancer transcriptome (TCGA) to pseudo time-series with time-qualified points of cancer cell lines used as reference (Ref.). The CCN expression pattern of TCGA samples and cell lines (Ref.) is also shown as an example (i.e., using an arbitrary scale) by means of heatmaps in the various pseudo-time points. **D)** Correlation heatmap of expression profile of core-clock genes (CCN) in SW480 cells time-course and TCGA-COAD samples. Color codes are as per the legend and represent the Pearson product-moment correlation (ρ) of z-score values.

E) TIMELESS and ZEB1 expression profile in the pseudo-time analysis of TCGA-COAD samples, together with results of Phase Set Enrichment Analysis (PSEA). Each circle represents the z-scores values of TIMELESS and ZEB1 expression and are as per the legend. The inner the circle the lower the z-score. Color codes are as per the legend. **F)** Upper panel, distribution of Colon Molecular Subtypes (CMS) in the pseudo-time analysis of TCGA-COAD samples. Lower panel, distribution of Microsatellite Status in the pseudo-time analysis of TCGA-COAD samples. P-values were computed by post hoc test following chi-square; ** $P \leq 0.01$; ***$P \leq 0.001$.

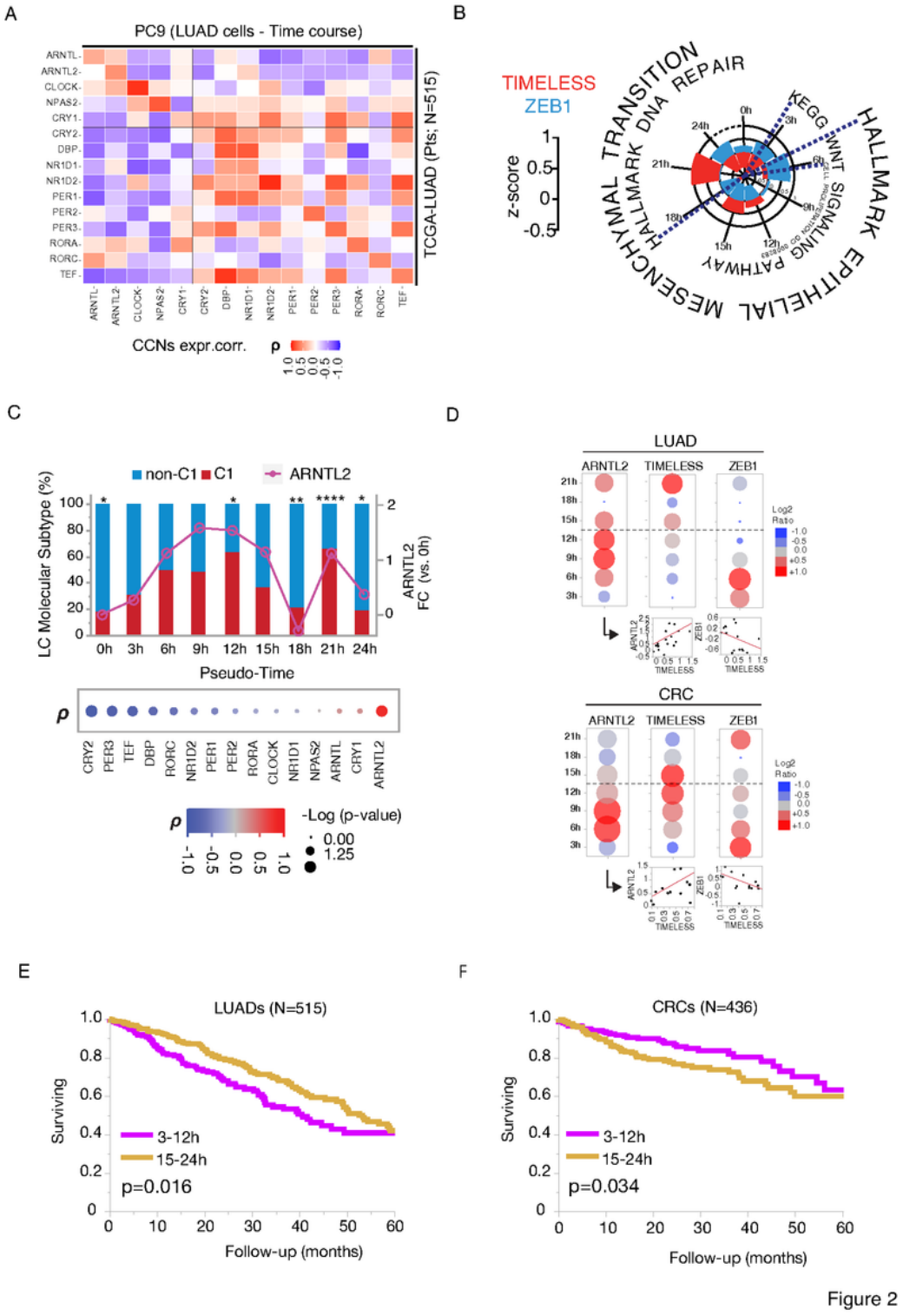

Figure 2

Figure 2

A) Correlation heatmap of expression profile of core-clock genes (CCN) in time-course analysis of PC9 cells and pseudo-time analysis of TCGA-LUAD samples. Color codes are as per the legend and represent the Pearson product-moment correlation (**ρ**) of z-score values. **B)** TIMELESS and ZEB1 expression profile in the pseudo-time analysis of TCGA-LUAD samples, together with results of Phase Set Enrichment Analysis (PSEA). Each circle represents the z-score value of TIMLESS and ZEB1 expression and are as

per the legend. The inner the circle the lower the z-score. Color codes are as per the legend. **C)** Distribution of Lung Cancer Molecular Subtypes (C1 and nonC1) in the pseudo-time analysis of TCGA-LUAD samples, together with the expression profile of ARNTL2 gene (secondary Y-axes; FC, fold change). P-values were computed by post hoc test following chi-square; *P≤0.05, ** P≤ 0.01, and ****P≤0.0001. Lower panel, bubble plot showing the Pearson correlation scores (**ρ**) of CCN genes with C1 subtype distribution; color codes and size of bubbles are as per the legend. **D)** Bubble plot analysis of ARNTL2, TIMELESS and ZEB1 expression profile in the pseudo-time series in the TCGA-LUAD dataset (upper panel) or in the TCGA-COAD (CRC) dataset (lower panel). Color codes are as per the legend. Size of bubbles are directly proportional to Log_2 Ratio of expression as per the legend. Correlation plots of the indicated genes are also shown (underneath the two panels). Dashed line indicates the two pseudo-time interval periods used to stratify TCGA samples as in (E). **E-F)** Survival analysis of TCGA-LUAD patients (E) and TCGA-COAD (CRC) patients (F) stratified by the indicated pseudo-time interval periods. Color codes are as per the legend. P-values were computed by log-rank test.

Supplementary Files

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

- SupplementalFigure1mazzoccolietal.pdf
- SupplementalFigure2mazzoccolietal.pdf
- SupplementalMaterialMazzoccolietal.docx
- TableS1mazzoccolietal.xlsx