non-coding RNA

MDPI

Brief Report

Is Evolutionary Conservation a Useful Predictor for Cancer Long Noncoding RNAs? Insights from the Cancer LncRNA Census 3

Adrienne Vancura [1,2,3,4], Alejandro H. Gutierrez [1,3], Thorben Hennig [4], Carlos Pulido-Quetglas [1,2,3], Frank J. Slack [4], Rory Johnson [1,3,5,6,*] and Simon Haefliger [1,3,*]

1 Department of Medical Oncology, Inselspital, Bern University Hospital, University of Bern, 3010 Bern, Switzerland
2 Graduate School of Cellular and Biomedical Sciences, University of Bern, 3012 Bern, Switzerland
3 Department for BioMedical Research, University of Bern, 3008 Bern, Switzerland
4 HMS Initiative for RNA Medicine, Department of Pathology, Beth Israel Deaconess Medical Center Cancer Center, Harvard Medical School, Boston, MA 02215, USA
5 School of Biology and Environmental Science, University College Dublin, D04 V1W8 Dublin, Ireland
6 Conway Institute of Biomedical and Biomolecular Research, University College Dublin, D04 V1W8 Dublin, Ireland
* Correspondence: rory.johnson@ucd.ie (R.J.); simon.haefliger@insel.ch (S.H.)

Abstract: Evolutionary conservation is a measure of gene functionality that is widely used to prioritise long noncoding RNAs (lncRNA) in cancer research. Intriguingly, while updating our Cancer LncRNA Census (CLC), we observed an inverse relationship between year of discovery and evolutionary conservation. This observation is specific to cancer over other diseases, implying a sampling bias in the selection of lncRNA candidates and casting doubt on the value of evolutionary metrics for the prioritisation of cancer-related lncRNAs.

Keywords: long noncoding RNA; cancer; conservation

Citation: Vancura, A.; Gutierrez, A.H.; Hennig, T.; Pulido-Quetglas, C.; Slack, F.J.; Johnson, R.; Haefliger, S. Is Evolutionary Conservation a Useful Predictor for Cancer Long Noncoding RNAs? Insights from the Cancer LncRNA Census 3. *Non-Coding RNA* **2022**, *8*, 82. https://doi.org/10.3390/ncrna8060082

Academic Editors: Balaji Krishnamachary and Oliver Treeck

Received: 8 November 2022
Accepted: 30 November 2022
Published: 7 December 2022

1. Introduction

Long noncoding RNAs play central, functional roles in cancer and are being developed as targets for RNA therapeutics [1–4]. Given the high costs of drug discovery studies, and the frequency of late-stage failures, it is imperative to collect and effectively prioritise lncRNAs with the greatest therapeutic value. We here present the third version of the successful Cancer lncRNA Census (CLC3), covering publications from the period from 2019 to late 2020, comprising altogether 702 unique GENCODE-annotated lncRNAs with functional cancer roles based on a variety of evidence.

2. Results

CLC3 incorporates and extends previous versions (CLC1, 118 lncRNAs, CLC2 374 lncRNAs) [1,2]. In addition to its size, CLC3 now incorporates for the first time lncRNAs involved in chemoresistance, with 10% of CLC lncRNAs exhibiting this functionality (Figure 1A). We previously observed that CLC lncRNAs carry a range of features distinguishing them from other lncRNAs. Amongst these were elevated rates for several measures of evolutionary conservation, similar to those previously observed for protein-coding cancer genes [1,5]. First, we evaluated the confidence level of experimental support for CLC genes, finding these to be consistent between versions with roughly 40% of lncRNAs validated by the highest-confidence in vivo evidence (Figure 1A). Therefore, any differences observed between CLC versions is not likely to arise from differences in confidence regarding their disease roles. Next, we comprehensively evaluated a range of features of CLC versions, comparing non-redundant gene sets to non-cancer lncRNAs ("nonCLC") (Figure 1B). For comparison, we also compared a collection of disease-associated lncRNAs, from which cancer genes were removed (EVlncRNA) [6]. All CLC versions and

EVlncRNAs display elevated levels of gene expression, expression ubiquity, overall gene length, spliced RNA length and proximity to nearest protein-coding genes (Figure 1B). Surprisingly, however, we noticed that CLC3 lncRNAs are not more evolutionarily conserved compared to other non-cancer lncRNAs (arrows). This is true not only for two different measures of conservation from the widely used PhastCons measure (average base-level score and percentage of exon coverage by conserved elements), but also for the promoter (average base-level), for which particularly elevated conservation has been observed in lncRNAs [7,8]. A more detailed gene-level inspection supported these findings (Figure 1C–F), showing a pronounced trend for the CLC3 lncRNAs to have comparable or even lower conservation than lncRNAs in general.

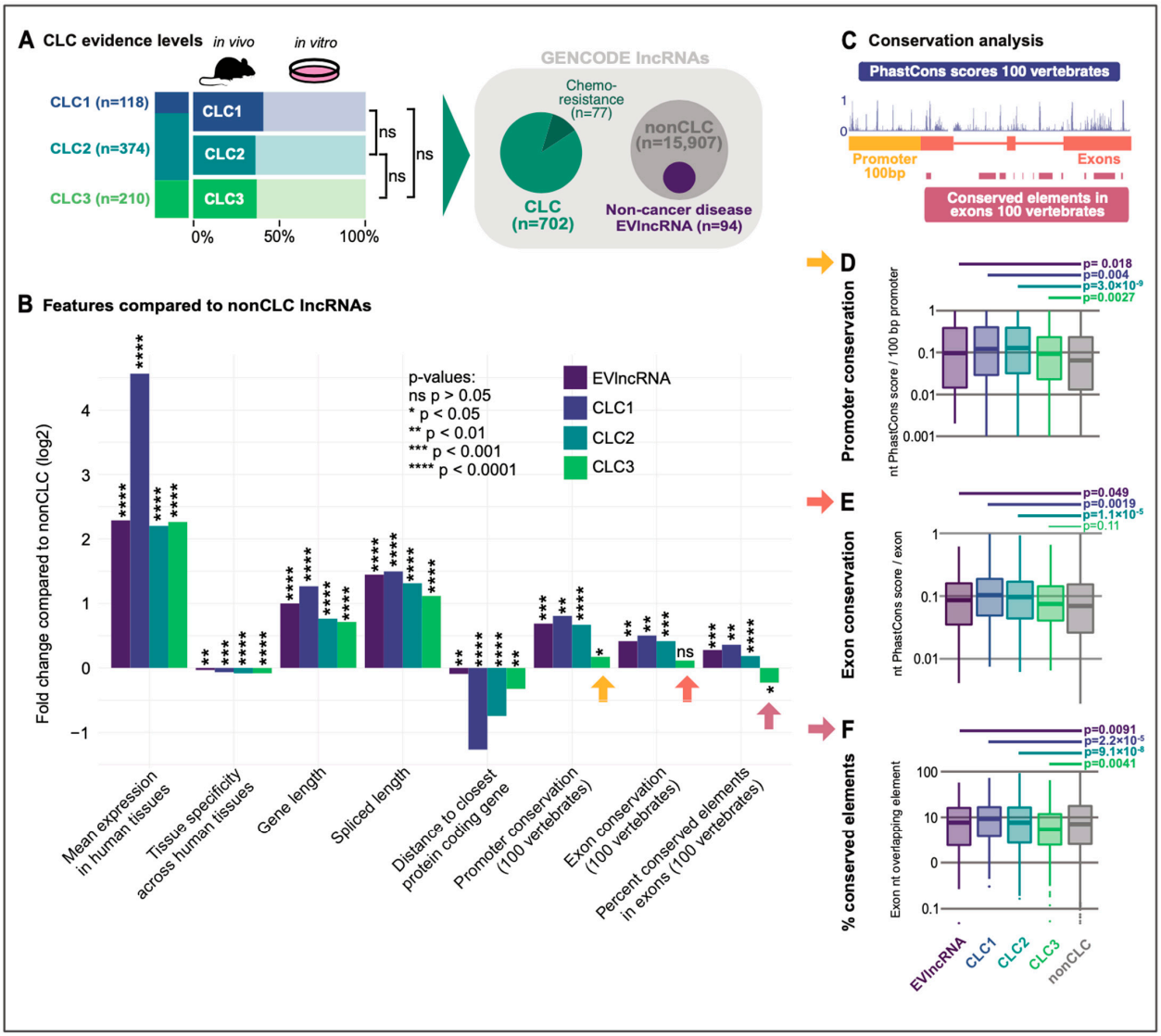

Figure 1. (**A**) Evidence levels for functional lncRNAs in CLC database versions. Dark colour indicates number of lncRNAs tested in an in vivo setting. No significant (ns) difference of in vivo enrichment is observed across the datasets. The full CLC consists of 702 lncRNAs with 77 lncRNAs exhibiting chemoresistance mechanisms. GENCODE lncRNAs are subdivided in CLC and nonCLC genes for further comparison. Non-cancer disease EVlncRNAs are nonCLC genes indicating a disease functionality but not represented in the CLC database. (**B**) Features in datasets compared to nonCLC lncRNAs using LnCompare. (**C**) Overview of conservation analysis using 100 vertebrates comparisons. (**D**) Promoter conservation analysis for all datasets. (**E**) Exon analysis for all datasets. (**F**) Conserved elements analysis for all datasets.

To strengthen these findings, we used an alternative method to evaluate evolutionary conservation: the existence of orthologous lncRNA genes in other species. Using the tool ConnectOR [9], we searched for orthologues of human lncRNAs in chimpanzees and mice (see Methods). Overall, we identified orthologues for 4102 and 4493 lncRNAs in chimpanzees and mice, respectively (lower rates in chimpanzees likely reflect less mature lncRNA annotations). Consistent with previous results, we observed that CLC3 lncRNAs have a significantly lower chance of having an identifiable orthologue than CLC1 and CLC2, at a level comparable to nonCLC lncRNAs (Figure 2A).

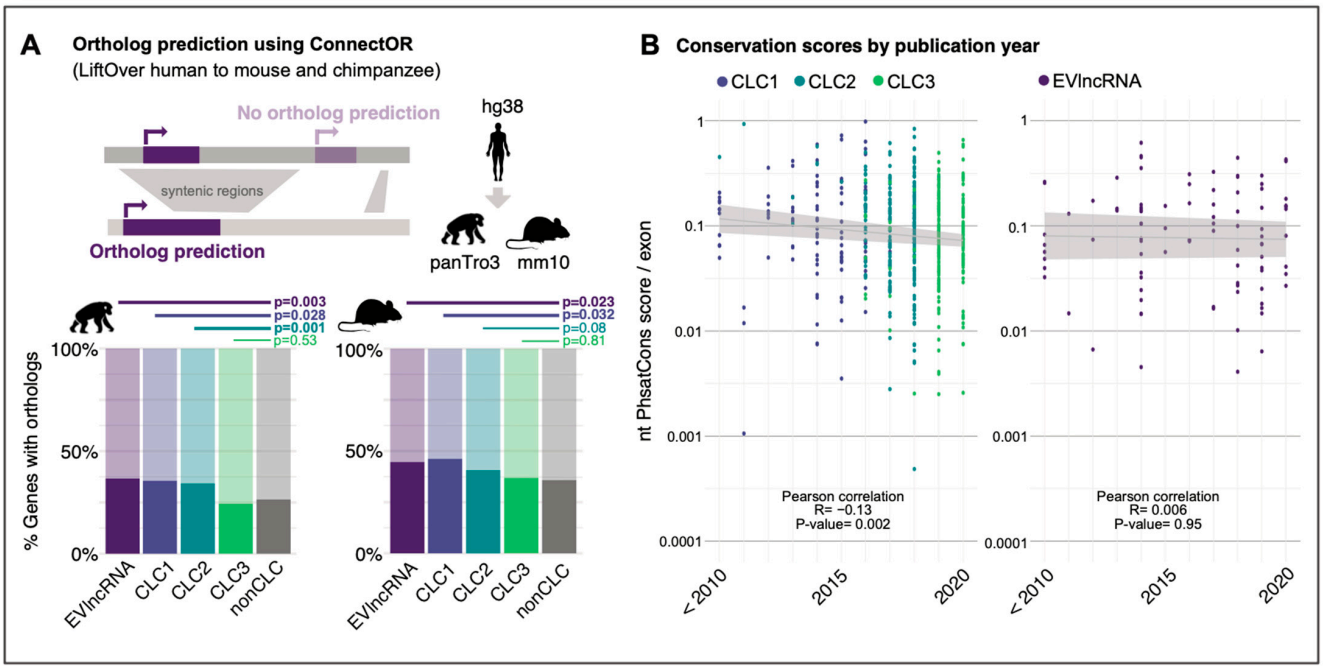

Figure 2. (**A**) Ortholog prediction using ConnectOR for chimpanzees (left) and mice (right). (**B**) Exon conservation scores by publication year for CLC versions (left) and EVlncRNAs (right).

Given that CLC3 lncRNAs were collected most recently, we hypothesised that the observed trend arose from a relationship between conservation and the moment when the lncRNA was studied. Indeed, we observed a significant negative correlation between conservation and year of discovery (Figure 2B, left). This trend appears to be specific to cancer, because EVlncRNAs from other diseases do not display this behaviour (Figure 2B, right). In other words, as time goes on, researchers are turning their attention to less conserved lncRNAs that nevertheless play functional cancer roles.

3. Discussion

In summary, we have presented the latest version of the Cancer lncRNA Census, a carefully curated resource of functional cancer-associated lncRNAs intended to serve as a useful true positive dataset for large-scale discovery and as a source for therapeutic development. We have made the surprising observation that evolutionary conservation of collected lncRNAs decreases with year of publication, and that recently published cancer lncRNAs have conservation levels similar to lncRNAs in general. Although previous studies have shown that protein-coding cancer genes are conserved more on average [1], it remains possible that a similar phenomenon affects these genes. This phenomenon appears to be specific for cancer, since catalogues of lncRNAs playing roles in other diseases do not display the same trend. Evolutionary conservation is a longstanding and widely used criterion for the selection of candidate lncRNAs for follow-up study [10–13]. However, there are numerous examples of functionally validated, non-conserved lncRNAs [14,15]. Supporting these findings, recent unbiased large-scale functional screens found no relationship between

conservation and hits [16]. The apparent specificity to cancer raises the possibility that tumours exploit lncRNA sequences that have no natural function. Indeed, a similar model was recently proposed by Adnane and colleagues [17]. These findings suggest that the scientific community may have suffered an unconscious bias in selecting evolutionarily conserved lncRNAs for study, thereby reinforcing the impression that conservation is a useful criterion for candidate selection [1]. This work suggests that filtering by evolutionary conservation may result in omission of important cancer related lncRNAs.

4. Materials and Methods

4.1. Literature Search and LnCompare for Feature and Repeat Analysis

This analysis was performed as described in CLC2 [2] and the full CLC3 gene list can be found here: https://zenodo.org/record/7075104#.YyCB0C1Q3T8 (accessed on 13 September 2022).

4.2. EVlncRNA Non-Cancer lncRNA Dataset

EVlncRNAs were downloaded from and were sorted for ENSG (GENCODE v28) and overlayed with CLC genes to exclude functional cancer lncRNAs.

4.3. Conservation Scores

Exons were collapsed using exon info from GENCODE v28, and PhastCons exon conservation scores (PhastCons100way.UCSC.hg28) were generated according to Vancura et al., 2021 using Bioconductor Genomic Scores R package.

PhastConsElements 100way were downloaded from genome.ucsc.edu using the table browser. PhastConsElements were intersected with datasets using intersectBed. Statistical evaluation was performed using Wilcoxon test.

4.4. Publication Year

PMID years for each lncRNA were extracted using the code from https://www.ncbi.nlm.nih.gov/books/NBK179288/ (accessed on 5 March 2022) and earliest publication years were used for subsequent analysis.

4.5. Orthologue Prediction

Orthologue prediction was performed using ConnectOR (https://github.com/Carlospq/ConnectOR (accessed on 20 May 2022)) based on LiftOver of syntenic regions from human (hg38) to mouse (mm10) or chimpanzee (panTro3). ConnectOR results "not lifted" and "one to none" were characterised as no orthology prediction. Statistical evaluation was performed using Fisher's one-sided *t*-test.

Author Contributions: R.J. conceived the project. A.V. and A.H.G. performed manual annotation of CLC3 genes. A.V. and T.H. performed evolutionary analysis; C.P.-Q. provided ortholog prediction tool. R.J., A.V., F.J.S. and S.H. drafted the manuscript and prepared the figures. All authors have read and agreed to the published version of the manuscript.

Funding: We thank Basak Ginsbourger (DMBR) for administrative support and Willy Hofstetter and Patrick Furer (DMBR) for logistical support. All computations were performed on the Bern Interfaculty Bioinformatics Unit computing cluster maintained by Rémy Bruggmann and Pierre Berthier. This work was funded by the Swiss National Science Foundation through the National Centre of Competence in Research (NCCR) "RNA & Disease" (51NF40-182880), project funding "The elements of long noncoding RNA function" (31003A_182337), Sinergia project "Regenerative strategies for heart disease via targeting the long noncoding transcriptome" (173738); SNF Doc.mobility project (P1BEP3_199932), by the Medical Faculty of the University of Bern and Inselspital University Hospital of Bern; by the Helmut Horten Stiftung, Swiss Cancer Research Foundation (4534-08-2018); by the Werner and Hedy Berger Janser—Foundation for cancer research; by Science Foundation Ireland through Future Research Leaders award 18/FRL/6194; and the NIH [R35 CA232105].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The CLC3 gene list can be found here: https://zenodo.org/record/7075104#.YyCB0C1Q3T8.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Carlevaro-Fita, J.; Lanzós, A.; Feuerbach, L.; Hong, C.; Mas-Ponte, D.; Pedersen, J.S.; Abascal, F.; Amin, S.B.; Bader, G.D.; Barenboim, J.; et al. Cancer LncRNA Census reveals evidence for deep functional conservation of long noncoding RNAs in tumorigenesis. *Commun. Biol.* **2020**, *3*, 56. [CrossRef] [PubMed]
2. Vancura, A.; Lanzós, A.; Bosch-Guiteras, N.; Esteban, M.T.; Gutierrez, A.H.; Haefliger, S.; Johnson, R. Cancer LncRNA Census 2 (CLC2): An enhanced resource reveals clinical features of cancer lncRNAs. *NAR Cancer* **2021**, *3*, zcab013. [CrossRef] [PubMed]
3. Arun, G.; Diermeier, S.D.; Spector, D.L. Therapeutic Targeting of Long Non-Coding RNAs in Cancer. *Trends Mol. Med.* **2018**, *24*, 257–277. [CrossRef] [PubMed]
4. Winkle, M.; El-Daly, S.M.; Fabbri, M.; Calin, G.A. Noncoding RNA therapeutics—Challenges and potential solutions. *Nat. Rev. Drug Discov.* **2021**, *20*, 629–651. [CrossRef] [PubMed]
5. Furney, S.J.; Madden, S.F.; Kisiel, T.A.; Higgins, D.G.; Lopez-Bigas, N. Distinct patterns in the regulation and evolution of human cancer genes. *In Silico Biol.* **2008**, *8*, 33–46. [PubMed]
6. Zhou, B.; Ji, B.; Liu, K.; Hu, G.; Wang, F.; Chen, Q.; Yu, R.; Huang, P.; Ren, J.; Guo, C.; et al. EVLncRNAs 2.0: An updated database of manually curated functional long non-coding RNAs validated by low-throughput experiments. *Nucleic Acids Res.* **2021**, *49*, D86–D91. [CrossRef]
7. Guttman, M.; Amit, I.; Garber, M.; French, C.; Lin, M.F.; Feldser, D.; Huarte, M.; Zuk, O.; Carey, B.W.; Cassady, J.P.; et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **2009**, *458*, 223–227. [CrossRef] [PubMed]
8. Chodroff, R.A.; Goodstadt, L.; Sirey, T.M.; Oliver, P.L.; Davies, K.E.; Green, E.D.; Molnár, Z.; Ponting, C.P. Long noncoding RNA genes: Caaonservation of sequence and brain expression among diverse amniotes. *Genome Biol.* **2010**, *11*, R72. [CrossRef] [PubMed]
9. Pulido-Quetglas, C. GitHub—Carlospq/ConnectOR: Multiple Species Orthology Finder. Available online: https://github.com/Carlospq/ConnectOR (accessed on 20 June 2022).
10. Siepel, A.; Bejerano, G.; Pedersen, J.S.; Hinrichs, A.S.; Hou, M.; Rosenbloom, K.; Clawson, H.; Spieth, J.; Hillier, L.D.W.; Richards, S.; et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **2005**, *15*, 1034–1050. [CrossRef] [PubMed]
11. Ulitsky, I.; Shkumatava, A.; Jan, C.H.; Sive, H.; Bartel, D.P. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* **2011**, *147*, 1537–1550. [CrossRef] [PubMed]
12. Iyer, M.K.; Niknafs, Y.S.; Malik, R.; Singhal, U.; Sahu, A.; Hosono, Y.; Barrette, T.R.; Prensner, J.R.; Evans, J.R.; Zhao, S.; et al. The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* **2015**, *47*, 199–208. [CrossRef] [PubMed]
13. Ponting, C.P. Biological function in the twilight zone of sequence conservation. *BMC Biol.* **2017**, *15*, 71. [CrossRef] [PubMed]
14. Ruan, X.; Li, P.; Chen, Y.; Shi, Y.; Pirooznia, M.; Seifuddin, F.; Suemizu, H.; Ohnishi, Y.; Yoneda, N.; Nishiwaki, M.; et al. In vivo functional analysis of non-conserved human lncRNAs associated with cardiometabolic traits. *Nat. Commun.* **2020**, *11*, 45. [CrossRef]
15. Cao, T.; O'Reilly, M.E.; Selvaggi, C.; Cynn, E.; Lumish, H.; Xue, C.; Jha, A.; Reilly, M.P.; Foulkes, A.S. Cis-regulated expression of non-conserved lincRNAs associates with cardiometabolic related traits. *J. Hum. Genet.* **2022**, *67*, 307–310. [CrossRef] [PubMed]
16. Liu, S.J.; Horlbeck, M.A.; Cho, S.W.; Birk, H.S.; Malatesta, M.; He, D.; Attenello, F.J.; Villalta, J.E.; Cho, M.Y.; Chen, Y.; et al. CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science* **2017**, *355*, eaah7111. [CrossRef]
17. Adnane, S.; Marino, A.; Leucci, E. LncRNAs in human cancers: Signal from noise. *Trends Cell Biol.* **2022**, *32*, 565–573. [CrossRef]