

This is the author-created version of the following work:

Wong, Justin J.L., and Schmitz, Ulf (2022) *Intron retention: importance, challenges, and opportunities*. Trends in Genetics, 38 (8) pp. 789-792.

Access to this file is available from:

<https://researchonline.jcu.edu.au/75294/>

© 2022 Elsevier Ltd. All rights reserved.

Please refer to the original source for the final version of this work:

<https://doi.org/10.1016/j.tig.2022.03.017>

Intron retention: importance, challenges and opportunities

Justin J.-L. Wong^{1,2}, Ulf Schmitz^{3,4}

¹ Epigenetics and RNA Biology Program Centenary Institute, The University of Sydney, Camperdown 2050, Australia

² Faculty of Medicine and Health, The University of Sydney, Camperdown 2050, Australia

³ Department of Molecular & Cell Biology, College of Public Health, Medical & Vet Sciences, James Cook University, Townsville 4811, Australia

⁴ Centre for Tropical Bioinformatics and Molecular Biology, Australian Institute of Tropical Health and Medicine, James Cook University, Cairns 4878, Australia

Abstract

Recent landmark discoveries have underpinned the physiological importance of intron retention across multiple domains of life and revealed an unexpected breadth of functions in a large variety of biological processes. Despite significant progress in the field some challenges remain. Once solved, opportunities will arise for discovering more functions of intron retention.

Keywords

alternative splicing, RNA processing, splicing regulation, epi-transcriptomics, intron retaining transcripts

Introduction

Intron retention (IR) is a well-conserved form of alternative splicing that has been studied extensively in plants, fungi, insects, and viruses. Its importance in regulating gene expression and essential biological processes in mammalian cells has only recently been acknowledged [1-3]. Yet, ever new studies keep surfacing that describe the involvement of IR in a multitude of cellular processes, some of which are surprising given that IR was for a long time considered transcriptional noise [4]. Figure 1 summarizes key concepts in relation to the importance of IR, challenges and opportunities.

Why is intron retention important?

A widely accepted function of IR is to regulate gene expression via nonsense-mediated decay (NMD), triggered by intronic premature termination codons (PTCs), or IR transcript detention and degradation in the nucleus [1,2,5,6]. IR-induced post-transcriptional gene regulation can be observed in stages of the cell cycle, cell differentiation, apoptosis, stress response, and reproduction [4]. However, there are noticeable intercellular differences in the abundance and location of intron-retaining transcripts (IRTs). While less frequently observed in muscle and embryonic stem cells in human and mouse there is a high prevalence of IRTs in neural, adipose, and immune cells. IR is well conserved across vertebrate species but seems more widespread in species with fewer protein coding genes wherein transcriptomic complexity is preserved by enhanced alternative splicing programs [3,7]. Retention of both major and minor introns are important in regulating gene expression and physiological processes [8].

IRTs that evade degradation due to the lack of PTCs, can be translated into novel protein isoforms with impact on protein function. This is the predominant fate of IRTs in plants. The specific function of retained introns is often determined by the cellular localization of the IRTs. Intron detention refers to IRTs residing in the nucleus. These transcripts can be degraded by mechanisms independent of NMD or await further processing (cytoplasmic export, splicing) upon specific stimuli [1,2]. IR can, in fact, stabilize nuclear transcripts, thereby extending their half-life [1].

IR is an important mechanism to respond to environmental stress such as thermal, nutritional, oxidative or CO₂-induced stress, and is a response to other extracellular stimuli and infection in plants [9]. IR affects genes that are implicated in temperature-dependent sex determination in multiple species including fish and reptiles [10], in growth and development of plants, in fission yeast meiosis, and in mammalian spermatogenesis [1].

Although not thoroughly verified yet, IR may contribute to the synthesis of novel non-coding RNAs (with additional structural domains). Retained introns can be carrier of cellular localization signals and signals receptive to alternative polyadenylation generating truncated proteins [4].

Studying IR is also important in the context of diseases. Aberrant IR patterns have been observed in diverse cancers [11]. Some recent studies attempted to shed light on the role of IR in cancer pathogenesis and mechanisms that lead to aberrant and pathologic IR in cancer and other diseases. It was found that IR can cause (i) tumor suppressor gene inhibition, (ii) oncoprotein expression, (iii) neoepitope synthesis, or (iv) therapy resistance, and serve as (v) biomarker, (vi) present a therapeutic vulnerability, or (vii) reduce cancer cell proliferation [12]. These sometimes disease-specific disparities stem from PTCs introduced via IR-associated single nucleotide variants (i and vi), novel functions of translated IRTs (ii), intron-derived peptides with major histocompatibility complex binding affinity (iii), IR-induced NMD in antigens (iv), IR profiles associated with disease outcomes (v), and therapeutically-restored spliceosome function (vi).

Challenges in intron retention discovery and understanding its roles

The discovery of IR requires a larger number of mRNA sequencing reads to detect affected genes that express a large percentage of IRTs subject to degradation. Over 100 million reads would be ideal as recommended by ENCODE for alternative splicing analysisⁱ. Poly-A enrichment methods should also be applied, instead of ribosomal RNA depletion, to avoid contamination with pre-mRNAs. Researchers should recognise that utilizing suboptimal sequencing data will inaccurately measure IR. Datasets from most publicly available datasets were generated without the intention to investigate IR and do not have the appropriate depth. For example, analysis using the lower depth sequencing data (<50M reads) from the Cancer Genome Atlas Research Network (TCGA) is likely to underestimate the levels of IR detected [11]. Results of IR analysis using ribosomal RNA-depleted RNA samples should be interpreted with caution and validated using a poly-A enrichment method. Strand-specific sequencing protocols are encouraged to exclude antisense RNAs, which may confound accurate detection of IR [13].

From a bioinformatics perspective, mapping tools need to be customized to cope with multi-mapping reads and repetitive sequences often present in retained introns, that would be discarded by standard mapping methods [13]. It is also important to ensure that reads spanning exon-intron junctions are sufficiently detected to avoid erroneous calling of intron-

derived non-coding RNAs as transcripts retaining introns. Recent pipelines for IR discovery include filters to circumvent these issues [13].

IR detection using short-read sequencing is unable to verify whether multiple introns are retained in the same transcript. In these cases, it would be necessary to perform long-read sequencing, using PacBio or Nanopore technologies, to determine the exact combination of exons and introns in the final transcripts. This step will also inform whether the IRTs contain in-frame PTCs to understand their fate. However, further improvement of long-read sequencing technologies is required to circumvent their limitation in IR detection due to relatively low depth, high error rates and the propensity towards detecting short transcripts.

IR can lead to diverse impacts on transcripts and proteins, making it challenging to understand its role in specific biological contexts. Given that many factors can regulate IR including RNA-binding proteins, epigenetic changes, transcriptional rate (see Box 1), it would be necessary to first identify the cause of specific IR events before functional studies can be performed. Nevertheless, it is challenging to efficiently modulate IR in laboratory settings without causing off-target effects as modulating these factors is likely to cause other transcriptional and post-transcriptional changes. Given that IRTs are often rapidly degraded via NMD or other pathways, it would be necessary to block relevant RNA degradation pathways to robustly detect them [2,5]. While NMD is the major mechanism that degrades IRTs in the cytoplasm, pathways occurring in the nucleus are largely unknown. Mechanisms that degrades IRTs may also operate in a cell type-specific manner. Recent studies indicate that the nuclear exosome machinery does not trigger degradation of IRTs in humans, although it has been reported in murine neuronal cells [2,5]. In addition, it remains unclear how IRTs are detained in the nucleus and are protected from degradation. We hypothesize that this process involves liquid-liquid phase separation that compartmentalises different transcripts for different molecular processes that occur within the phase-separated bodies.

Future directions in intron retention research

The mechanisms underpinning the regulation of IR in different biological contexts remain unclear. Nuclear-enriched small non-coding RNAs (including microRNAs) that bind near splice junctions in nascent RNAs indicate the possibility of their involvement in IR regulation [4], however, a direct effect remains to be tested experimentally. A recent report found that m⁶A RNA modifications accumulate in retained introns and that global depletion of m⁶A promotes splicing of these introns, which suggests an important role of RNA modifications in IR regulation [14]. It would be interesting to identify the mechanisms by which m⁶A directs IR beyond the few reported so far [15]. The factors that define the specificity of IR regulation by one mechanism or another, and the contribution of environmental cues to this process, remain to be uncovered.

Studies that explore the functions of IR are relatively few compared to those focusing on IR discovery. Many proposed functions of IR remain to be validated experimentally including IR providing templates to store and derive non-coding RNAs, and preserving alternative polyadenylation sites to generate truncated proteins with oncogenic potential. Another hypothesis suggests that IR may be a source of microRNA response elements that allow IRTs to act as microRNA decoy (ie. “sponge”) or competing endogenous RNA. The roles of IR in the pathogenesis of diverse human diseases including metabolic diseases, cardiovascular diseases and immune-related disorders remain to be determined.

Concluding remarks

Given the widespread of IR across taxa and functional implications IR bears, we strongly encourage the inclusion of algorithms for alternative splicing detection and quantification in all transcriptomic analyses. Considerations for IR analysis should begin at the experiment planning phase as appropriate choices for library preparation (e.g. poly-A enrichment) and sequencing (e.g. sufficient read depth and/or long reads) should be made upfront. By doing so, investigators will be rewarded with high-resolution data, disclosing the full breadth of transcriptomic complexity within their samples. This step will facilitate IR detection at high sensitivity and an accurate quantitation of IR in protein coding and non-coding genes. Nevertheless, IR analyses require a careful approach that navigates around common pitfalls, which could lead to false IR predictions. Hence our advice is to closely monitor technological and algorithmic advances as the field of IR progresses.

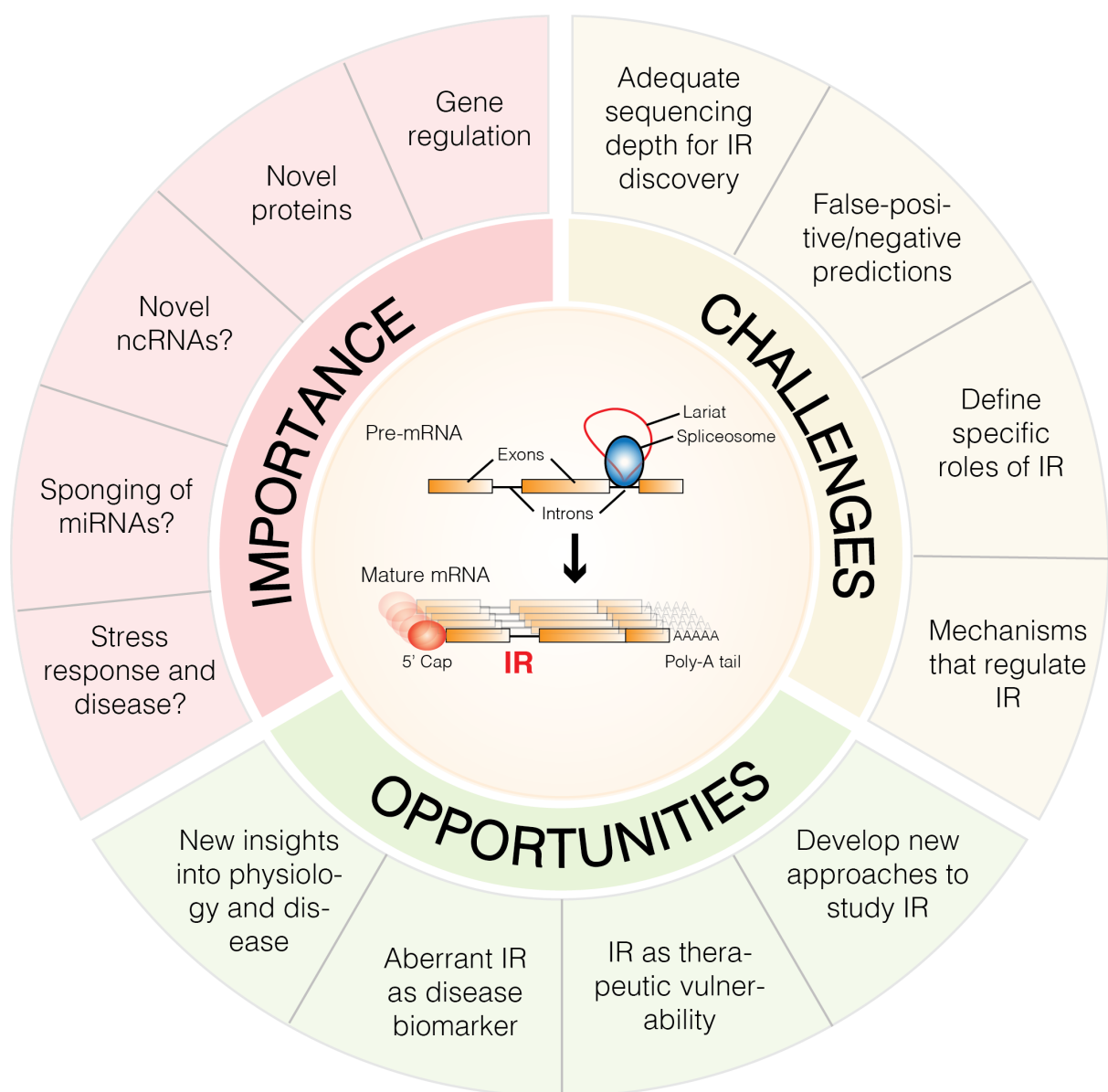


Figure 1: Illustration of intron retention and associated importance, key challenges, and opportunities. Concepts with a question mark (“?”) demand further validation. ncRNA – non-coding RNA.

Box 1: Regulators of intron retention

Retained introns are strongly associated with well-conserved intrinsic characteristics, which include a high GC content, short length, and weak splice sites and branch points [3]. *Trans*-regulatory elements modulating IR levels include RNA binding proteins for which binding motifs can often be found within or in the vicinity of retained introns. Loss of SR protein expression has a particularly pronounced effect on global IR. There are also a number of epigenetic factors that modulate IR levels [4]. For example, reduced DNA methylation around introns is associated with IR and so are changes in methyltransferase expression (e.g. DNMT1) or the occupancy of proteins with a methyl-CpG-binding domain (e.g. MeCP2). Post-translational histone modifications have been associated with IR as well, particularly trimethylation at lysine 36 in histone 3 (H3K36me3) has been reported in association with IR in mammals and plants [3]. In the latter, chromatin accessibility was also linked to IR. Most of the epigenetic modifications mentioned impact upon RNA polymerase II elongation rates. It is therefore no surprise, that IR has been associated with RNA Pol II accumulation and accelerated RNA Pol II elongation rates [4]. It was proposed recently that RNA modifications could also impact upon IR regulation [15].

Funding:

The research in the laboratory of J.J.-L.W. is supported by the National Health and Medical Research Council of Australia (Grants #APP1128175, 1158998, 2010647), the Cancer Council of NSW (Grant #RG19-05) and the Office of Health and Medical Research NSW (Grant #46014). U.S. received support from the National Health and Medical Research Council (Investigator Grant 1196405) and the Cancer Council NSW (project grant RG20-12).

Resources:

- i) https://genome.ucsc.edu/ENCODE/protocols/dataStandards/ENCODE_RNAseq_Standards_V1.0.pdf

References

1. Naro C, Jolly A, Di Persio S, Bielli P, Setterblad N, Alberdi AJ, Vicini E, Geremia R, De la Grange P, Sette C (2017) An orchestrated intron retention program in meiosis controls timely usage of transcripts during germ cell differentiation. *Dev Cell* 41 (1):82-93 e84. doi:10.1016/j.devcel.2017.03.003
2. Green ID, Pinello N, Song R, Lee Q, Halstead JM, Kwok CT, Wong ACH, Nair SS, Clark SJ, Roediger B, Schmitz U, Larance M, Hayashi R, Rasko JEJ, Wong JJ (2020) Macrophage development and activation involve coordinated intron retention in key inflammatory regulators. *Nucleic Acids Res* 48 (12):6513-6529. doi:10.1093/nar/gkaa435
3. Braunschweig U, Barbosa-Morais NL, Pan Q, Nachman EN, Alipanahi B, Gonatopoulos-Pournatzis T, Frey B, Irimia M, Blencowe BJ (2014) Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res* 24 (11):1774-1786. doi:10.1101/gr.177790.114
4. Monteuuis G, Wong JJJ, Bailey CG, Schmitz U, Rasko JEJ (2019) The changing paradigm of intron retention: regulation, ramifications and recipes. *Nucleic Acids Res* 47 (22):11497-11513. doi:10.1093/nar/gkz1068

5. Yap K, Lim ZQ, Khandelia P, Friedman B, Makeyev EV (2012) Coordinated regulation of neuronal mRNA steady-state levels through developmentally controlled intron retention. *Genes Dev* 26 (11):1209-1223. doi:10.1101/gad.188037.112
6. Braun CJ, Stanciu M, Boutz PL, Patterson JC, Calligaris D, Higuchi F, Neupane R, Fenoglio S, Cahill DP, Wakimoto H, Agar NYR, Yaffe MB, Sharp PA, Hemann MT, Lees JA (2017) Coordinated splicing of regulatory detained introns within oncogenic transcripts creates an exploitable vulnerability in malignant glioma. *Cancer Cell* 32 (4):411-426 e411. doi:10.1016/j.ccell.2017.08.018
7. Schmitz U, Pinello N, Jia F, Alasmari S, Ritchie W, Keightley MC, Shini S, Lieschke GJ, Wong JJ, Rasko JEJ (2017) Intron retention enhances gene regulatory complexity in vertebrates. *Genome Biol* 18 (1):216. doi:10.1186/s13059-017-1339-3
8. Inoue D, Polaski JT, Taylor J, Castel P, Chen S, Kobayashi S, Hogg SJ, Hayashi Y, Pineda JMB, El Marabti E, Erickson C, Knorr K, Fukumoto M, Yamazaki H, Tanaka A, Fukui C, Lu SX, Durham BH, Liu B, Wang E, Mehta S, Zakheim D, Garippa R, Penson A, Chew GL, McCormick F, Bradley RK, Abdel-Wahab O (2021) Minor intron retention drives clonal hematopoietic disorders and diverse cancer predisposition. *Nat Genet* 53 (5):707-718. doi:10.1038/s41588-021-00828-9
9. Jabre I, Reddy ASN, Kalyna M, Chaudhary S, Khokhar W, Byrne LJ, Wilson CM, Syed NH (2019) Does co-transcriptional regulation of alternative splicing mediate plant stress responses? *Nucleic Acids Res* 47 (6):2716-2726. doi:10.1093/nar/gkz121
10. Georges A, Holleley CE (2018) How does temperature determine sex? *Science* 360 (6389):601-602. doi:10.1126/science.aat5993
11. Dvinge H, Bradley RK (2015) Widespread intron retention diversifies most cancer transcriptomes. *Genome Med* 7 (1):45. doi:10.1186/s13073-015-0168-9
12. Smart AC, Margolis CA, Pimentel H, He MX, Miao D, Adeegbe D, Fugmann T, Wong KK, Van Allen EM (2018) Intron retention is a source of neoepitopes in cancer. *Nat Biotechnol* 36 (11):1056-1058. doi:10.1038/nbt.4239
13. Lorenzi C, Barriere S, Arnold K, Luco RF, Oldfield AJ, Ritchie W (2021) IRFinder-S: a comprehensive suite to discover and explore intron retention. *Genome Biology* 22 (1):307. doi:10.1186/s13059-021-02515-8
14. Kortel N, Ruckle C, Zhou Y, Busch A, Hoch-Kraft P, Sutandy FXR, Haase J, Pradhan M, Musheev M, Ostareck D, Ostareck-Lederer A, Dieterich C, Huttelmaier S, Niehrs C, Rausch O, Dominissini D, Konig J, Zarnack K (2021) Deep and accurate detection of m6A RNA modifications using miCLIP2 and m6Aboost machine learning. *Nucleic Acids Res* 49 (16):e92. doi:10.1093/nar/gkab485
15. Pendleton KE, Chen B, Liu K, Hunter OV, Xie Y, Tu BP, Conrad NK (2017) The U6 snRNA m(6)A methyltransferase METTL16 regulates SAM synthetase intron retention. *Cell* 169 (5):824-835 e814. doi:10.1016/j.cell.2017.05.003