

Univerzita Karlova
Přírodovědecká fakulta

Speciální chemicko-biologické obory
Molekulární biologie a biochemie organismů



Božena Klodová

Význam dlouhých nekódujících RNA v rostlinách
The role of long non-coding RNAs in plants

Bakalářská práce

Vedoucí práce: doc. RNDr. David Honys, Ph.D.

Konzultant: RNDr. Jan Fíla, Ph.D.

Praha, 2017

Děkuji svému školiteli doc. RNDr. Davidu Honysovi, Ph.D. za odborné konzultace a vedení. Dále mé poděkování patří RNDr. Honzovi Fílovi, Ph.D. za opětovanou pomoc, nekonečnou trpělivost a drahocenné rady, které pomohly dovést mojí práci k závěru. V neposlední řadě bych chtěla poděkovat mé babičce s dědou za jejich podporu v udržení nervů.

Prohlašuji, že jsem tuto bakalářskou práci vypracovala sama z uvedené literatury a na základě konzultací se svým školitelem a konzultantem.

Praha 2017.

Božena Klodová

Abstrakt a klíčová slova

Dlouhé nekódující RNA (lncRNA) jsou transkripty o délce větší než 200 nukleotidů s nízkým kódujícím potenciálem. Jedná se o skupinu regulačních ribonukleových kyselin (RNA) zatím v raném stádiu poznání, nicméně s potenciálně širokou rolí napříč organismy a biologickými procesy. U živočichů bylo popsáno několik regulačně významných lncRNA a jejich role roste mimo jiné i v souvislosti s mnohými nádorovými onemocněními. Přestože mnohé dlouhé nekódující transkripty byly odhaleny, jejich konkrétní mechanismy působení zůstávají neznámé. U rostlin se výzkum soustředí převážně na ekonomicky a zemědělsky významné organismy jako je rýže, kukuřice nebo sója. Jedná se ale zatím jen o omezené množství experimentálně ověřených transkriptů. Nicméně se zdokonalováním sekvenačních a bioinformatických metod začíná být významnost role lncRNA v mnohých biologických procesech (např. rozmnožování či rostlinné rezistenci k patogenům) naprosto zjevná. Tato bakalářská práce shrnuje dosavadní znalosti o dlouhých nekódujících ribonukleových kyselinách (lncRNA) s důrazem na jejich roli v rostlinách a taktéž se detailněji zabývá současnými problémy a nastiňuje budoucí potenciál výzkumu lncRNA.

Klíčová slova: nekódující RNA, lncRNA, krytosemenné rostliny, vývoj květů, regulace genové exprese, epigenetika

Abstract and key words

Long non-coding RNAs (lncRNA) represent group of transcripts with length greater than 200 nucleotides with low coding potential. It is a group of regulatory ribonucleic acid (RNA) still not fully understood but with significant potential in many biological processes across different species. For animals, many important lncRNA regulators and their roles in range of events including their involvement in carcinogenic diseases have been reported. Yet concrete mechanisms of functions are often yet to be discovered. Considering plants, economically important species such as rice, maize or soybean are of particular interest. There is still only a few fully annotated transcripts. However, with the constant improvement of sequencing and bioinformatic methods, the importance of lncRNA (for example in pathogen resistance or plant reproduction becomes clear. This bachelor thesis reviews up to date knowledge about lncRNA and its roles in plants. It also describes the difficulties of lncRNA research and discusses their future potential.

Key words: lncRNA, non-coding RNA, angiosperms, floral development, regulation of gene expresion, epigenetics

Seznam použitých zkratk

RNA	ribonukleová kyselina
ncRNA	nekódující RNA (angl. <i>non-coding RNA</i>)
lncRNA	dlouhá nekódující RNA (angl. <i>long non-coding RNA</i>)
nt	nukleotidy
ORF	otevřený čtecí rámec (angl. <i>open reading frame</i>)
miRNA	mikroRNA
siRNA	malé interferující RNA (angl. <i>small interfering RNA</i>)
mRNA	mediátorová RNA
PolIV	DNA-dependentní RNA polymeráza IV
PolV	DNA-dependentní RNA polymeráza V
ilncRNA	intronové dlouhé nekódující RNA (angl. <i>intronic long non-coding RNA</i>)
lincRNA	mezigenové dlouhé nekódující RNA (angl. <i>long intergenic non-coding RNA</i>)
RepA	sekvence A repetice (angl. <i>repeat A sequence</i>)
PRC2	polycomb repressive complex 2
H3K27	histon 3, lysin 27
FLC	flowering locus C
DNMT1	DNA methyl transferáza 1
RNAPII	DNA-dependentní RNA polymeráza II
NAT	transkript z opačného vlákna (angl. <i>natural antisense transcript</i>)
ARE	AU-bohaté elementy (angl. <i>AU-rich elements</i>)
3'UTR	3' nepřekládaná oblast (angl. <i>3' untranslated region</i>)
iNOS	inducibilní syntáza oxidu dusnatého (angl. <i>inducible nitric oxide synthase</i>)
AS	protismyslný (angl. <i>antisense</i>)
IPS1	indukovaný nedostatkem fosfátu (angl. <i>induced by phosphate starvation 1</i>)
ceRNA	kompetitivní endogenní RNA (angl. <i>competitive endogenous RNA</i>)
MRE	elementy odpovídající na mikroRNA (angl. <i>microRNA response elements</i>)
dsRNA	dvouvláknová RNA (angl. <i>double strand RNA</i>)
LDMAR	RNA asociovaná se samčí plodností za dlouhého dne (angl. <i>long-day -specific male-fertility-associated RNA</i>)
PSMS	samčí sterilita citlivá na fotoperiodu (angl. <i>photoperiod-sensitive male sterility</i>)
BcMF11	<i>Brassica campestris</i> Male Fertility 11
WT	divoká rostlina (angl. <i>wild type</i>)

TE	transponovatelné elementy
RBP	Proteiny vázající ribozomy (angl. <i>ribosome-binding proteins</i>)
NSR	nuclear speckle RNA-binding proteins
CDS	kódující sekvence DNA
AP2	APETALA 2
ELENA1	dlouhá nekódující RNA indukovaná Elf18

OBSAH

1. ÚVOD	1
2. CO JSOU LNCRNA	3
3. KLASIFIKACE LNCRNA	5
3.1 Rozdělení na základě genomového kontextu	6
3.2 Rozdělení na základě molekulárního mechanismu	6
3.2.1 <i>Epigenetická regulace a represe transkripce</i>	7
3.2.2 <i>Aktivace transkripce</i>	8
3.2.3 <i>Post-transkripční regulace</i>	9
3.2.4 <i>Maskování cílů miRNA</i>	9
3.2.5 <i>miRNA prekurzory</i>	11
4. IDENTIFIKACE LNCRNA	12
4.1 Celogenomová analýza	12
5. FUNKCE LNCRNA U ROSTLIN	15
5.1 Reprodukce	15
5.2 Nedostatek fosforečnanů, vývoj kořenů a prýtu	17
5.3 Abiotický stres	18
5.4 Biotický stres.....	18
5.5 Další funkce	19
6. ZÁVĚR	21
7. SEZNAM POUŽITÝCH CITACÍ.....	1

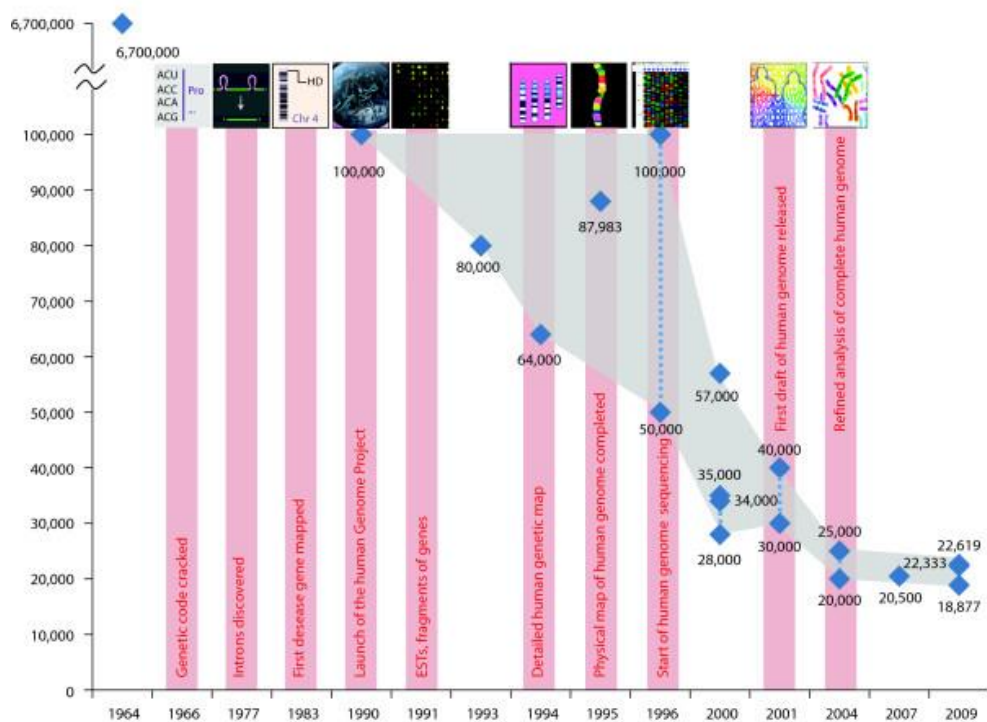
1. Úvod

Ribonukleová kyselina (RNA) je molekula, o které s jistotou můžeme tvrdit hlavně to, že o ní zatím nevíme zdaleka vše. Byla objevena na začátku minulého století P. A. Levenem a následné experimenty vedly až k definování centrálního dogmatu molekulární biologie popsanému v roce 1957 Francisem Crickem, stavící RNA do pozice adaptoru mezi DNA a proteiny (Crick 1958). Dnes se, mimo funkce mRNA jako mediátoru v centrálním dogmatu, vědecké bádání potýká s otázkou komplexity genomu a zejména regulačních funkcí dalších typů RNA nekódujících proteiny.

V roce 1971 postuloval C. A. Thomas takzvaný C-value paradox (Thomas 1971), který nemůže být vysvětlen pouze centrálním dogmatem. Výsledky experimentů totiž naznačují, že komplexita organismu sice koreluje s množstvím DNA u bezobratlých, u obratlovců ovšem takový trend nabýval spíše opačné tendence (Mirsky and Ris April 1951). Dalšímu paradoxu dala vzniknout práce Friedricha Vogela, který se jako první pokusil v roce 1964 odhadnout počet genů v lidském genomu na základě počtu aminokyselin v α a β řetězci hemoglobinu. On sám byl zaskočen počtem přibližně 6,7 milionu genů (Vogel 1964). V průběhu posledních padesáti let se na základě nových objevů a později hlavně díky nástupu sekvenačních metod počet lidských genů zpřesňoval, tedy klesal (obr 1.). Chybou Vogelova experimentu byla právě presumpce kódujícího potenciálu celé DNA. Dnes se předpokládá, že přes 60 procent genomu je složeno z nekódujících sekvencí, v šedesátých letech s populárním označením „*junk DNA*“, do češtiny volně přeloženo jako „zbytečná DNA“. Teorií funkce *junk DNA* bylo mnoho. Někteří zastávali názor, že se jedná jen o zbylý evoluční odpad s případnou strukturní funkcí například pro tvorbu oblasti centromer (Ohno 1972). V sedmdesátých letech bylo objeveno, že je transkribována výrazně větší část genomu (cca 65%), než je nutné pro tvorbu bílkovin. Začaly se objevovat názory, že regulační funkce nekódujících RNA (ncRNA) by vedle alternativního sestřihu mohla být právě nositelem výhody vyšších eukaryot (Mattick 2001). Jednalo by se tak o další krok k vysvětlení Thomasova paradoxu. Nekódující sekvence totiž zabírají většinovou část transkriptomu (Okazaki et al. 2002). V posledních letech proto došlo k podstatnému nárůstu zájmu o nekódující složky genomu a odchýlení od názoru zastávajícího existenci jen několika ojedinělých adaptorových molekul jako RNA XIST nebo ribozomální RNA (Bussotti

et al. 2013). Se zlepšováním metod dochází i k postupnému experimentálnímu zjišťování konkrétních funkcí ncRNA.

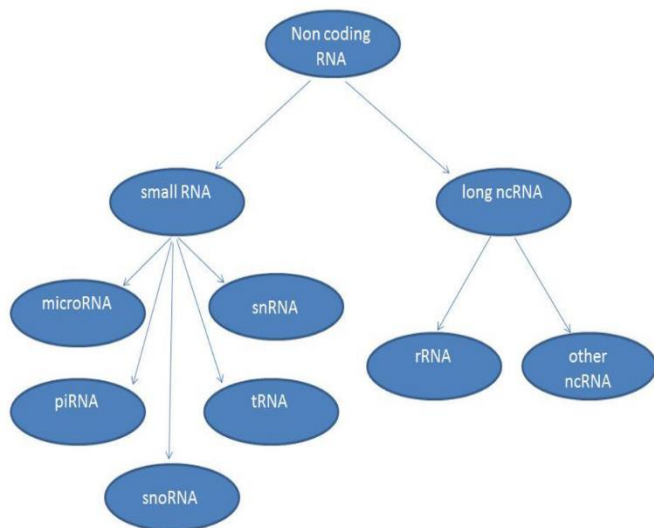
Jednou z recentně vyčleněných skupin jsou dlouhé nekódující RNA (lncRNA), definované jako transkripty s nízkým kódujícím potenciálem a délkou větší než 200 nukleotidů (nt). Navzdory nedostatku experimentálně ověřených dat je pravděpodobné, že zastávají mnohé krucální funkce u celé řady organismů. U člověka jsou právě mutace ve funkcích nebo regulaci lncRNA považovány za jedny z potenciálně hlavních příčin onemocnění, například nádorových (v přehledovém článku: Wapinski and Chang 2011). U rostlin je celá problematika méně probádaná, zjistilo se ale, že lncRNA se mohou podílet na mnoha procesech od stresových odpovědí, přes reprodukci až po syntézu důležitých látek. V této práci bych chtěla výsledky dosavadního výzkumu lncRNA u rostlin shrnout a nastínit jeho problematiku a potenciál.



Obr. 1 vývoj výzkumu odhadu počtu genů u člověka začínající Vogelovým výpočtem v roce 1964. Upraveno podle: Pertea and Salzberg 2010

2. Co jsou lncRNA

LncRNA představují skupinu RNA řadící se do širší kategorie nekódujících transkriptů, tzn. těch s nízkým kódujícím potenciálem obsahující pouze krátké, nebo žádné otevřené čtecí rámce – ORF (angl. *open reading frame*). Jednou z možností rozdělení nekódujících transkriptů je dělení na základě jejich délky, jak ostatně odráží i jejich název. Dlouhé nekódující RNA jsou v tomto rozdělení definovány jako transkripty delší než 200 nukleotidů (nt). Skutečná



Obr. 2 Způsob dělení nekódujících RNA na základě jejich délky, krátké RNA do 200 nukleotidů, dlouhé nad 200 nukleotidů. Skupina, o které pojednává tato práce je na obrázku zobrazena jako *other ncRNA*. Upraveno podle: Fenoglio et al. 2013

výpovědní hodnota této charakteristiky je ale fádni, jedná se totiž čistě o uměle vytvořenou definici platnou do té doby, než bude nahrazena novou variantou postavenou na experimentálních výsledcích (Mattick and Rinn 2015). Nebo je pravděpodobné, že pojem lncRNA přestane existovat za vzniku menších, lépe definovatelných skupin. Vyčlenění všech lncRNA jednotnou definicí je totiž, jak si troufám usuzovat, pro variabilitu vlastností (ať už se jedná o místo vzniku, mechanismus účinku,

nebo funkci) jednotlivých lncRNA téměř nemožné a s přibývajícím množstvím experimentálních dat časem nepotřebné. Momentální délkové vymezení vzniklo kvůli vyčlenění různých lépe definovaných skupin kratších transkriptů, jako jsou například miRNA a siRNA. Celkové rozdělení typů RNA je uvedeno na obrázku 2. Druhou možností dělení transkriptů bez kódujícího potenciálu je konzervativnější způsob na základě míry exprese. Rozdělují se pak na dříve probádané regulační RNA s více méně konstantní měrou exprese a regulační transkripty s měrou exprese proměnlivou. První skupina zastává v organismu udržovací funkce (patří sem např. rRNA nebo tRNA) a zahrnuje RNA esenciální pro život. Naopak proměnlivost exprese druhé skupiny často souvisí se specifickými podmínkami. LncRNA v tomto rozdělení spadají do kategorie regulační.

Jak jsem uvedla výše, popis obecných vlastností lncRNA je komplikovaný. Z dosavadních výzkumů lze ale vyvodit některé obecně platné fenomény, které uvádím v následujícím odstavci. Nakolik jsou tyto vlastnosti platné pro veškeré lncRNA, je diskutabilní. Jejich souhrn je ale užitečný k pochopení široké diverzity mechanismů a funkcí

lncRNA v kontextu srovnání s mRNA. Transkripce lncRNA je zprostředkována DNA-dependentní RNA polymerázou I, II a III, u rostlin navíc i specifickými DNA-dependentními RNA polymerázami IV a V (PolIV, PolV)¹. Bylo zjištěno, že transkripty vzniklé aktivitou polIV postrádají polyA konec, neobsahují introny a na 5'konci obsahují monofosfát (Li et al. 2015). Tím se odlišují od transkriptů polII, které jsou většinou polyadenylované s trifosfátovou čepičkou na 5'konci stejně jako mRNA. Posttranskripční úpravy se tudíž mezi jednotlivými lncRNA mohou lišit. LncRNA mají ve srovnání s mRNA nižší obecný stav exprese, jedná se ale o transkripty s výrazně vyšší tkáňovou specifikou a se větším rozdílem v expresi za specifických podmínek (Di et al. 2014; Hirsch et al. 2006; H. Wang et al. 2014). Tento fenomén naznačuje jejich potenciální funkci v regulaci odpovědi na vnější stimuly. Ze strukturní analýzy je prokázáno, že lncRNA jsou většinou kratší než mRNA, obsahují méně, většinou delších, exonů (Khemka et al. 2016; Yuan et al. 2016; Zou et al. 2016) a mívají málo homologů mezi příbuznými druhy (Pang et al. 2006). LncRNA jsou navíc pravděpodobně přísněji inhibovány methylací, což dokazuje preciznost jejich regulace (M. Wang et al. 2015). Dále obsahují hodně transponovaných elementů, které by potenciálně mohly být zodpovědné jak za jejich funkci, tak evoluci (Kapusta et al. 2013; D. Wang et al. 2017). LncRNA mohou být lokalizované v jádře nebo v cytoplasmě, a proto jsou schopné regulovat genovou expresi na více úrovních. Myslím, že souhrnně je možné lncRNA označit jako specificky exprimované, strukturně variabilní, přísně regulované transkripty se širokým polem účinku.

¹ PolIV je také klíčová pro tvorbu většiny siRNA (Zhang et al. 2007), polV figuruje pravděpodobně v umlčování genů nikoliv přímo transkripcí siRNA, ale právě mechanismem tvorby dlouhých nekódujících sekvencí fungujících jako lešení pro siRNA (Wierzbicki 2008).

3. Klasifikace lncRNA

Navzdory rychlému nárůstu objevů nových lncRNA u nejrůznějších organismů, je stále množství experimentálně ověřených dat nízké. Existuje mnoho databází shrnujících různé informace o lncRNA, v tabulce 1 (Fritah et al. 2014) jsou některé uvedeny. „The functional lncRNA Database“, která zahrnuje výhradně transkripty s experimentálně ověřenou funkcí, čítá pouze 209 záznamů, na rozdíl například od největší databáze NONCODE (v 3.0) s více než 30 000 záznamů zahrnujících i lncRNA pouze s předpokládanou funkcí. Z výše uvedených důvodů proto nelze vytvořit jednotící parametr pro lncRNA a možností klasifikace je více. Například na základě funkce, délky, asociace s DNA elementy se známou funkcí, podobnosti s kódujícími RNA, stability, struktury a sekvence, na podmínkách exprese nebo genomovém kontextu (St Laurent et al. 2015). V této práci jsem se rozhodla rozvést rozdělení na základě genomového kontextu a mechanismu působení. V poslední kapitole dále rozvádím dělení podle konkrétních funkcí. Rozdělení na základě genomového kontextu je elegantní pro svoji objektivitu a univerzalitu. Popis na základě podobnosti k již známým genům umožňuje snadno navrhnout jasně definovatelné skupiny. Mnohé práce se pak často zabývají již jen konkrétními skupinami lncRNA. Na druhou stranu zatím není prokázána významná spojitost mezi funkcí (nebo jejím mechanismem) a umístěním dané lncRNA v genomu. Důvodem může být schopnost lncRNA ovlivňovat i vzdálené lokusy. Proto jsem se rozhodla věnovat více prostoru rozdělení na základě mechanismu, které sice není tak jasně kategorizovatelné, umožňuje ale lepší pochopení vlastností lncRNA.

Databáze	DIANA- lncBase	LNCipedia	lncRNA Disease	lncRNadb	Noncode v3.0	The Functional lncRNA Database
Počet lncRNA	56110	32183	1564	>150	33829	204
Obsah databáze	miRNA vazebná místa	Struktura, kódující potenciál, miRNA vazebná místa	lncRNA asociované se 166 onemocněními	Detailní popis a informace	Detailní popis, potenciální funkce	Popsaná buněčná funkce, prekurzory miRNA, repetice

Tab. 1: Tabulka popisující některé databáze obsahující informace o lncRNA s popisem popsaných kritérií. Upraveno podle: Fritah et al. 2014

3.1 Rozdělení na základě genomového kontextu

LncRNA mohou být transkribovány z různých částí genomu. S využitím již popsaných genů jako markerů je možné lncRNA rozdělit na základě jejich umístění. LncRNA mohou pocházet buď z intronové oblasti kódujících genů (intronové lncRNA; ilncRNA), z mezigenových nekódujících oblastí (mezigenové lncRNA; lincRNA) nebo se může jednat o transkripty úplně či částečně překrývající kódující oblasti (překrývající lncRNA; angl. *overlapping lncRNA*). Jednotlivé lncRNA pak můžou vykazovat funkci jak v trans, tak cis, ovlivňovat buď geny stejného lokusu nebo i ve velkých vzdálenostech. Byly prokázány i transkripty pocházející z pseudogenů, ale jejich detailní analýza musí být ještě provedena.

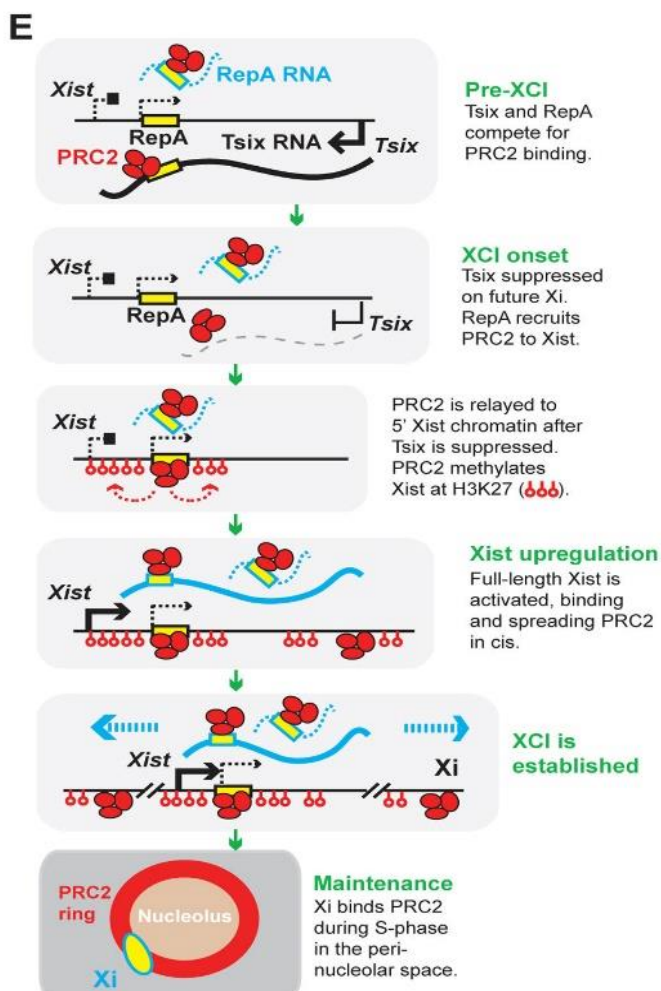
Nadřazenou předešlému dělení a funkčně významnou skupinu tvoří lncRNA spadající do skupiny NAT (přírozně se vyskytující protismyslný transkript; angl. *natural antisense transcript*). NAT je příslušný transkript přepisovaný z opačného vlákna, než patřičný kódující gen. Jedná se o heterogenní skupinu kódujících i nekódujících RNA schopných zpětně ovlivňovat expresi svého partnerského genu, případně nějakých dalších (Khorkova et al. 2014). NATs je potom možno rozdělit na (1) shodné regulátory, jejichž exprese roste s expresí genu na komplementárním vlákně a (2) opačné regulátory, jejichž exprese roste s klesající expresí partnerského genu či obráceně (H. Wang et al. 2014).

3.2 Rozdělení na základě molekulárního mechanismu

Molekulární mechanismus účinku byl prokázán pouze u malého množství lncRNA, povětšinou u živočichů. Na rozdíl od krátkých nekódujících sekvencí nelze funkce odvozovat ze struktury transkriptů nebo způsobu jejich vzniku. Pro svoji délku a variabilitu mohou lncRNA zastávat mnohé funkce. Mohou působit jako signály v rychlé odpovědi bez nutnosti translace. Dalším archetypem jsou lncRNA fungující jako návnady pro RNA-vazebné molekuly schopné akumulovat molekuly do komplexů, ovlivňovat transkripci vyvazováním transkripčních faktorů, nebo naopak adresovat důležité molekuly do cílových lokací jak v cis, tak trans komplementaritě (K. C. Wang and Chang 2011). V následující kapitole rozvádím konkrétní mechanismy účinku, nicméně hranice mezi jednotlivými skupinami jsou řádné a bude potřeba dalšího výzkumu pro jasnou kategorizaci.

3.2.1 Epigenetická regulace a represe transkripce

Regulace epigenetických změn struktury chromatinu jsou úzce spojeny s působením RNA (Bernstein and Allis 2005). LncRNA se může podílet vazbou a následným shlukováním faktorů vedoucích k umlčení aktivity cílových genů. Mechanismus účinku byl popsán jak u živočichů, tak u rostlin. Lidská lncRNA Xist je zodpovědná za inaktivaci chromosomu X u samičích pohlavních buněk. Xist pokrývá inaktivovaný chromosom a pomocí RepA (repeat A) je na 5' Xist zrekrutován PRC2 (Polycomb binding complex 2) vázaný svojí Ezh2 doménou k RNA (Chaumeil et al. 2006). Dochází k rozložení PRC2 po celé délce Xist a trimetylaci H3-K27 (histon 3, lysin 27) vedoucí k umlčení transkripce². Mechanismus včetně potenciální regulace pomocí krátkého transkriptu Tsix je uveden na obrázku č. 4. Role Xist je ale pravděpodobně



Obr. 4 Mechanismus působení Xist v rekrutování PRC2 na RepA a jeho následné šíření po celé délce Xist. Tsix je blokován v umlčovaném chromosomu, v případě aktivace tvoří PRC2-Tsix komplex a brání umlčení. Upraveno podle: Zhao et al. 2008

mnohem komplexnější zastávající úlohu například při genové translokaci do umlčených částí chromosomu X vedoucí k jeho metylaci (Chaumeil et al. 2006).

Podobný princip rekrutování PRC2 je nutný pro rostlinnou vernalizaci³ v případě epigenetického umlčování genu FLC (*FLOWERING LOCUS C*) bránícího kvetení, pomocí H3K27m3. Dvě chladem indukované lncRNA byly objeveny v této vernalizační dráze – COLDAIR a COOLAIR. Bylo potvrzeno, že působením chladu dochází k transkripci 1,1 kbp dlouhého transkriptu COLDAIR z kryptického promotoru ve stejném směru (*sense*) jako FLC. Metylace FLC je indukována histon methyl-

²Xist vyřazuje z funkce RNA polymerázu II a brání transkripčním faktorům v iniciaci transkripce. Jedná se o krok nezávislý na XistRepA sekvenci (Chaumeil et al 2006).

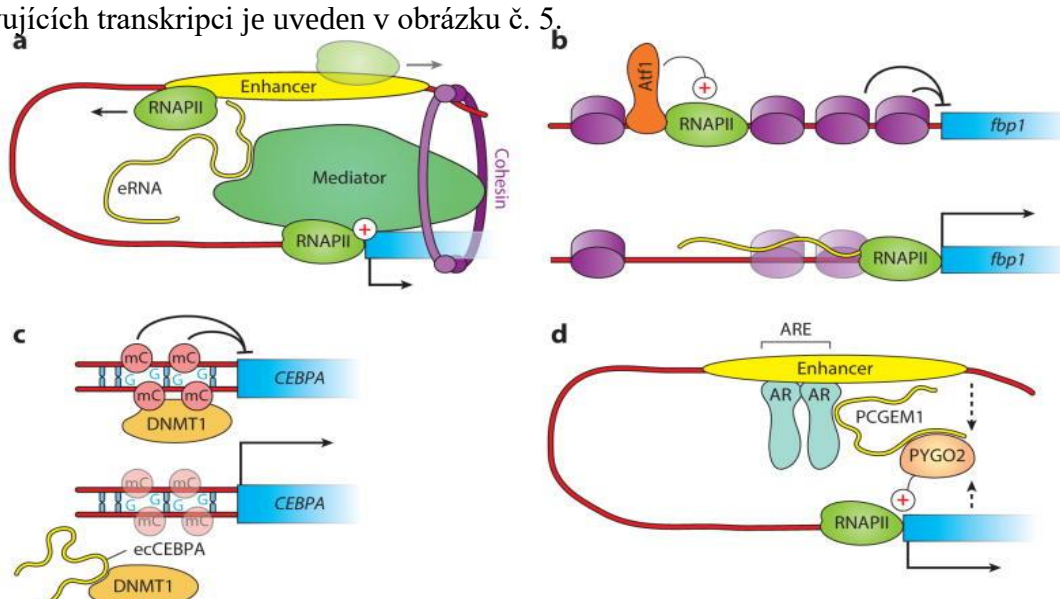
³Vernalizace je proces nutný pro vykvetení rostliny na jaře, při němž musí být rostlina po delší dobu vystavena chladu.

živočichů nazývanou E(z), u rostlin zastupována 3 homologními doménami (CLF, SWN, MEDEA). Pomocí RNA interference bylo potvrzeno, že COLDAIR je nutný pro represi FLC a indukci kvetení pravděpodobně právě mechanismem rekrutování CLF k FLC vedoucí k jeho umlčení metylací (Heo and Sung 2011).

3.2.2 Aktivace transkripce

Funkce lncRNA v aktivaci transkripce může být provedena několika mechanismy často využívajících proteinovou afinitu k RNA či opět chromatinovou remodelaci. Příkladem je inaktivace metylace promotoru daného genu pomocí RNA interagujících s DNMT1 (DNA metyl transferáza 1) (Di Ruscio et al. 2013). Výzkum byl proveden na modelovém genu CEBPA, jehož exprese je metylačně sensitivní a inaktivována pomocí DNMT1. *Sense upstream* lncRNA nazvaná ecCEBPA váže DNMT1 na specifickou sekundární sekvenci, což umožňuje zahájení transkripce CEBPA genu. DNMT1 totiž prokazuje vyšší vazebnou afinitu k RNA než k DNA. Povedlo se prokázat, že podobný mechanismus funguje i u dalších lokusů, může se tedy jednat o základní princip regulace transkripce ovlivněním metylace (Di Ruscio et al. 2013).

Ve výzkumu transkripce genu *fbp1*⁺ indukovaném při glukózovém hladovění u *Schizosaccharomyces pombe* bylo objeveno, že *sense upstream* nekódující transkripty jsou potřebné pro chromatinovou remodelaci umožňující RNA polymeráze II (RNAPII) transkripci genu *fbp1*⁺ za přítomnosti transkripčních faktorů (Atf1 a Rst2) nutných mj. pro odstranění represorů (tup11 a tup12) (Hirota et al. 2008). Mechanismus těchto a dalších mechanismů aktivujících transkripci je uveden v obrázku č. 5.

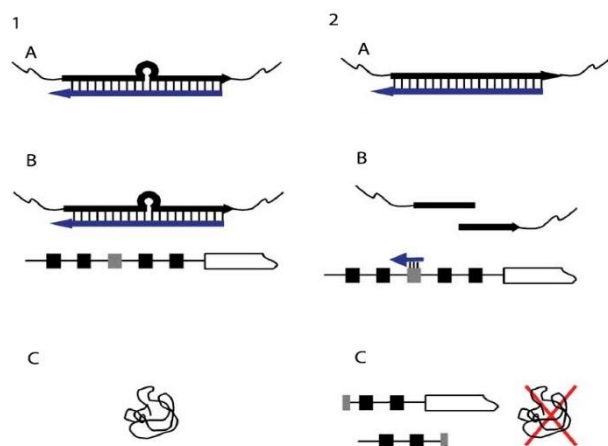


Obr. 5 Mechanismy působení lncRNA v aktivaci transkripce. a) lncRNA (eRNA) slouží jako přímé propojení mediátoru a zesilovače (angl. enhanceru) pomocí tvorby smyčky za přítomnosti kohesinu. b) lncRNA je nutná k aktivaci RNAPII a následné remodelaci chromatinu pro přepis *fbp1* genu u *Schizosaccharomyces pombe*. c) lncRNA umožňuje hypometylací promotoru vyvázáním RNA afinitního represoru DNMT1. d) lncRNA interaguje s proteiny stimulujícími transkripci a váže se do komplexu. Upraveno podle: Bonasio and Shiekhattar 2014

3.2.3 Post-transkripční regulace

Způsobů, kterými lncRNA může post-transkripčně ovlivňovat mRNA je více, obecně lncRNA většinou opět umožňuje vazbu mezi cílovým transkriptem a regulačními elementy, nebo jí naopak brání. Nejstarším známým mechanismem je překryv sestřihových míst NATs vedoucí ke změně v alternativním sestřihu, nebo přímá interakce se sestřihovými faktory⁴ (Kung et al. 2013; Morrissy et al. 2011; Munroe 1988). Další možností regulace je stabilizace mRNA zabráněním vazby destabilizujících faktorů na AU bohaté elementy (angl. *AU-rich elements* – AREs) v 3'UTR transkriptu. Jedním z popsaných mechanismů je stabilizace iNOS (inducibilní syntáza oxidu dusnatého, angl. *inducible nitric oxide synthase*) pomocí jejího NAT. V případě obsazení 3'UTR *sense* oligonukleotidy, které soutěží o vazebná místa pro *antisense* (AS) transkript docházelo k rychlejší degradaci iNOS. Naopak v případě *over-exprese* AS transkriptu docházelo k větší stabilitě iNOS měřené pomocí reportérové luciferázy na 3'UTR mRNA (Matsui et al. 2008). Další možností ovlivnění je rekrutování stabilizujících ARE vazebných faktorů pomocí AS transkriptu na 3'UTR mRNA (Matsui et al. 2008).

3.2.4 Maskování cílů miRNA



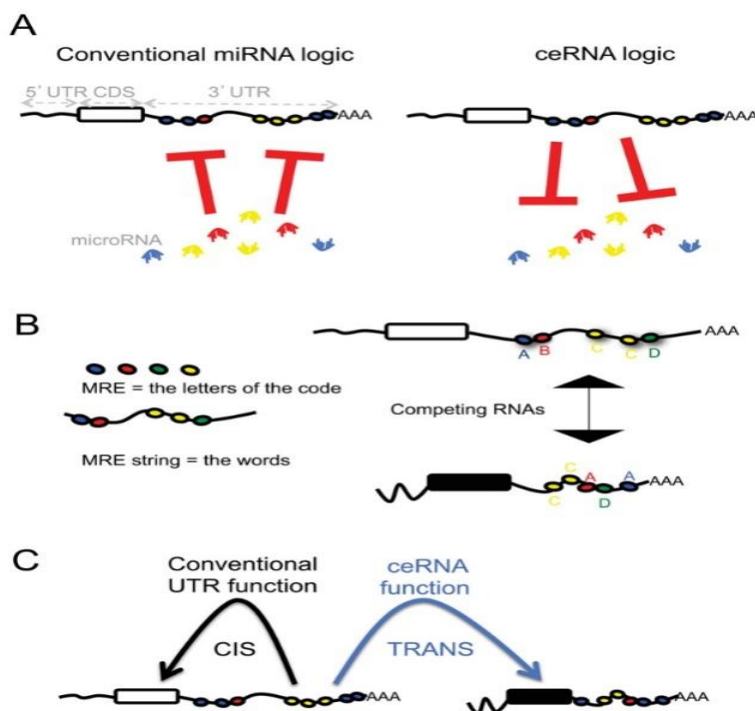
Obr. 6 Mechanismus degradace PHO2 pomocí miRNA-399. V dráze č. 1 je IPS1 (černě) nedegradovatelná díky nepárové smyčce, na ní navázaná miRNA-399 proto nemůže štěpit PHO2 mRNA (čtverečkovaná) a dochází k translaci. V dráze č. 2 IPS1 neobsahuje smyčku a je díky navázání miRNA-399 rozštěpena. Následně je rozštěpena i PHO2 mRNA a nedochází k translaci. Upraveno podle: Franco-Zorrilla et al. 2007

LncRNA může obsahovat sekvenci homologní k mnohým kratším regulačním transkriptům, například miRNA a zároveň dokáže být imunní k jejich štěpné aktivitě. Funguje potom jako falešný cíl, který vyvazuje danou miRNA a brání jí v regulaci jejího původního cíle. Příkladem lncRNA, fungující tímto mechanismem, je IPS1 (indukovaný nedostatkem fosfátu, angl. *INDUCED BY*

⁴Jedná se například o savčí lncRNA *MALAT1* a *Gomafu/MIAT* exprimované v jádře, přičemž obě potenciálně související s rakovinnými onemocněními.

PHOSPHATE STARVATION 1). IPS1 obsahuje motiv komplementární k miRNA-399 a v případě nedostatku fosfátu váže miRNA-399. V případě vysoké exprese miRNA-399 totiž dochází k snížení akumulace PHO2, mRNA zodpovědné za pokles koncentrace fosfátu v kořenových vláscích a jeho případnou translokaci v případě jeho nedostatku. U experimentů se zvýšenou expresí IPS1 docházelo k zábraně v degradaci PHO2 vedoucí ke stejnému fenotypu, a proto se předpokládá, že IPS1 funguje jako pseudosubstrát miRNA-399 (Franco-Zorrilla et al. 2007). Mechanismus funkce je uveden v obrázku č. 6 s popisem potenciální obrany lncRNA proti štěpení komplementární miRNA pomocí nepárové smyčky v nekonzervované struktuře lncRNA.

Na tomto principu je založena teorie kompetitivních endogenních RNA (ceRNA). Podle této teorie je dvojice miRNA a k nim komplementární MRE (MicroRNA response elements – elementy zodpovědné za komunikaci mezi transkripty) sekvence daných kódujících i nekódujících RNA základem v komunikaci a vzájemné regulaci mezi jednotlivými transkripty (Salmena et al. 2011). Na základě této teorie je pak podobnost MRE mezi jednotlivými transkripty pojítkem mezi jejich funkcí a regulací, a tudíž by z toho bylo možné usuzovat i vzájemné ovlivnění kódujících a nekódujících sekvencí (obr. 7).



Obr. 7 Principy fungování ceRNA jako jazyka transkriptů. A) působení mezi mRNA a miRNA je reciproké, tudíž za využití miRNA jako prostředníka může mRNA ovlivňovat jinou mRNA. B) MRE fungují jako písmena kódu skládajícího se v sekvenci „slov“. Z míry jejich shody mezi jednotlivými transkripty lze vyvozovat podobnost jejich regulace. C) MRE v UTR oblasti mohou regulovat RNA v cis a pomocí miRNA by měly být schopny regulovat i v trans. Upraveno podle: Salmena et al. 2011

3.2.5 miRNA prekurzory

Krátké transkripty s délkou 21–24 nt označované jako miRNA fungují nejen u rostlin jako důležití transkripční i post-transkripční regulátory vznikající sestřihem dvouvláknové RNA (dsRNA angl. *double-stranded RNA*). Jako matrice pro vznik krátkých RNA mohou sloužit jak kódující, tak dlouhé nekódující sekvence asociací se svým NAT (natural-antisense transcript) pomocí alternativního sestřihu. Tento fenomén byl popsán u rostlin i živočichů a s přibývajícím množstvím genomových analýz nabývá prekurzorová funkce lncRNA na významu (Florez-Zapata et al. 2016; Hirsch et al. 2006; JX Tian et al. 2016b).

4. Identifikace lncRNA

Jedním ze základních kamenů výzkumu nejen lncRNA, ale i obecně nekódujících transkriptů, je problematika jejich odlišení od mRNA a jejich identifikace v genomu. Na rozdíl od mRNA není primární struktura nekódujících transkriptů dostatečně informativní a statisticky významná. Neexistuje zde totiž rozsáhlý kód jednotlivých aminokyselin, pro popis si musíme vystačit pouze se čtveřicí bází a to komplikuje i určení homologie mezi jednotlivými transkripty (Bussotti et al. 2013). Za zmínku stojí i komplikace při určování sekundární struktury RNA. Kvůli chemickým vlastnostem je náročné určování struktury RNA metodou krystalizace náročné, výzkum se proto většinou uchyluje jen k bioinformatické předpovědi struktury. Ačkoliv zatím není dostatek dat k tomu, aby se na základě znalosti sekundární struktury lncRNA předpovídala jejich funkce, jedná se o další aspekt složitosti výzkumu nekódujících RNA v porovnání s těmi kódujícími. Vlastnosti nekódujících sekvencí svoji identifikaci dále komplikují například nízkou expresí a malou měrou vzájemné homologie, vysokou tkáňovou a podmínkovou specifitou nebo přílišnou podobností s mRNA danou možným společným původem.

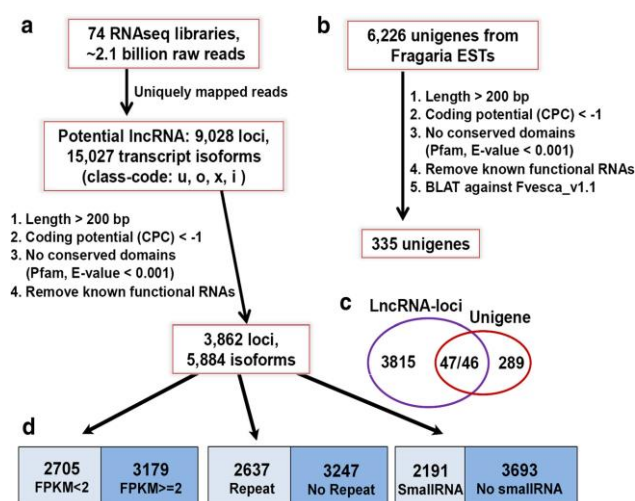
K přesnosti identifikace lncRNA přispěl rozvoj sekvenovacích metod umožňujících popis celého transkriptomu. Jednou z metod, která je hojně využívána v celogenomových analýzách lncRNA (Kang and Liu 2015; Kwenda et al. 2016; Y. Zhang et al. 2014) je sekvenování nové generace (NGS, angl. *next generation sequencing*), kterou je možné detekovat krátké exony, nepřekládané oblasti (UTR), nebo nekonzervované oblasti. Na rozdíl od DNA mikročipů je například také schopná rozlišit i drobné rozdíly v sekvencích a přesné ohraničení 5' a 3' hranic exonu (Nagalakshmi et al. 2008). Jedná se také o metodu schopnou zachytit transkripty s nízkou expresí (Bussotti et al. 2013).

S již zmiňovaným nedostatkem experimentálně ověřených lncRNA se stále objevují případy chybné interpretace a špatného zařazení některých transkriptů, takže je vždy nutné k výsledkům přistupovat s jistým nadhledem. Chybné zařazení se může objevit i v jednotlivých databázích (v přehledovém článku: Bussotti et al. 2013; Dinger et al. 2008), ze kterých potom vychází další výzkumy.

4.1 Celogenomová analýza

Konkrétní postupy jednotlivých výzkumných skupin se liší, princip metody detekce nových lncRNA zůstává ale podobný. Obecně vychází analýza z původních dat, většinou velmi

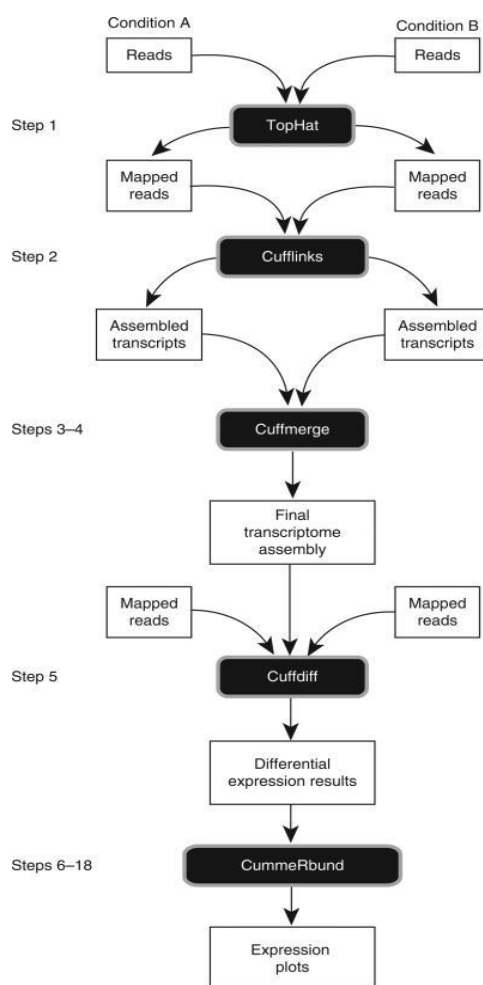
rozsáhlých, převzatých z databází. V prvním kroku jsou data analyzována pomocí dostupných softwarů na analýzu hrubých RNA-seq dat. Dále jsou transkripty vyřazovány na základě kritérií pro konkrétní výzkum. Výsledkem je získání potenciálních lncRNA exprimovaných za daných podmínek v určitých tkáních. Pro pozdější výzkum tak vzniká databáze kandidátů ovlivňujících danou dráhu.



Obr. 8 Postup v identifikaci lncRNA u genomové analýzy jahodníku obecného (*Fragaria vesca*). Upraveno podle: Kang and Liu 2015

Následovalo užití různých softwarových programů, které seřadily a rozdělily hrubá data z RNA-seq. V tomto článku byl použit hojně užívaný protokol Tuxedo, skládající se z TopHat a Cufflinks analýzy. TopHat má na starosti alignment jednotlivých sekvencí podle osekvenovaného genomu daného zástupce (též mapování)⁶. Tophat je dále schopný rozeznat setřihová místa, takže správně přirovná sestřihnuté sekvence z RNA-seq k nesestřihnuté genomové sekvenci. Cufflinks následně spojí segmenty získané z RNA-seq do známých i nových transkriptů a změří míru jejich exprese. Pokud ale existuje příliš velká variace alternativních sestřihů jednoho genu, snižuje se přesnost tohoto kroku. Pomocí Cuffmerge (spolu s Cuffcompare jsou součástí programu Cufflinks) jsou

Jako příklad uvádím postup u vybraného článku zabývajících se globální analýzou a identifikací lncRNA u jahodníku obecného (*Fragaria vesca*) v průběhu kvetení a vývoje plodů (Kang and Liu 2015), zobrazano na obrázku 8. Na začátku bylo získáno 74 souborů dat z RNA-seq pocházejících z pletiv 37 různých jedinců.



Obr. 9 Postup v provádění bioinformatické analýzy za použití Tuxedo protokolu. Upraveno podle: Trapnell et al. 2012

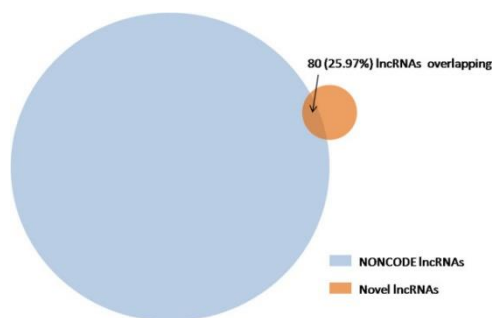
⁶ Existují i softwary schopné tzn. *de novo* alignmentu, bez potřeby templátu, např. software Trinity

transkripty jednotlivých zástupců spojeny v jeden a s použitím Cuffcompare porovnány s anotacemi a charakterizovány. V tomto kroku je možné selektovat transkripty podle potřebných podmínek, například program Cuffdiff měří a statisticky vyhodnocuje rozdíly exprese v určených podmínkách, nicméně v případě výše uvedeného článku nebyla tato diagnóza provedena. Cumberbund je pak vizualizační krok protoku, který utvoří potřebné grafy (Trapnell et al. 2012). Celý proces protokolu Tuxedo je ještě znázorněn v obrázku 9.

Využitím protokolu Tuxedo bylo získáno 15 027 kandidátů na ncRNA u jahodníku obecného. Další selekce byla provedena tak, aby byly splněny podmínky lncRNA, takže byly z analýzy vyřazeny transkripty kratší než 200 nt, transkripty s kódujícím potenciálem větším než -1 (za použití *Coding potential calculator*) a transkripty obsahující konzervované proteinové domény (na základě databáze Pfam), případně všechny RNA se známou funkcí. Na základě této selekce zbylo 5 884 potenciálních lncRNA. Další selekční krok je ukázkou specifického přístupu u tohoto konkrétního článku, kdy autoři dále vyřadili prekuzory známých miRNA, transkripty obsahující repetice a transkripty s příliš nízkou expresí, v článku označované jako hc-lncRNA (high-confident lncRNA).

Na základě takovéto celogenomové analýzy je možné zkoumat vlastnosti lncRNA jako je například rozmístění v genomu, korelace exprese se sousedními nebo vzdálenými geny, délka exonů a intronů, druhová specifita či konzervovanost.

Jednotlivé kroky analýzy mají stále své hranice a společně s experimentální anotací lncRNA je k postupu v jejich výzkumu nutné i zdokonalování metod za účelem jejich identifikace. Důležitým krokem je i aktualizace existujících databází na základě nejnovějších metod a sjednocení používaných metod, aby nedocházelo k mylným anotacím. Na to poukazuje i Sun a kol. (2012), kde byl popsán nízký překryv mezi nově identifikovanými lncRNA a těmi vloženými v databázi (obr. 10).



Obr. 10 Srovnání množství lncRNA vložených v databázi NONCODE 3.0 (36991 zápisů k datu) s lncRNA objevenými bioinformatickou analýzou (308 transkriptů) při identifikaci nových lncRNA mutovaných v genu *klf1* u myších linií. Upraveno podle: Sun et al. 2012

5. Funkce lncRNA u rostlin

Jen omezené množství rostlinných lncRNA má experimentálně prokázanou funkci. S postupem času je ale prováděno čím dál více bioinformatických analýz vedoucích k odhalení funkčního potenciálu lncRNA v různých biologických procesech. V následující kapitole uvádím dosavadní výsledky predikcí a experimentů zabývajících se funkční analýzou lncRNA u jednotlivých rostlinných organismů.

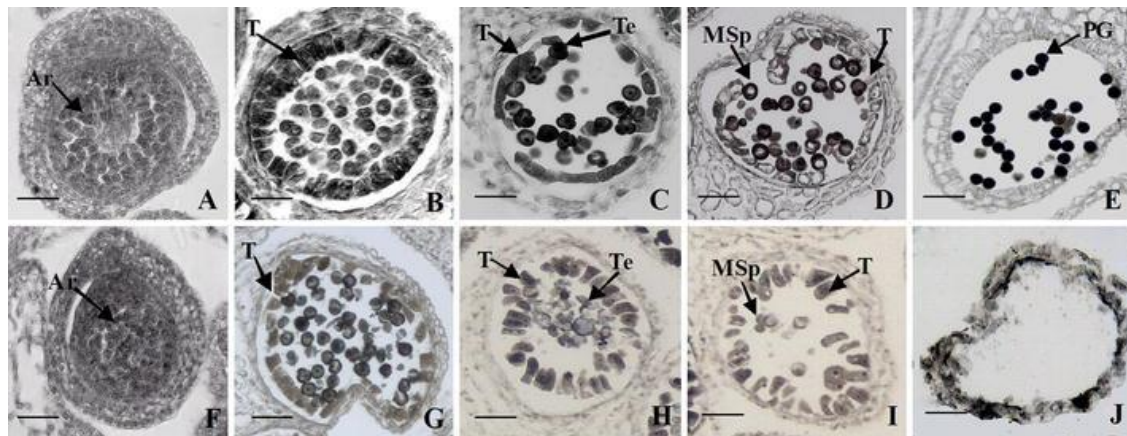
5.1 Reprodukce

Reprodukce organismů je velmi komplexní proces s mnohými stupni regulace, kdy z původního organismu vzniká další generace. Rostliny jsou schopné jak pohlavního, tak nepohlavního rozmnožování, obě varianty jsou detailně studovány nejen kvůli ekonomickému významu. Byla objevena řada miRNA podílejících se na různých fázích ontogenetického vývoje a růstu rostlin, jako je například vývoj tapeta či tvorba a uvolňování pylu. LncRNA COOLAIR, *antisense* transkript FLC genu, a COLDAIR, intronový transkript FLC genu, byly objeveny jako první dlouhé transkripty regulující vernalizaci přes supresi FLC (viz Epigenetická regulace a represe transkripce).

Vysoce významná pro výzkum plodin a šlechtění je 1 236 bazí dlouhá LDMAR (long-day-specific male-fertility-associated RNA). Pravděpodobně se jedná centrální regulátor u mutantů s fenotypem samčí sterility v závislosti na osvětlení (PSMS angl. photoperiod-sensitive male sterility). U mutantů se spontánní G–C mutací v promotorové oblasti RNA docházelo k poklesu tvorby LDMAR a následně k předčasné programované buněčné smrti u prašníku za dlouho-denních podmínek a tudíž k samčí sterilitě (Ding et al. 2012).

Dalším objevem učiněným na čínském zelí (*Brassica campestris*) je BcMF11 (*Brassica campestris* Male Fertility 11) o délce 828 bazí (Song et al. 2007). Bylo zjištěno, že mutanti se sníženou expresí BcMf11 vykazují mnohé aberace ve vývoji prašníků, dozrávání pylu (obr. 11 a celkové plodnosti (Song et al. 2013). Stejný fenomén u vývoje mikrospor byl pozorován i u transgenních rostlin kukuřice seté (*Zea mays*) se sníženou transkripcí Zm401 (Ma et al. 2008), 1,4 kilobazí dlouhou nekódující RNA specifickou v květech. Podobný fenotyp byl pozorován i prašníků a vývoje květu. U obou mutantů bylo dokázáno, že po zkřížení mutantní samice s divokým samcem nedochází ke snížení samičí fertility, mutace pravděpodobně ovlivňují pouze vývoj u samčích linií. Ani u jedné lncRNA nebyl zatím prokázán mechanismus

funkce, jedná se ale o mezidruhově konzervované sekvence. U Zm401 bylo dokázáno, že se nepodílí na vzniku miRNA, naopak u mutantů dochází ke změně v transkripci genů esenciálních pro vývoj pylu. Vedle významu pro sterilitu samců byla specificky u apomiktických rostlin rodu *Boecheerra* identifikována lncRNA BspUPG2, potenciálně klíčový



Obr. 11 Mikroskopické snímky srovnávající vývoj prašníků u transgenního čínské zeli (mutant v dolní řadě). A, F) archesporie. Bez rozdílu mezi transgenní rostlinou a WT. B, G) meiotická telofáze I. Nerovnoměrná tvorba a uspořádání tapeta u transgenního prašníku. C, H) tetrádová fáze. U transgenních rostlin nebylo možné pozorovat tetrády a docházelo ke zvětšování tapeta. D, I) pozdní jednojaderná fáze. U transgenních rostlin nedocházelo k rozpadu tapeta, mikrospory byly deformované a bez cytoplasmy. E, J) zralý pyl. U transgenní rostliny se pylová zrna nedokázala od sebe rozdělit a zůstávala v pylovém vaku spolu se zbytky tapeta. Ar – archeosporotická buňka, T – tapetum, Te – tetráda, Msp – mikrospory, PG – pylové zrno. Měřítka 20 μ m. Upraveno podle: Song et al. 2013

faktor pro dráhu neredukovaného vývoje pylu (Mau et al. 2013).

lncRNA mohou být i potenciální klíčoví hráči v regulaci vývoje meicytů a zároveň by mohly být faktorem genomového rozdílu divokých a domestikovaných rostlin či živočichů (Florez-Zapata et al. 2016). U výzkumu prováděném na slunečnici (*Helianthus annuus*), bylo zjištěno, že rozdíl v expresi lncRNA je výrazně vyšší než rozdíl v expresi u kódujících sekvencí právě mezi domestikovanými a divokými druhy. Tento jev by mohl být i důvodem rozdílu ve frekvenci homologní rekombinace. Jeden z faktorů hypoteticky zodpovědný za tento trend by mohl představovat vysoký obsah transponovatelných elementů (TE) v sekvencích lncRNA. Integrace TE do kódujících sekvencí může narušovat integritu genomu. Naopak integraci do sekvencí lncRNA nedochází k ohrožení organismu a TE jsou schopné se udržet a zachovat si funkci podmiňující potenciální rozdíl mezi domestikovaným a divokým jedincem. Jedná se ale zatím pouze o hypotetický účinek lncRNA na vývoj meicytů a jejich význam pro vlastnosti domestikovaných druhů (Florez-Zapata et al. 2016).

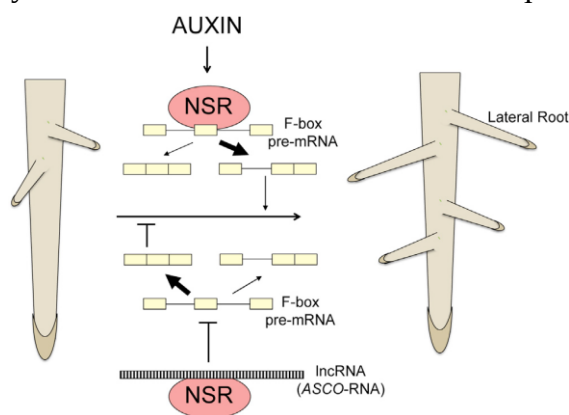
Jasný mechanismus působení a přesná funkce musí být opět teprve objeveny, na základě již zjištěných dat je ale s určitostí možné usoudit, že podíl dlouhých nekódujících RNA na vývoji květů a celkové reprodukci u rostlin je zásadní a budoucí výzkum lncRNA může vést

ke zdokonalení zemědělské produkce a celkovému porozumění komplexity celého rozmnožovacího procesu.

5.2 Nedostatek fosforečnanů, vývoj kořenů a prýtu

Fosfor je klíčový prvek pro většinu organismů podílející se na mnohých energetických a metabolických procesech. Jeho příjem a výdej je regulován složitou sítí faktorů jejichž nedílnou součástí tvoří pravděpodobně i lncRNA. MiR399 je jeden z hlavních post-transkripčních regulátorů fosfátové homeostáze (Fujii et al. 2005), jehož aktivita je ovlivněna dvěma lncRNA (IPS1 a AT4) výše zmíněným mechanismem maskování cílových miRNA. Xu a kolektiv provedli celogenomovou analýzu potenciálních lncRNA u rostlin trpících nedostatkem fosforečnanů. Za předpokladu kompetice jednotlivých lncRNA (případně i mRNA) o regulační miRNA a jejich analýzou rozdělili kompetentní endogenní RNA do skupin a pomocí RNA seq odhalili více než 1000 lncRNA souvisejících s nedostatkem fosforečnanů. Podobná analýza byla provedena i u *Arabidopsis thaliana* a vedla k objevu 227 nových lncRNA indukovaných či reprimovaných v případě nedostatku fosforečnanů (Xu et al. 2016; Yuan et al. 2016).

Mechanismus účinku byl objeven u dlouhých nekódujících sekvencí *ENOD40* a *ASCO-RNA* (původně lnc351). *ENOD40* je mimo jiné exprimován v hlízkách tvořící se v kořenech rostlin se symbiotickými bakteriemi fixujícími dusík (Campalans et al. 2004). Jeho účinek je mj. translokace RBP (proteinů vázajících ribozom, angl. *ribosome binding proteins*), MtRBP1, v diferencujících buňkách. *ASCO-RNA* ovlivňuje alternativní sestřih u auxinem indukovaného růstu laterálních kořenů mechanismem vyvazování regulačních proteinů, konkrétně NSR (nuclearspeckle RNA-binding proteins). V přítomnosti *ASCO-RNA* je NSR vyvazován a tvoří se isoforma mRNA nepodporující růst laterálních kořenů, zobrazeno v obr.



12. (Bardou et al. 2014). Růst laterálních kořenů je totiž v případě nedostatku fosforu podpořen na úkor růstu kořeně hlavního a umožňuje větší absorpční plochu.

Obr. 12 Regulace růstu laterálních kořenů mechanismem alternativního sestřihu. Upraveno podle: Kornbliht 2014

5.3 Abiotický stres

Na základě genomové analýzy bylo odhaleno, že abiotickým stresem (sucho, teplo, zasolení, zima, osvětlení) dochází k většímu rozdílu v síle exprese lncRNA než je rozdíl v expresi mRNA (Di et al. 2014) a několik RNA bylo i funkčně ověřeno v podílu na stresové odpovědi (Ben Amor et al. 2009). Mnohé lncRNA by navíc mohly být potenciálními cíli stresem indukovaných transkripčních faktorů, což by mohlo vést k objevu nových regulačních drah ve stresové odpovědi obecně. V letošní studii vyšla identifikace lncRNA exprimovaných ve stresových podmínkách (chlad, sucho) a povedlo se analyzovat 318 potenciálních transkriptů a na základě srovnávání se známými geny, které jsou zároveň exprimovány, byly odvozeny i případné mechanismy funkce, které potvrdily pravděpodobnou všestrannost působení lncRNA v regulačních drahách – tvorba miRNA, inhibiční funkce miRNA mechanismem maskování cílů, tvorba siRNA, trans a cis regulace transkripce (Li et al. 2017). LncRNA by tudíž mohly být klíčovou cestou ke zvyšování odolnosti plodin k abiotickým stresům, a tak ke zlepšení zemědělských produkcí.

5.4 Biotický stres

Dalším jevem, ovlivňujícím produkci s potenciálním rizikem v případě epidemií, je nákaza plodin virovými, bakteriálními nebo houbovými infekcemi. Opět byli objeveni mnozí kandidáti mezi nekódujícími RNA s potenciální rolí v regulaci obrany rostliny proti patogenu. Výzkum byl proveden na houbových infekcích pšenice způsobující padlí (*Blumeria graminis f. sp. tritici*) odhalující 71 potenciálních lncRNA se zvýšenou expresí v případě infekce (Xin et al. 2011). Jako zajímavou hypotézu dále uvedli, že dvě lncRNA by mohly působit jako prekursorů stejné miRNA v rozdílných pletivech, naznačující úzce propojený regulační systém. Podobný výzkum byl proveden u nákazy rzí plevové (*Puccinia striiformis f. sp. tritici*) napadající také pšenici. Odhalil 3 lncRNA indukované patogenem (H. Zhang et al. 2013). Další výzkumy byly provedeny u houbových infekcí *Fusarium oxysporum* a *Sclerotinia sclerotiorum* u huseníčku rolního (*Arabidopsis thaliana*) a brukve řepky (*Brassica napus*) odhalující jak prekursorů miRNA, tak antisense páry lncRNA-mRNA (Joshi et al. 2016; Zhu et al. 2014).

U bakteriálních infekcí Kwenda et al. (2016) provedli genomovou analýzu lincRNA u lilku bramboru (*Solanum tuberosum*) v odpovědi na infekci enterobakterií *P. carotovorum subsp. brasiliense*. Opět potvrdili funkci lncRNA jako prekursoru miRNA a regulátoru mRNA. Dále objevili 17 potenciálních lincRNA, jejichž exprese byla vysoce asociovaná s CDS geny

(Kwenda et al. 2016). U studia genové exprese *APETALA2* (*AP2*) v případě *virové* infekce mozaiky vodnice (Turnip crinkle virus) byla zjištěna záporná korelace v expresi *AP2* a sousedícího transkriptu nazvaném *Linc-AP2*, jehož přepis je v případě infekce indukován a mohl by být potenciální regulátor jak tvorby květu, tak obrany proti virovým onemocněním (Gao et al. 2016).

Mezi konkrétní příklady patří v letošním roce identifikovaná 589 bazí dlouhá lncRNA *ELENA1* (Elf18-induced long-noncoding RNA1). Rostliny se sníženou expresí *ELENA1* byly náchylnější k bakteriálnímu patogenu *Pseudomonas syringae*. Také u nich docházelo k poklesu exprese genu *PR1* (Patogenesis-related gene1). Zvýšená exprese *ELENA1* vedla k opačnému fenotypu a docházelo ke zvýšení exprese dalších genů indukovaných expresí *Elf18*. Jedním z pravděpodobných mechanismů funkce je aktivace transkripce asociací s mediátory. *ELENA1* patří k lncRNA ovlivňující nejen geny v blízkosti svého lokusu, potvrdilo se i její působení v trans (Seo et al. 2017).

5.5 Další funkce

Analýzy zabývající se objevem potenciálních lncRNA jsou prováděny na spektru, často zemědělsky či ekonomicky významných zástupců. Mezi rostliny, na které se výzkum lncRNA soustředí, patří kupříkladu bavlna (*Gossypium sp.*), topol černý (*Populus tomentosa*) nebo aktinidie lahodná (*Actinidia deliciosa*, kiwi). Bylo objeveno několik kandidátů lncRNA podílejících se na tvorbě vláken bavlny mechanismem prekurzorů miRNA nebo aktivátorů spolu exprimovaných genů (M. Wang et al. 2015; Zou et al. 2016). Ve výzkumu lncRNA indukovaných giberelinovou stimulací u topolu byly navrženy hypotetické funkce jako tvorba buněčné stěny, laterálních kořenů, podíl na auxinové signalizaci (kyselina giberilinová a auxin interakce) nebo syntéza pektinu a celulózy mechanismem ovlivnění cílových genů a stresových odpovědí v případě miRNA prekurzorů (J. Tian et al. 2016a). U topolu *Populus tomentosa* byly dále vybrány lncRNA zastávající funkci v metabolismu ligninu (Chen et al. 2015). Jako konkrétní zástupce byla objevena 2 286 bazí dlouhá lncRNA *NERDL*. Její funkce musí být ještě plně potvrzena, nicméně spolu s transkriptem *PtoNERDL* z cis lokusu tvoří regulační dvojici s vysokou expresí ve floému a xylému. Pravděpodobně se jedná o regulaci sekundárního tloušťnutí (Shi et al. 2017).

Zkratka	Název	Délka (bp)	Funkce/mutace	Mechanismus	model
COOLAIR	cold induced long antisense intragenic RNA	700	Chladem indukované umlčování FLC1	Interference transkripce (?)	<i>Arabidopsis thaliana</i>
COLDAIR	Cold assisted intronic noncoding RNA	1 100	Umlčování FLC po vernalizaci	Epigenetická represe	<i>Arabidopsis thaliana</i>
LDMAR	long-day-specific male-fertility-associated RNA	1 236	PSMS	neznámý	<i>Oryza sativa</i>
BspUPG2	Boecherra spp. UPGRADE 2	3 156	Iniciace neredukovaného vývoje pylu	premiRNA(?), změna DNA metylace(?)	<i>Boecherra sp.</i>
BcMF11	Brassica campestris Male Fertility 11	828	Vývoj pylu a prašníků	neznámý	<i>Brassica campestris</i>
Zm401	Zea mays 401	1 400	Vývoj pylu a prašníků	neznámý	<i>Zea mays</i>
At4	Arabidopsis thaliana 4	750	Fosfátová homeostáze	Maskování cílů miRNA	<i>Arabidopsis thaliana</i>
IPS1	Induced by phosphate starvation 1	540	Fosfátová homeostáze	Maskování cílů miRNA	<i>Arabidopsis thaliana</i>
ASCO-RNA	Alternative splicing competitor RNA		Inhibice růstu laterálních kořenů u Auxin-indukované odpovědi	Alternativní sestřih	<i>Medicago truncatula</i>
Enod40	Early nodulin gene 40		Lokalizace RNP částic	neznámý	<i>Medicago truncatula</i>
Linc-AP2	Long intergenic non-coding RNA APETALA2		Tvorba květu, obrana proti patogenům	neznámý	<i>Arabidopsis thaliana</i>
ELENA1	Elf18-induced long-noncoding RNA1	589	Rezistence k patogenu <i>Pseudomonas syringae</i>	Aktivace transkripce asociací s mediátorem	<i>Tomato sp.</i>
NERDL	NERD	2 286	Sekundární tloušťnutí (?)	neznámý	<i>Populus tomentosa</i>

Tab. 2: Shrnutí lncRNA rozebíraných v této kapitole.

6. Závěr

Cílem této bakalářské práce bylo shrnout dosavadní znalosti a výsledky výzkumu týkajícího se dlouhých nekódujících RNA se specializací na rostliny. Ačkoliv ještě není objeveno a funkčně analyzováno takové množství lncRNA jako u živočichů, je jasné, že jejich role v rostlinách je klíčová. Pro svou tkáňovou specifitu a možnost rychlé a přesné regulace se lncRNA podílejí na odpovědi na stresové podmínky, a to jak biotické, tak abiotické. Fungují jako regulátory drah odpovídající vlivům prostředí a jsou klíčovými hráči v reprodukci rostlin. Mohou působit na několika úrovních, počínaje regulací transkripce, post-transkripčních procesů, translace a konče samotstabilitou samotných proteinů. Dále mohou fungovat jako prekursori krátkých nekódujících RNA, taktéž krucálních elementů regulace genové exprese u rostlin.

Momentálním cílem a budoucností výzkumu lncRNA je experimentální analýza funkcí jednotlivých transkriptů vedoucí k většímu porozumění mechanismu jejich účinku. Na základě dalších dat bude možné sestavit přesnější klasifikace a parametry definující tuto skupinu. Nedílnou součástí postupu je i vývoj metod pro rozlišení nekódujících sekvencí od kódujících a zpřesnění jejich bioinformatické analýzy. Myslím, že nejen lncRNA, ale nekódující sekvence obecně, přesahují svým potenciálem v celkové komplexitě regulace organismů veškeré dosavadní znalosti a přinesou mnohé nové objevy.

7. Seznam použitých citací

- Bardou, F., et al. (2014), 'Long noncoding RNA modulates alternative splicing regulators in *Arabidopsis*', *Dev Cell*, 30 (2), 166-76.
- Ben Amor, B., et al. (2009), 'Novel long non-protein coding RNAs involved in *Arabidopsis* differentiation and stress responses', *Genome Res*, 19 (1), 57-69.
- Bernstein, E. and Allis, C. D. (2005), 'RNA meets chromatin', *Genes Dev*, 19 (14), 1635-55.
- Bonasio, R. and Shiekhattar, R. (2014), 'Regulation of transcription by long noncoding RNAs', *Annu Rev Genet*, 48, 433-55.
- Bussotti, G., Notredame, C., and Enright, A. J. (2013), 'Detecting and comparing non-coding RNAs in the high-throughput era', *Int J Mol Sci*, 14 (8), 15423-58.
- Campalans, A., Kondorosi, A., and Crespi, M. (2004), 'Enod40, a short open reading frame-containing mRNA, induces cytoplasmic localization of a nuclear RNA binding protein in *Medicago truncatula*', *Plant Cell*, 16 (4), 1047-59.
- Chaumeil, J., et al. (2006), 'A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced', *Genes Dev*, 20 (16), 2223-37.
- Chen, J., Quan, M., and Zhang, D. (2015), 'Genome-wide identification of novel long non-coding RNAs in *Populus tomentosa* tension wood, opposite wood and normal wood xylem by RNA-seq', *Planta*, 241 (1), 125-43.
- CRICK, F. H. (1958), 'On protein synthesis', *Symp Soc Exp Biol*, 12, 138-63.
- Di, C., et al. (2014), 'Characterization of stress-responsive lncRNAs in *Arabidopsis thaliana* by integrating expression, epigenetic and structural features', *Plant J*, 80 (5), 848-61.
- Di Ruscio, A., et al. (2013), 'DNMT1-interacting RNAs block gene-specific DNA methylation', *Nature*, 503 (7476), 371-6.
- Ding, J. H., et al. (2012), 'A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice', *P Natl Acad Sci USA*, 109 (7), 2654-59.
- Dinger, M. E., et al. (2008), 'Differentiating protein-coding and noncoding RNA: challenges and ambiguities', *PLoS Comput Biol*, 4 (11), e1000176.
- Fenoglio, C., et al. (2013), 'An emerging role for long non-coding RNA dysregulation in neurological disorders', *Int J Mol Sci*, 14 (10), 20427-42.
- Florez-Zapata, N. M. V., Reyes-Valdes, M. H., and Martinez, O. (2016), 'Long non-coding RNAs are major contributors to transcriptome changes in sunflower meiocytes with different recombination rates', *Bmc Genomics*, 17.
- Franco-Zorrilla, J. M., et al. (2007), 'Target mimicry provides a new mechanism for regulation of microRNA activity', *Nat Genet*, 39 (8), 1033-7.
- Fritah, S., Niclou, S. P., and Azuaje, F. (2014), 'Databases for lncRNAs: a comparative evaluation of emerging tools', *RNA*, 20 (11), 1655-65.
- Fujii, H., et al. (2005), 'A miRNA involved in phosphate-starvation response in *Arabidopsis*', *Curr Biol*, 15 (22), 2038-43.
- Gao, R., et al. (2016), 'Upregulation of LINC-AP2 is negatively correlated with AP2 gene expression with Turnip crinkle virus infection in *Arabidopsis thaliana*', *Plant Cell Rep*, 35 (11), 2257-67.
- Heo, J. B. and Sung, S. (2011), 'Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA', *Science*, 331 (6013), 76-9.
- Hirota, K., et al. (2008), 'Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs', *Nature*, 456 (7218), 130-4.

- Hirsch, J., et al. (2006), 'Characterization of 43 non-protein-coding mRNA genes in *Arabidopsis*, including the MIR162a-derived transcripts', *Plant Physiol*, 140 (4), 1192-204.
- Joshi, R. K., et al. (2016), 'Genome Wide Identification and Functional Prediction of Long Non-Coding RNAs Responsive to *Sclerotinia sclerotiorum* Infection in *Brassica napus*', *PLoS One*, 11 (7), e0158784.
- Kang, C. Y. and Liu, Z. C. (2015), 'Global identification and analysis of long non-coding RNAs in diploid strawberry *Fragaria vesca* during flower and fruit development', *Bmc Genomics*, 16.
- Kapusta, A., et al. (2013), 'Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs', *PLoS Genet*, 9 (4), e1003470.
- Khemka, N., et al. (2016), 'Genome-wide analysis of long intergenic non-coding RNAs in chickpea and their potential role in flower development', *Sci Rep*, 6, 33297.
- Khorkova, O., et al. (2014), 'Natural antisense transcripts', *Hum Mol Genet*, 23 (R1), R54-63.
- Kornblihtt, A. R. (2014), 'A long noncoding way to alternative splicing in plant development', *Dev Cell*, 30 (2), 117-9.
- Kung, J. T., Colognori, D., and Lee, J. T. (2013), 'Long noncoding RNAs: past, present, and future', *Genetics*, 193 (3), 651-69.
- Kwenda, S., Birch, P. R., and Moleleki, L. N. (2016), 'Genome-wide identification of potato long intergenic noncoding RNAs responsive to *Pectobacterium carotovorum* subspecies brasiliense infection', *BMC Genomics*, 17 (1), 614.
- Li, S., et al. (2015), 'Detection of Pol IV/RDR2-dependent transcripts at the genomic scale in *Arabidopsis* reveals features and regulation of siRNA biogenesis', *Genome Res*, 25 (2), 235-45.
- Li, S., et al. (2017), 'Genome-wide identification and functional prediction of cold and/or drought-responsive lncRNAs in *cassava*', *Sci Rep*, 7, 45981.
- Ma, J., et al. (2008), 'Zm401, a short-open reading-frame mRNA or noncoding RNA, is essential for tapetum and microspore development and can regulate the floret formation in maize', *J Cell Biochem*, 105 (1), 136-46.
- Matsui, K., et al. (2008), 'Natural antisense transcript stabilizes inducible nitric oxide synthase messenger RNA in rat hepatocytes', *Hepatology*, 47 (2), 686-97.
- Mattick, J. S. (2001), 'Non-coding RNAs: the architects of eukaryotic complexity', *EMBO Rep*, 2 (11), 986-91.
- Mattick, J. S. and Rinn, J. L. (2015), 'Discovery and annotation of long noncoding RNAs', *Nat Struct Mol Biol*, 22 (1), 5-7.
- Mau, M., et al. (2013), 'The conserved chimeric transcript UPGRADE2 is associated with unreduced pollen formation and is exclusively found in apomictic *Boechera* species', *Plant Physiol*, 163 (4), 1640-59.
- Mirsky, A. E. and Ris, Hans (April 1951), 'The desoxyribonucleic acid content of animal cells and its evolutionary significance', *J Gen Physiol*, 34 (4), 451-62
- Morrissey, A. S., Griffith, M., and Marra, M. A. (2011), 'Extensive relationship between antisense transcription and alternative splicing in the human genome', *Genome Res*, 21 (8), 1203-12.
- Munroe, S. H. (1988), 'Antisense RNA inhibits splicing of pre-mRNA in vitro', *EMBO J*, 7 (8), 2523-32.
- Nagalakshmi, U., et al. (2008), 'The transcriptional landscape of the yeast genome defined by RNA sequencing', *Science*, 320 (5881), 1344-9.
- Ohno, Susumu (1972), 'Origin of the term "Junk DNA"', 23, 366-70.

- Okazaki, Y., et al. (2002), 'Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs', *Nature*, 420 (6915), 563-73.
- Pang, K. C., Frith, M. C., and Mattick, J. S. (2006), 'Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function', *Trends Genet*, 22 (1), 1-5.
- Perteau, M. and Salzberg, S. L. (2010), 'Between a chicken and a grape: estimating the number of human genes', *Genome Biol*, 11 (5), 206.
- Salmena, L., et al. (2011), 'A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language?', *Cell*, 146 (3), 353-8.
- Seo, J. S., et al. (2017), 'ELF18-INDUCED LONG-NONCODING RNA Associates with Mediator to Enhance Expression of Innate Immune Response Genes in *Arabidopsis*', *Plant Cell*, 29 (5), 1024-38.
- Shi, W., et al. (2017), 'The Interactions between the Long Non-coding RNA NERDL and Its Target Gene Affect Wood Formation in *Populus tomentosa*', *Front Plant Sci*, 8, 1035.
- Song, J. H., Cao, J. S., and Wang, C. G. (2013), 'BcMF11, a novel non-coding RNA gene from *Brassica campestris*, is required for pollen development and male fertility', *Plant Cell Rep*, 32 (1), 21-30.
- Song, J. H., et al. (2007), 'BcMF11, a putative pollen-specific non-coding RNA from *Brassica campestris* ssp. *chinensis*', *J Plant Physiol*, 164 (8), 1097-100.
- St Laurent, G., Wahlestedt, C., and Kapranov, P. (2015), 'The Landscape of long noncoding RNA classification', *Trends Genet*, 31 (5), 239-51.
- Sun, L., et al. (2012), 'Prediction of novel long non-coding RNAs based on RNA-Seq data of mouse Klf1 knockout study', *BMC Bioinformatics*, 13, 331.
- Thomas, C. A. (1971), 'The genetic organization of chromosomes', *Annu Rev Genet*, 5, 237-56.
- Tian, J., et al. (2016a), 'Population genomic analysis of gibberellin-responsive long non-coding RNAs in *Populus*', *J Exp Bot*, 67 (8), 2467-82.
- Trapnell, C., et al. (2012), 'Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks', *Nat Protoc*, 7 (3), 562-78.
- Vogel, F. (1964), 'A Preliminary Estimate of the Number of Human Genes', *Nature*, 201, 847.
- Wang, D., et al. (2017), 'Transposable elements (TEs) contribute to stress-related long intergenic noncoding RNAs in plants', *Plant J*, 90 (1), 133-46.
- Wang, H., et al. (2014), 'Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in *Arabidopsis*', *Genome Res*, 24 (3), 444-53.
- Wang, K. C. and Chang, H. Y. (2011), 'Molecular mechanisms of long noncoding RNAs', *Mol Cell*, 43 (6), 904-14.
- Wang, M., et al. (2015), 'Long noncoding RNAs and their proposed functions in fibre development of cotton (*Gossypium* spp.)', *New Phytol*, 207 (4), 1181-97.
- Wapinski, O. and Chang, H. Y. (2011), 'Long noncoding RNAs and human disease', *Trends Cell Biol*, 21 (6), 354-61.
- Wierzbicki, A. T., Haag, J. R., and Pikaard, C. S. (2008), 'Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes', *Cell*, 135 (4), 635-48.
- Xin, M., et al. (2011), 'Identification and characterization of wheat long non-protein coding RNAs responsive to powdery mildew infection and heat stress by using microarray analysis and SBS sequencing', *BMC Plant Biol*, 11, 61.
- Xu, X. W., et al. (2016), 'Functional analysis of long intergenic non-coding RNAs in phosphate-starved rice using competing endogenous RNA network', *Sci Rep*, 6, 20715.
- Yuan, J. P., et al. (2016), 'Systematic characterization of novel lncRNAs responding to phosphate starvation in *Arabidopsis thaliana*', *Bmc Genomics*, 17.

- Zhang, H., et al. (2013), 'Long non-coding genes implicated in response to stripe rust pathogen stress in wheat (*Triticum aestivum* L.)', *Mol Biol Rep*, 40 (11), 6245-53.
- Zhang, X., et al. (2007), 'Role of RNA polymerase IV in plant small RNA metabolism', *Proc Natl Acad Sci U S A*, 104 (11), 4536-41.
- Zhang, YC, et al. (2014), 'Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice', *Genome Biol*, 15 (12).
- Zhao, J., et al. (2008), 'Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome', *Science*, 322 (5902), 750-6.
- Zhu, Q. H., et al. (2014), 'Long noncoding RNAs responsive to *Fusarium oxysporum* infection in *Arabidopsis thaliana*', *New Phytol*, 201 (2), 574-84.
- Zou, C., et al. (2016), 'Transcriptome analysis reveals long noncoding RNAs involved in fiber development in cotton (*Gossypium arboreum*)', *Sci China Life Sci*, 59 (2), 164-71.