Bioinformatique des ARNs
Boussens
February 24 - 26, 2010

Small RNAs and X inactivation

Constance Ciaudo, Olivier Voinnet's lab
Constance.Ciaudo@ibmp-ulp.u-strasbg.fr
Institut de Biologie Moléculaire des Plantes du CNRS
12, rue du Général Zimmer
67084 Strasbourg Cedex
France

Constance Ciaudo, Edith Heard's lab
Constance.Ciaudo@curie.fr
Mammalian Developmental Epigenetics Group
UMR 3215 - Nuclear Dynamics and Genome Plasticity
Curie Institute - Section of Research
BDD - 11-13 rue Pierre et Marie Curie
75248 Paris Cedex 05 - FRANCE

Outline

1. Introductions:
    1. RNAi pathways
    2. ES cells model
    3. Next generation sequencing approaches

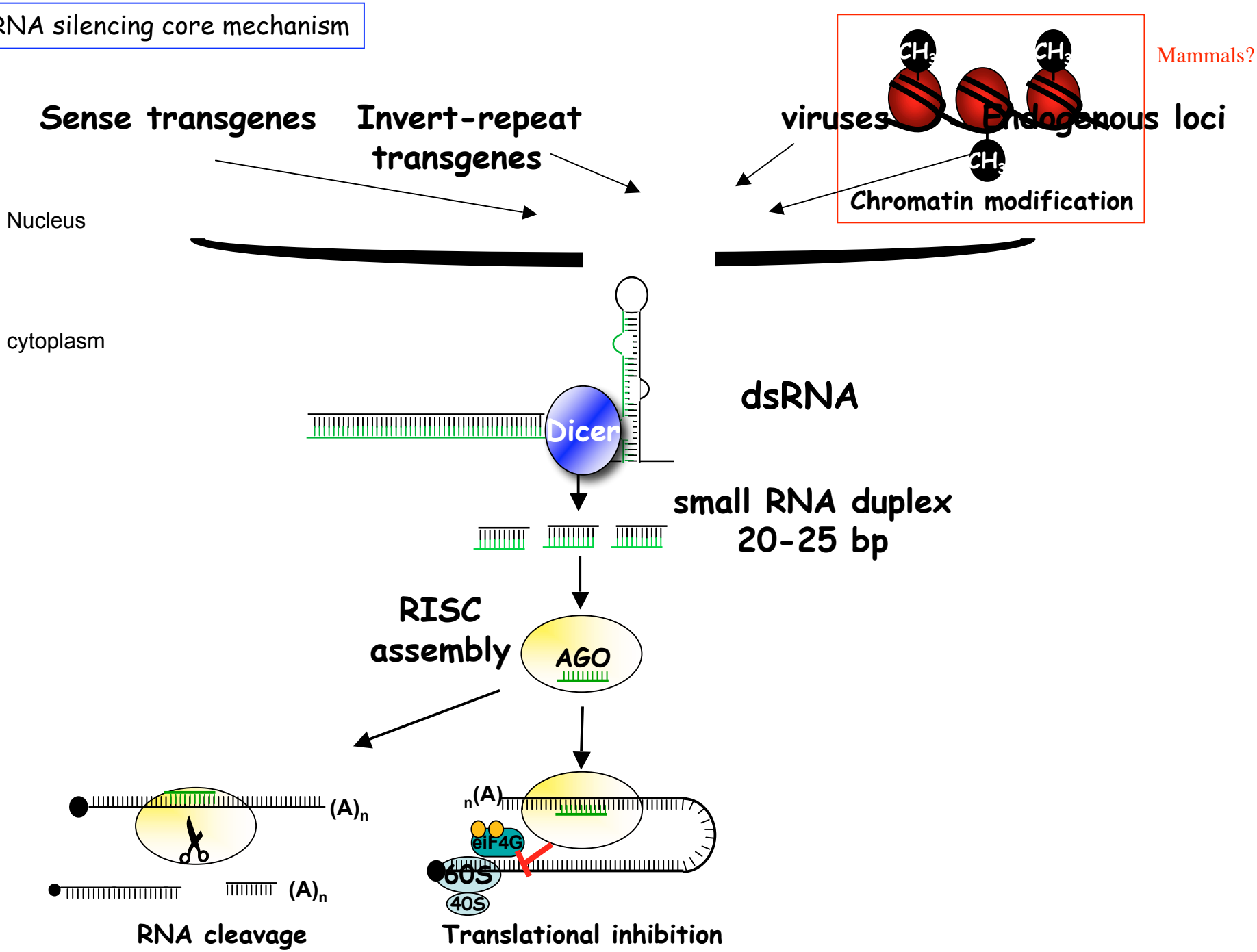
2. Bio-informatic workflow used for small RNA run analysis
    1. Terminology
    2. Reads mapping and annotation
    3. Comparison of tools & sequencing approaches

3. Profiling of microRNAs during ES cells differentiation

4. Small RNAs involved in X inactivation process?

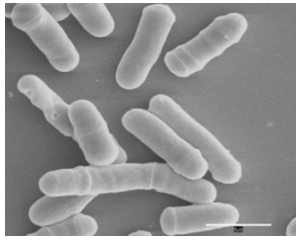
1. Introductions

RNA silencing core mechanism

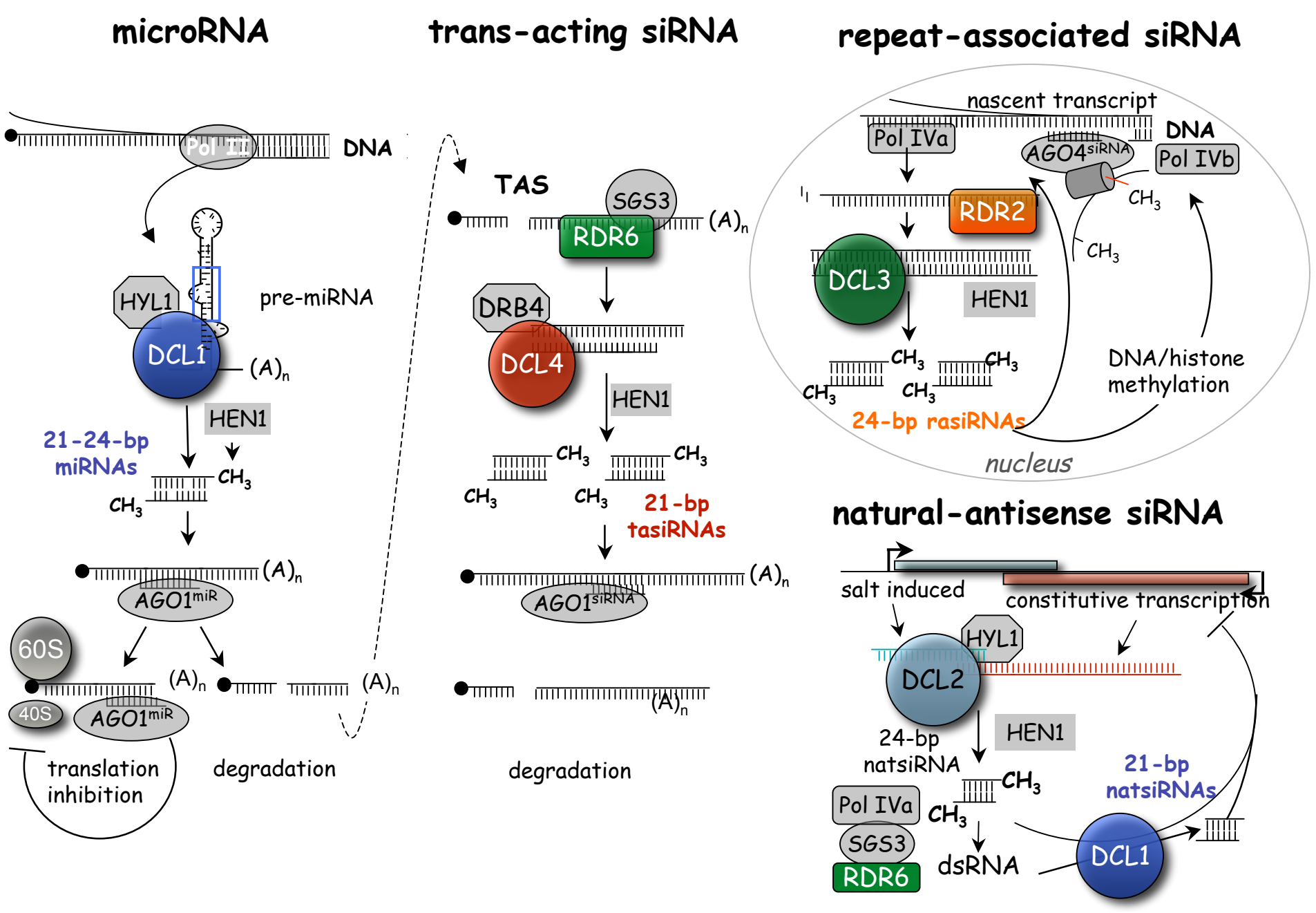


RNAi proteins in different organisms

	Human	Mice	Drosophila	C. elegans	Pombe	A. thaliana
Dicer	1	1	2	1	1	4
Argonaute	4+4	4+3	3	27	1	10
RdRp	None	None	None	3	1	6



RNAi in plant



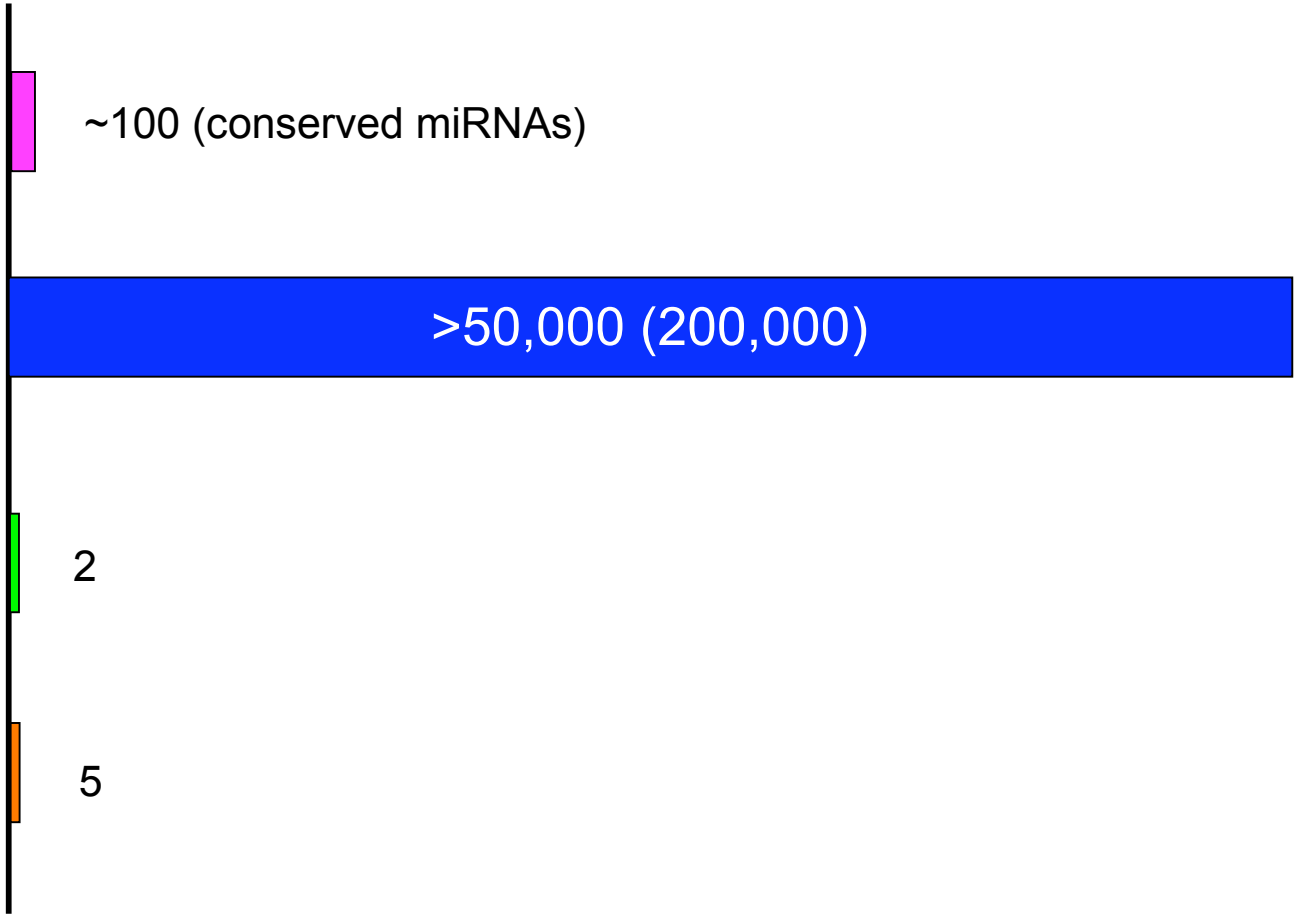
Four Dicers in *Arabidopsis thaliana*

DCL1: 20-25nt miRNA
-Development
-many other processes

DCL3: 24nt siRNA
-DNA methylation
-Heretochromatin
-TGS

DCL2: 22nt siRNA
-viral-derived siRNAs

DCL4: 21nt siRNA
-viral-derived siRNAs
-TasiRNAs
- 'Young' miRNAs



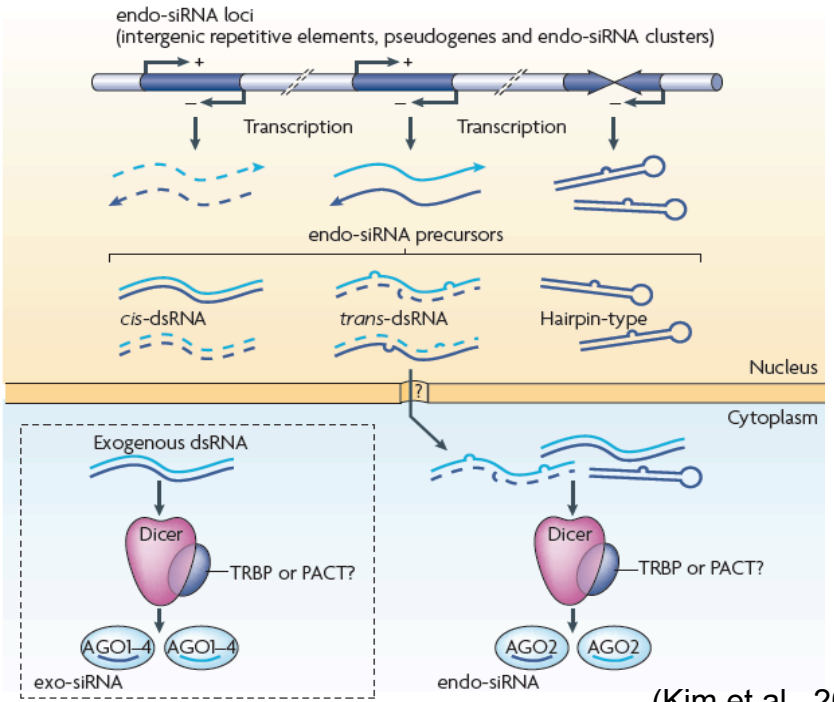
One Dicer in *Mus Musculus*

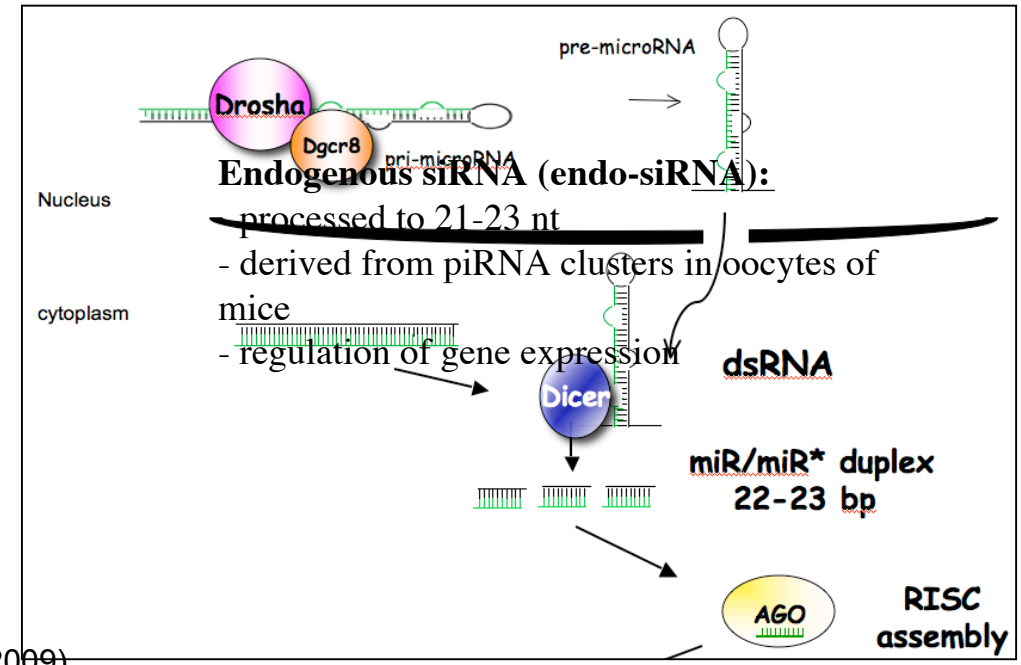
DCL1: - 22-23nt miRNA

miRNA cloned in undifferentiated ES cells

- ≈21nt endo-siRNA?

Putative siRNA dependent of Dicer?





Endogenous siRNA (endo-siRNA):

- processed to 21-23 nt
- derived from piRNA clusters in oocytes of mice
- regulation of gene expression

(Kim et al., 2009)

Next generation sequencing approaches

Genome

- *De Novo* Sequencing
- Targeted Resequencing
- Whole Genome Resequencing

Transcriptome

- Gene Expression Profiling
- Small RNA Analysis
- Whole Transcriptome Analysis

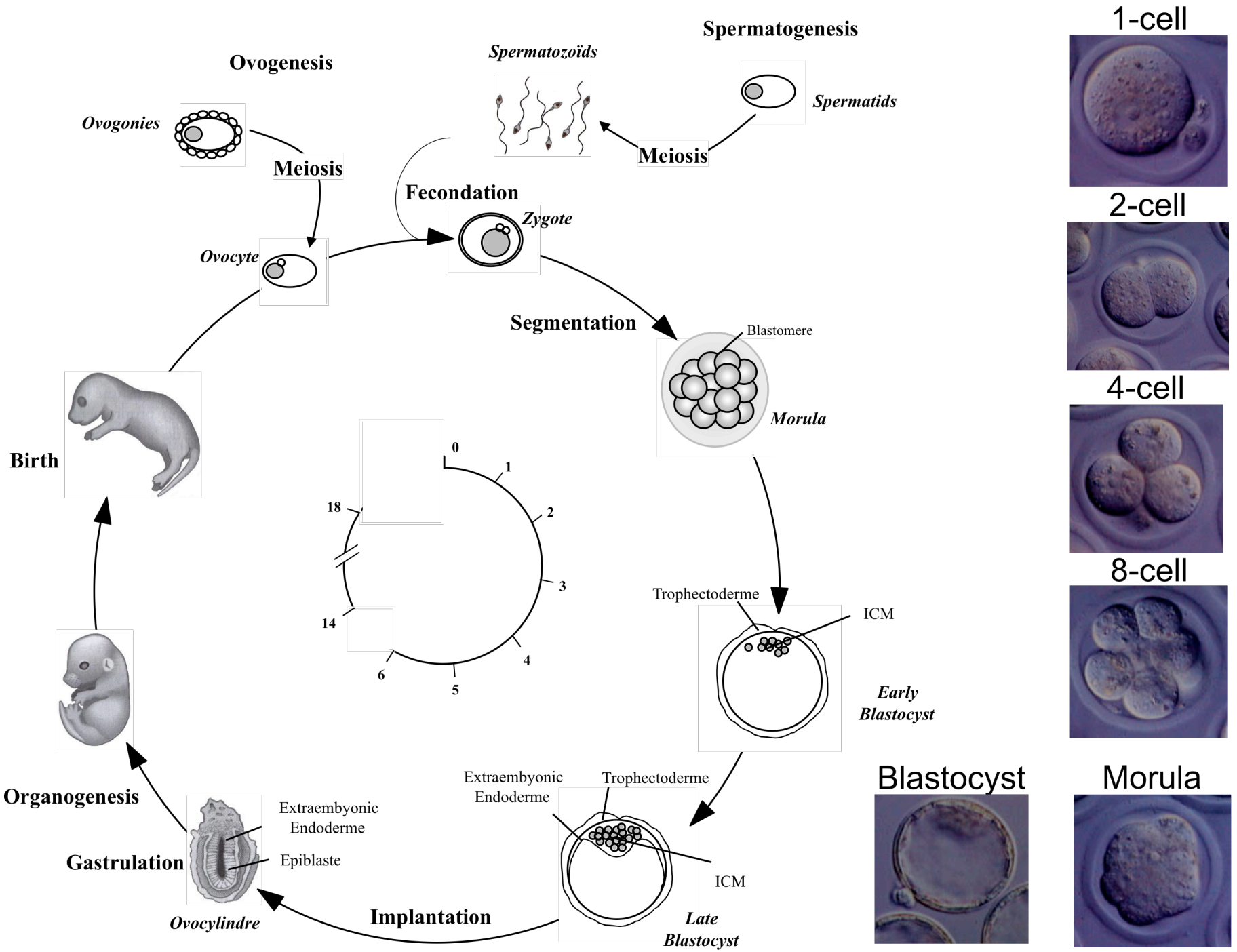
Interactome
- 4-5C analysis

Epigenome

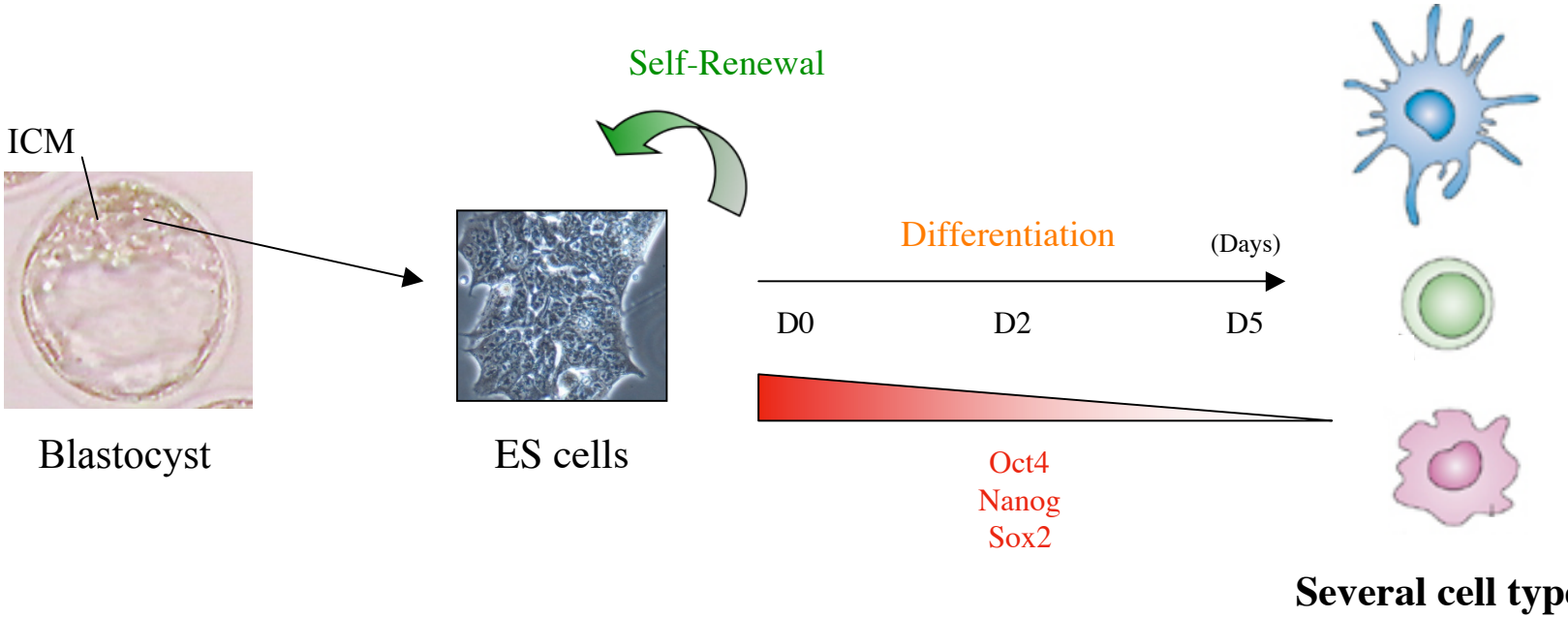
- Chromatin Immunoprecipitation Sequencing (ChIP-Seq)
- Methylation Analysis

(Modified from ABI website)

Small RNA profiling in mouse embryonic stem cells

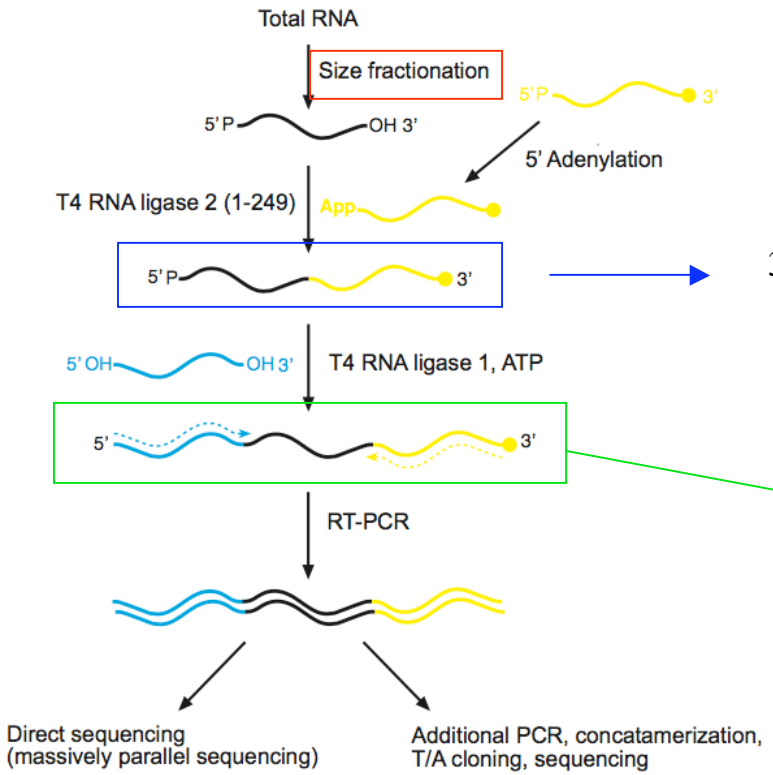

Spermatogenesis
Ovogenesis
Spermatozoïds
Spermatids
Ovogonies
Meiosis
Meiosis
Fecondation
Zygote
Ovocyte
Segmentation
Blastomere
Morula
Birth
0
1
2
3
4
5
6
14
18
Trophectoderme
ICM
Early Blastocyst
Organogenesis
Extraembyonic Endoderme
Trophectoderme
Extraembyonic Endoderme
Epiblaste
Gastrulation
ICM
Ovocylindre
Implantation
Late Blastocyst
1-cell
2-cell
4-cell
8-cell
Blastocyst
Morula

Dynamics of small RNAs during ES cell differentiation



Cell lines used: XX ES cells / XY ES cells

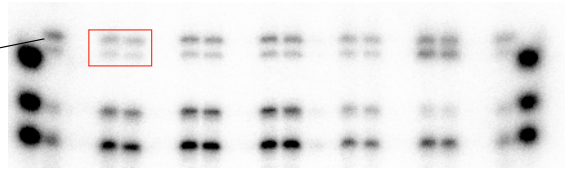
Small RNA libraries protocol



(1 week more for this "test sequencing" step)

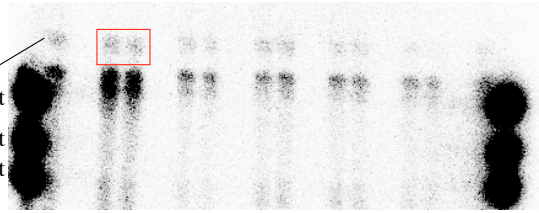
Isolation small RNA

3' ligated markers



3' ligation

5' ligated markers



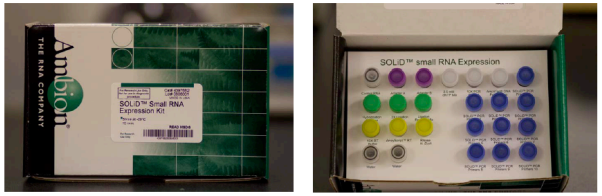
5' ligation

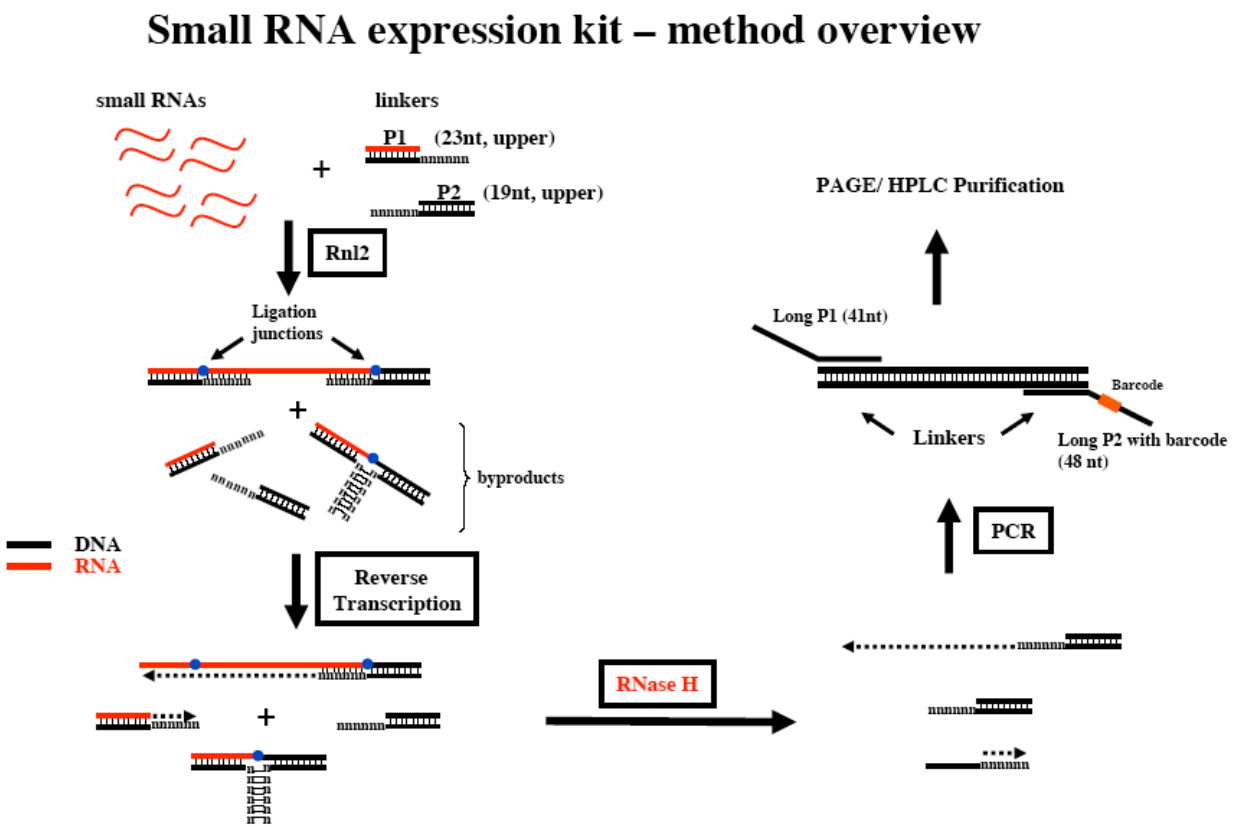
- 5 days
- 3 radioactive polyacrylamid gel
- did with minimum 4 ug of total RNA

(Pfeffer et al., *Methods and Enzymol*, 2007)

SOLID small RNA library kit

The SOLiD Small RNA Expression kit





No radioactivity
Done in ONE DAY
Start with low quantity of total RNA

Sequencing methods

Table 1

Summary of massively parallel sequencing technologies

Method	Amplification	Read length (base pairs)	Templates per run	Data production/ day	Sequence reaction	Reference
Commercially available technologies						
ABI 3730xl	PCR	~900 to 1,100	96	1 Mb/day	Sanger method	www.appliedbyosystems.com
454 FLX Roche	Emulsion PCR	~400	1,000,000	400 Mb/run/ 7.5 to 8 hours	Pyrosequencing	www.rocheapplied-science.com
Illumina (Solexa) Genome Analyzer	Bridge PCR	36 to 175	40,000,000	>17 Gb/run/ 3 to 6 days	Reverse terminator	www.illumina.com
ABI SOLiD	Emulsion PCR	~50	85,000,000	10 to 15 Gb/ run/6 days	Ligation sequencing	www.appliedbyosystems.com
Helicos Heliscope	None	30 to 35	800,000,000	21 to 28 Gb/ run/8 days	Single molecule sequence by synthesis	www.helicosbio.com
Technologies in development						
Pacific Biosciences	None	>1,000	NA	NA	Single molecule real-time DNA sequencing	www.pacificbiosciences.com
Intelligent Biosciences	Yes*	NA	NA	NA	Sequence by synthesis	www.intelligentbiosystems.com
Visigen Biotechnologies	None	NA	NA	NA	Base-specific FRET emission	www.visigenbio.com
ZS Genetics	None	NA	NA	NA	ZSG atomic labelling and electron microscopy	www.zsgenetics.com

NA, not available at present. *Amplification method not yet standardised. FRET, Förster resonance energy transfer.

(Jorge S Reis-Filho., 2009)

Sequencing methods

454 sequencing:

Emulsion PCR
Pyrosequencing of ~400 bp
1 000 000 reads per run

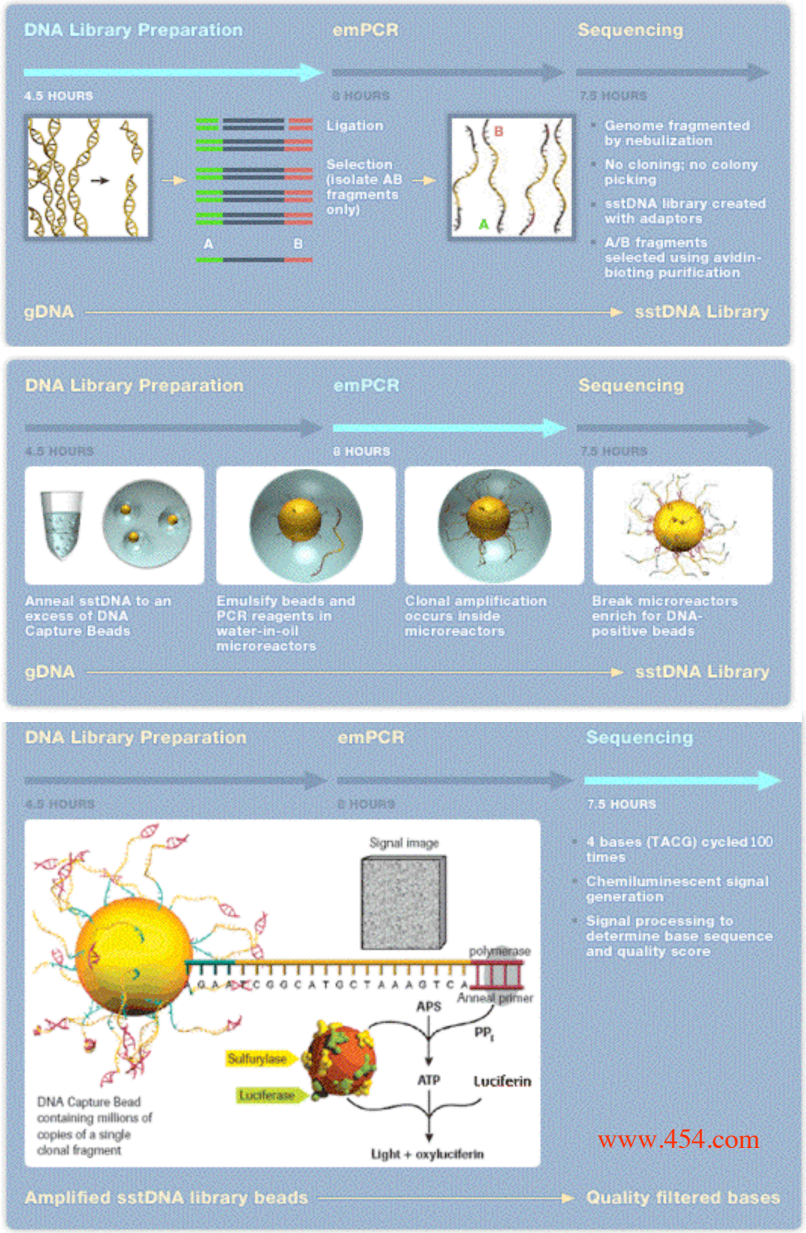
Example of barcoding libraries:

5' primer

1	CAG**GCAT**CGGAATTCCTCACTAAA
2	CAG**GATC**CGGAATTCCTCACTAAA
3	CAG**TGCA**CGGAATTCCTCACTAAA
4	CAG**TAGC**CGGAATTCCTCACTAAA
5	CAG**TGAC**CGGAATTCCTCACTAAA
6	CAG**TACG**CGGAATTCCTCACTAAA

3' primer

1'	GAC**CGTA**TGGAATTCGCGGTTAAA
2'	GAC**CTAG**TGGAATTCGCGGTTAAA
3'	GAC**ACGT**TGGAATTCGCGGTTAAA
4'	GAC**ATCG**TGGAATTCGCGGTTAAA
5'	GAC**ACTG**TGGAATTCGCGGTTAAA
6'	GAC**TAGC**TGGAATTCGCGGTTAAA

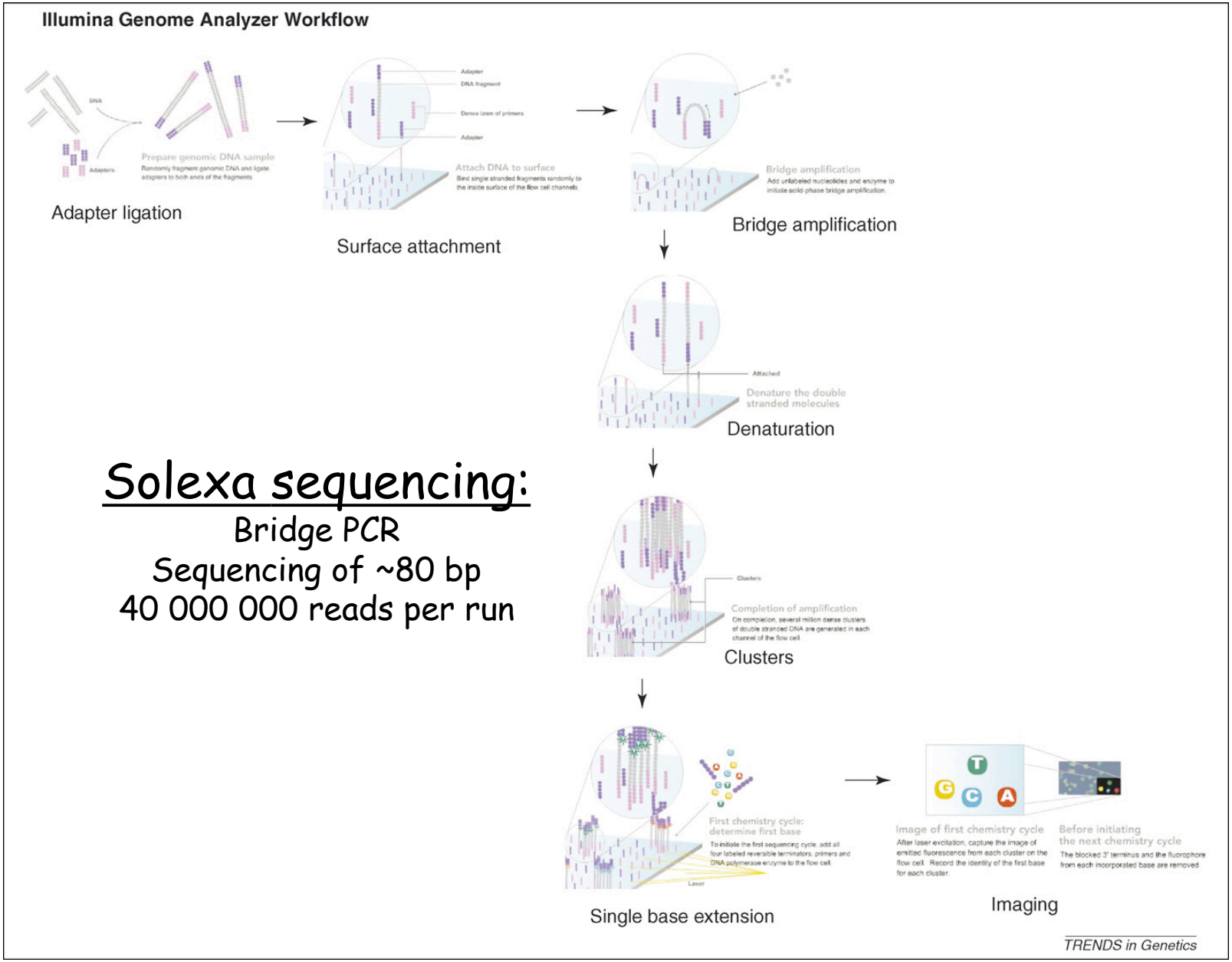


Some sequencing bias can be introduce using some barcodes (see Binladen et al., 2007)

(Genoscope sequencing)

Sequencing methods

www.illumina.com



Solexa sequencing:

Bridge PCR

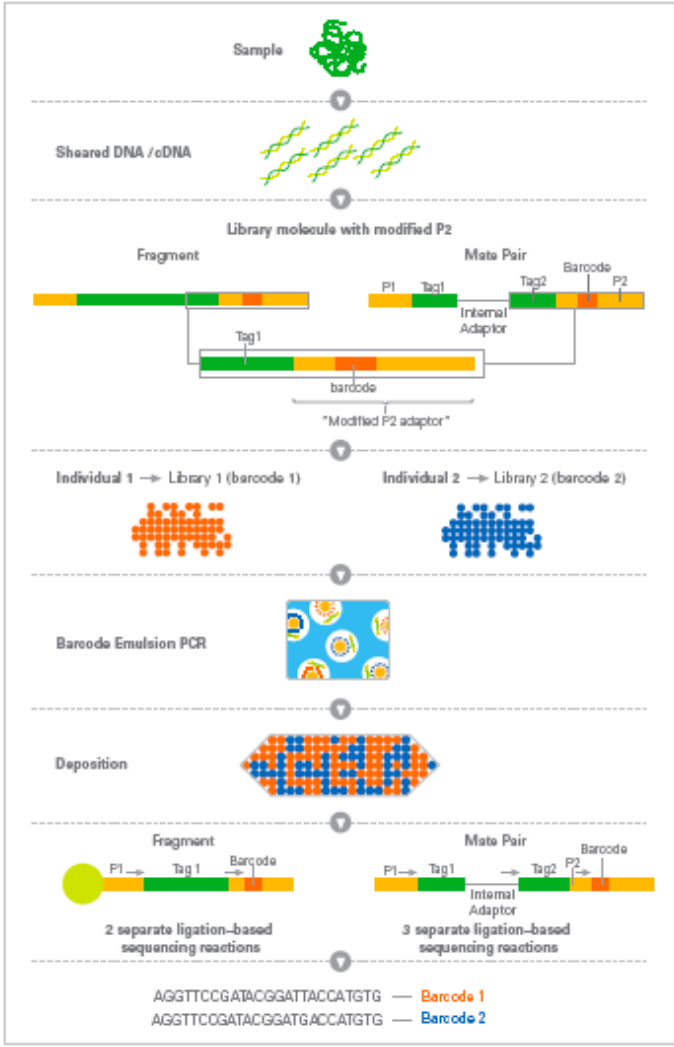
Sequencing of ~80 bp

40 000 000 reads per run

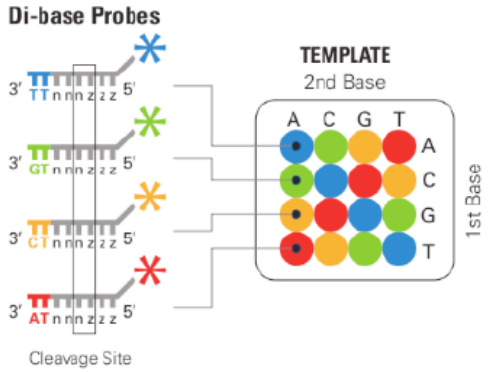
(Mardis, 2008)

Sequencing methods

SOLID sequencing:
Emulsion PCR
Sequencing of ~50 bp
85 000 000 reads per run



- séquençage par ligation séquentielle de sondes fluorescentes
  - 16 sondes (dinucléotides) → 4 fluorescences

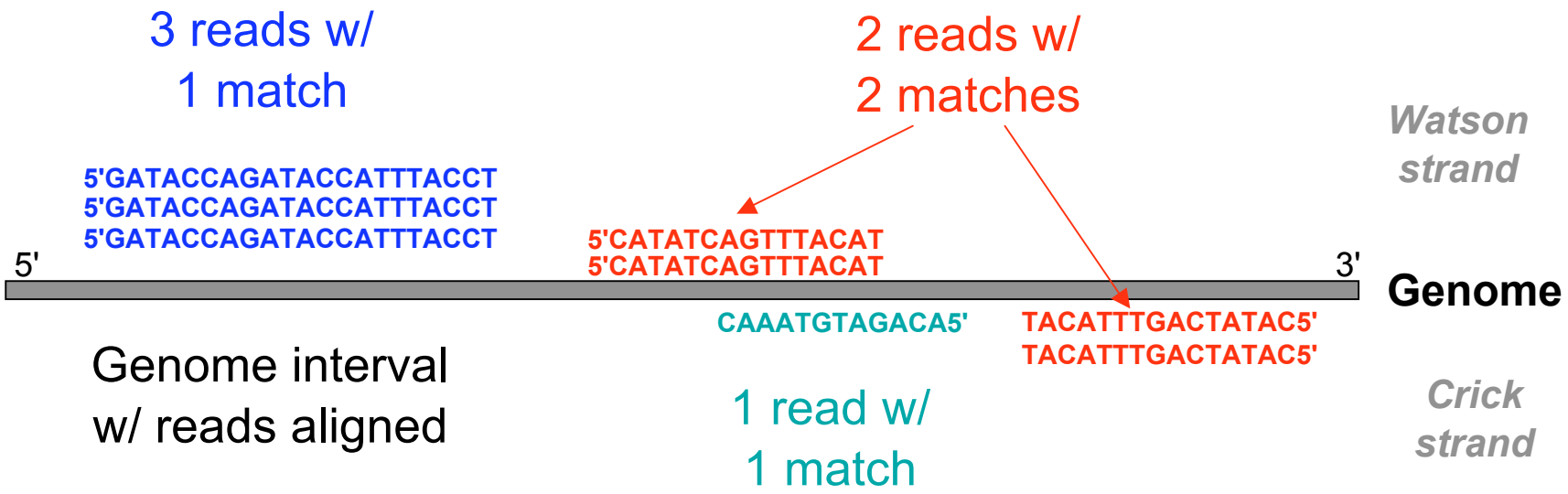


(Mardis, 2008)

2. Bio-informatic workflow used for small RNA run analysis

Terminology

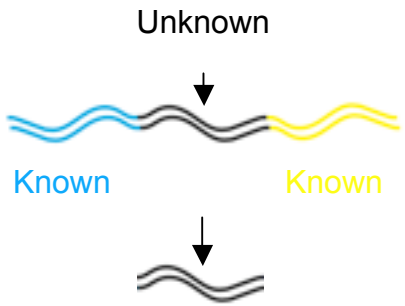
- *Read*: short nucleotide sequence output by the sequencer
- *Mapping*: Reads are *aligned* to genome, allowing mismatches (sequencing errors, SNPs)
- Aligned reads match one or more *genome intervals*
- *Annotation*: determine the type of read (miRNA, rRNA, etc.)



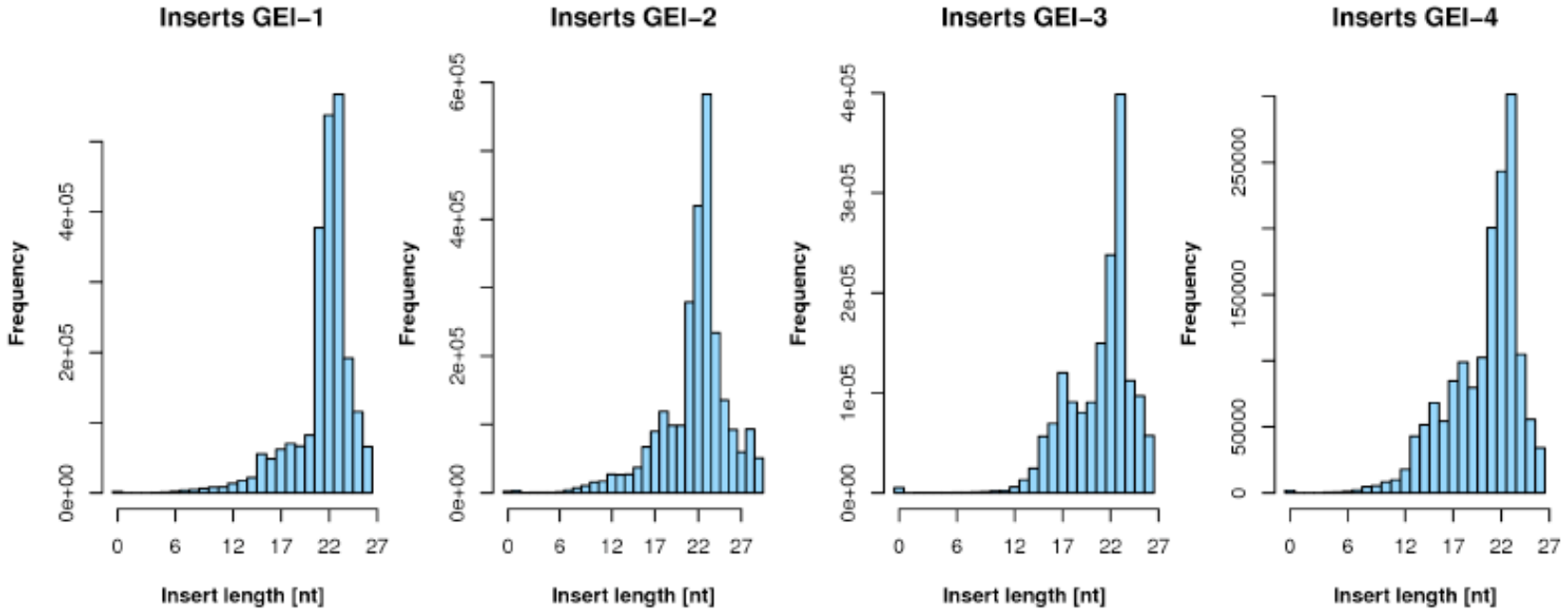
(J. Toedling)

Small RNAs analysis – workflow

1. Adaptor removal



Size distribution of reads



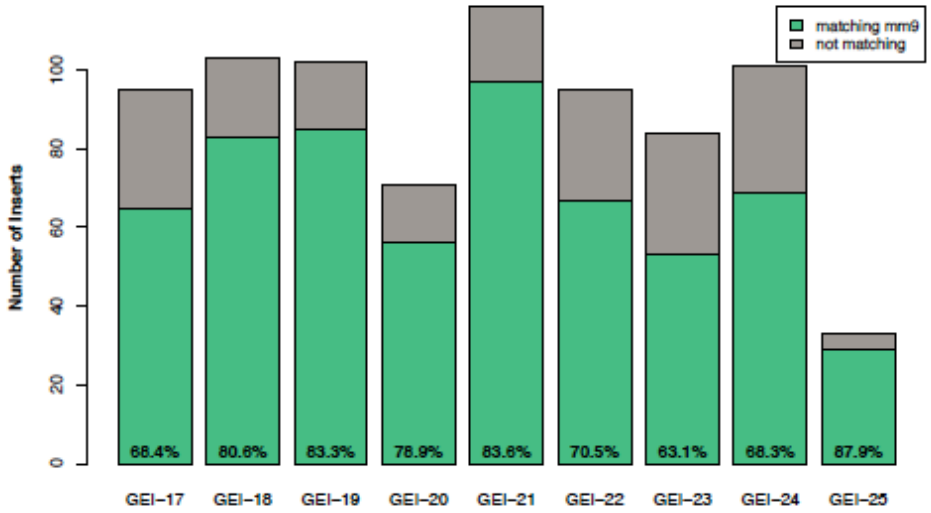
(J. Toedling)

Small RNAs analysis – workflow

2. Mapping of reads
   - **map reads to the genome and**
   - **determine type of read by (lack of) overlap with annotated genome elements**

Several tools available :
- **Mapreads - RNA2MAP (ABI)**
- **MAQ** - BWA
- SOCS
- **SHRIMP**
- BFAST
- PASS
- ZOOM
- **BOWTIE**

% of mapping per libraries

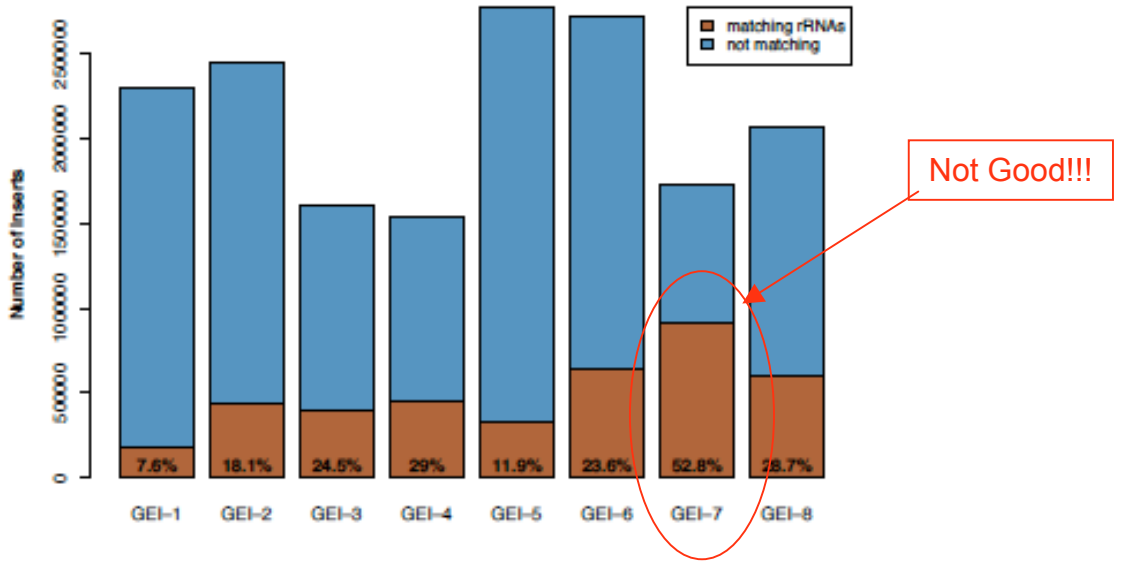


(J. Toedling)

Small RNAs analysis – workflow

3. Annotation of : (database used miRBase, MGI, RFAM)
   - rRNAs
   - microRNAs
   - tRNAs
   - Repeats
   - Genome siRNAs
   - Unknown sequences

% of rRNAs per libraries



(J. Toedling)

Small RNAs analysis – workflow

3. Annotation of : (database used miRBase, MGI, RFAM)
   - rRNAs
   - microRNAs

Ex:

- Length between 19 and 26.
- One unique match position (best)
- For which the match position overlap with the mature/pre-mir sequence (miRBase.13.0) on the same strand.

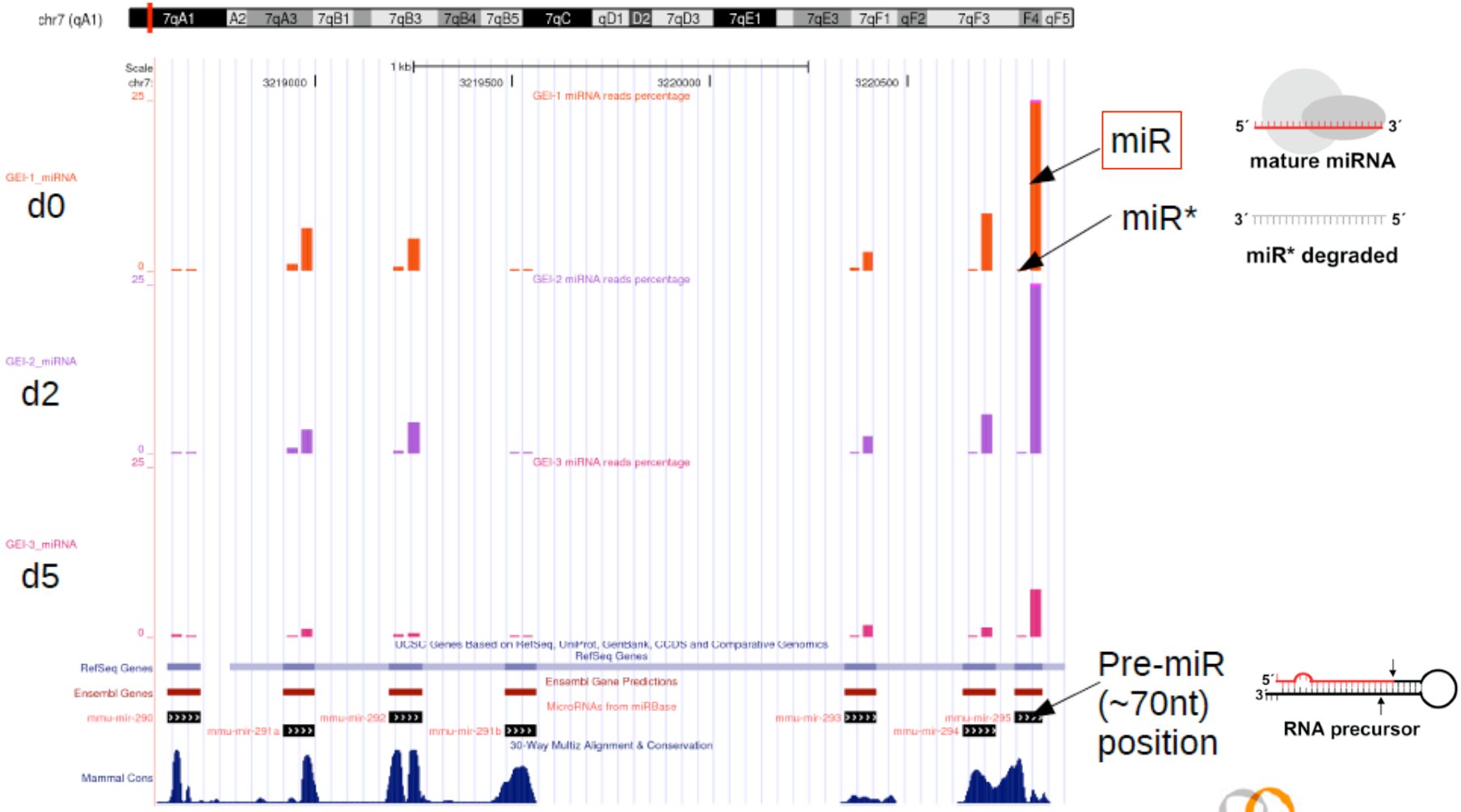
	R5	R6	R7	R7bis
All reads	25274245	21040771	31704644	22220727
18.2.6	7375333	4059383	4275178	3191503
	29.18%	19.29%	13.48%	14.36%
Best/Unique [19-26]	3069965	1491875	1516692	1071622
	41.62%	36.75%	35.47%	33.57%
Pre-miR	1884599	893935	799926	578914
	61.38%	59.92%	52.74%	54.02%
Mat-miR	1854796	880647	788194	570249
	60.41%	59.02%	51.96%	53.21%

Table 2: Reads annotation against miRBase13.0. Percentage of Best/Unique according to the total number of reads. Percentage of miRNAs according to the number of Best/Unique.

(N. Servant)

Small RNAs analysis – workflow

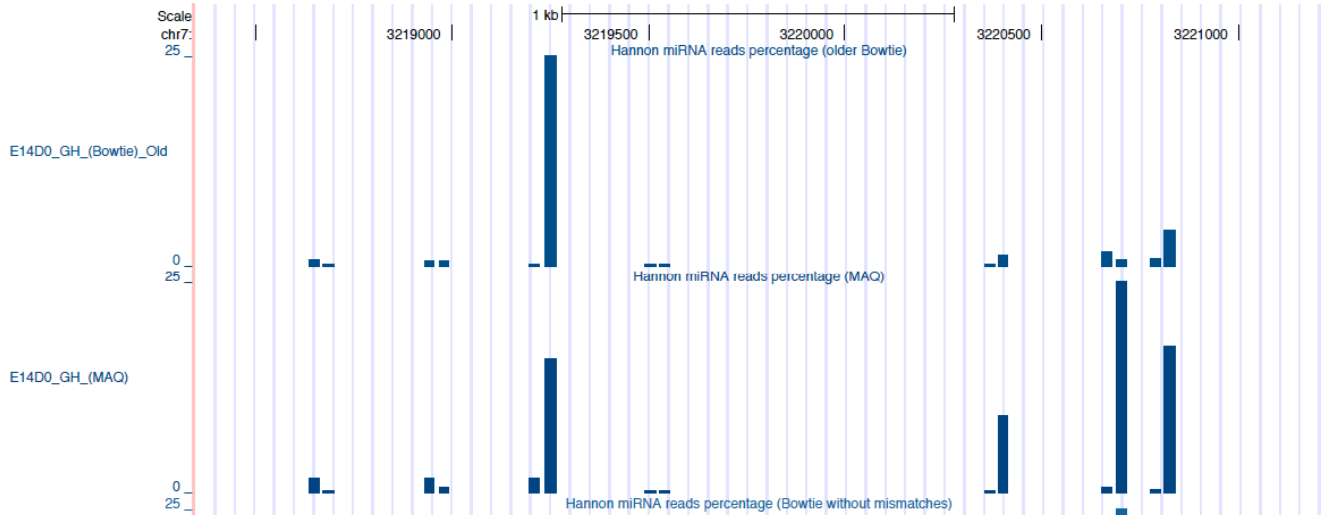
4. Visualisation of microRNA profiles on UCSC

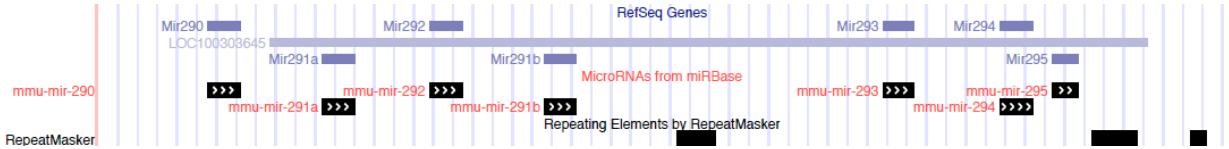


visualised using UCSC genome browser

(Ciaudo et al., 2009)

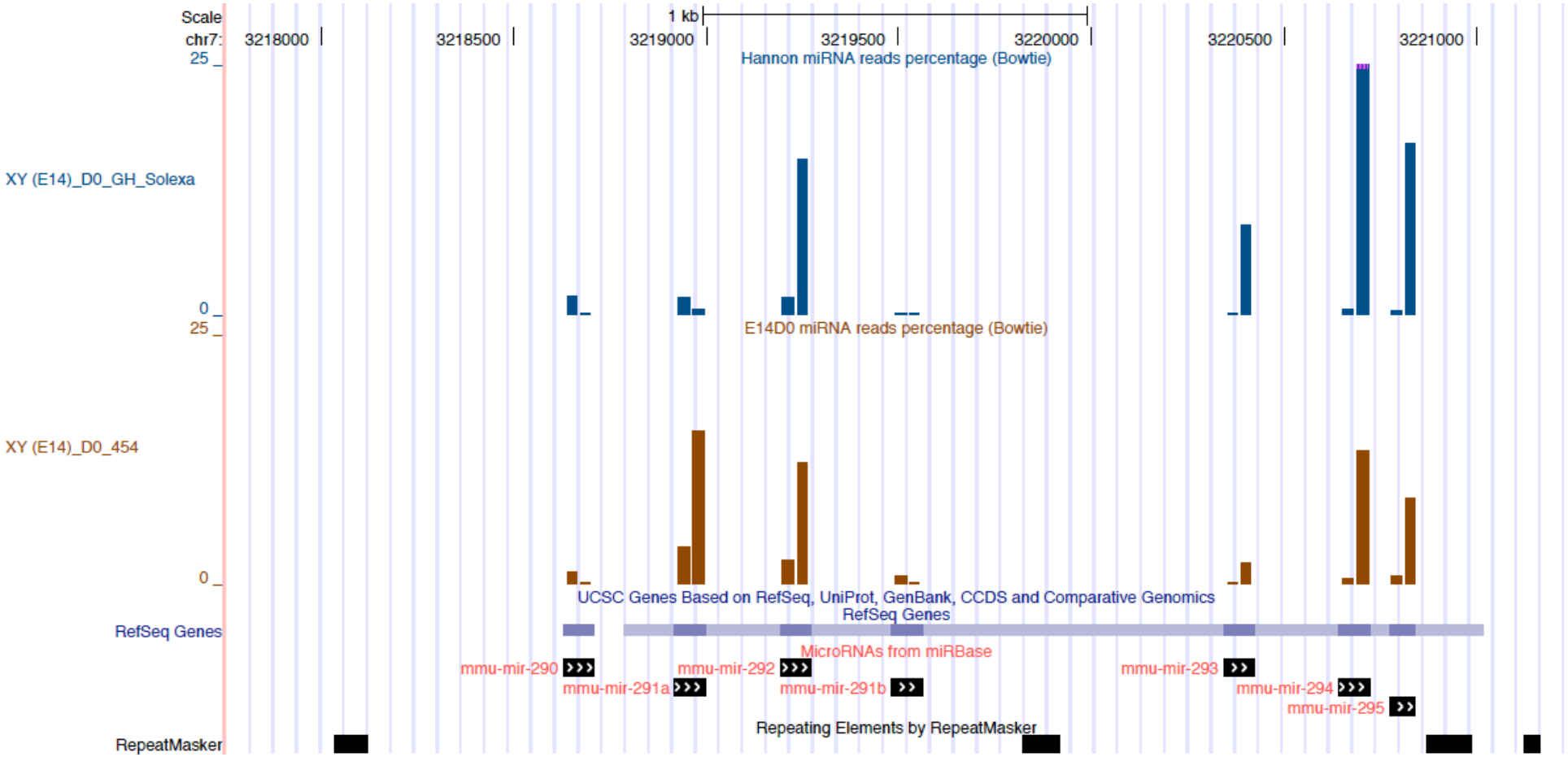
Comparison of tools





Parameters used for each tool have a big impact on results!!!

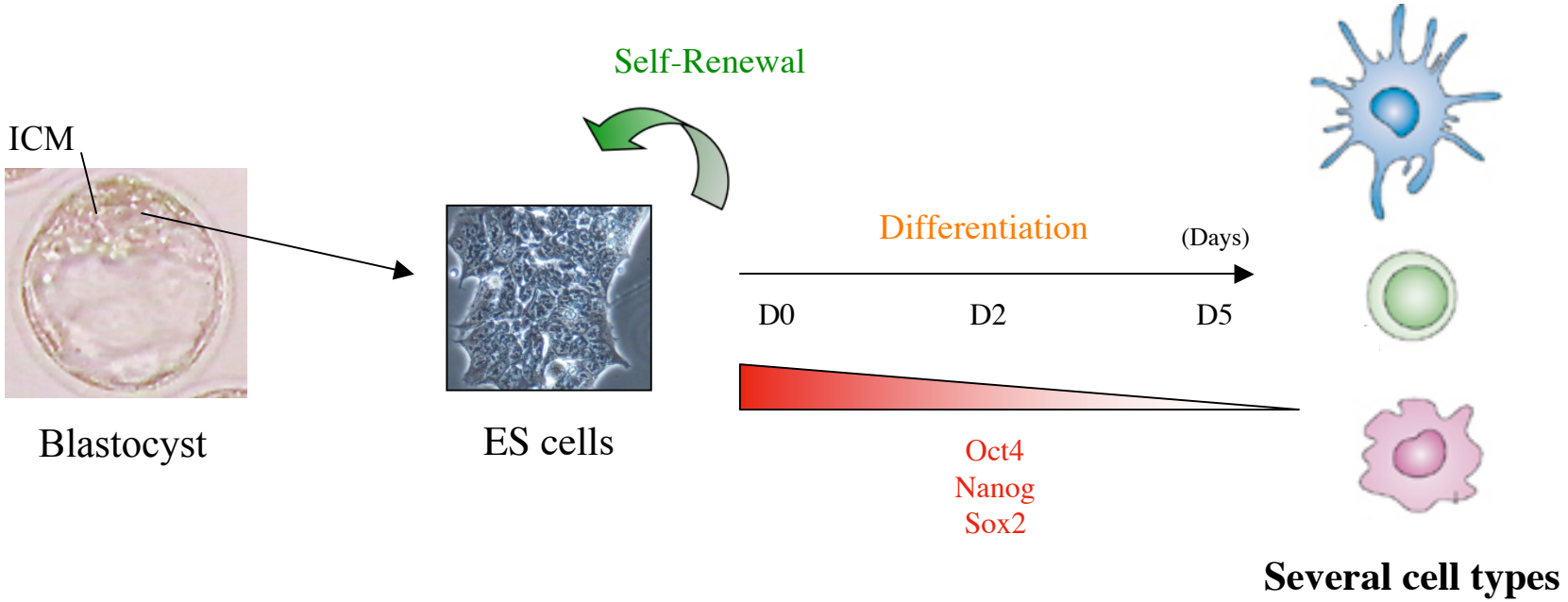
Comparison of sequencing machines



Some differences are observed (library preparation?)

3. Profiling of microRNAs during ES cells differentiation

Dynamics of small RNAs during ES cell differentiation



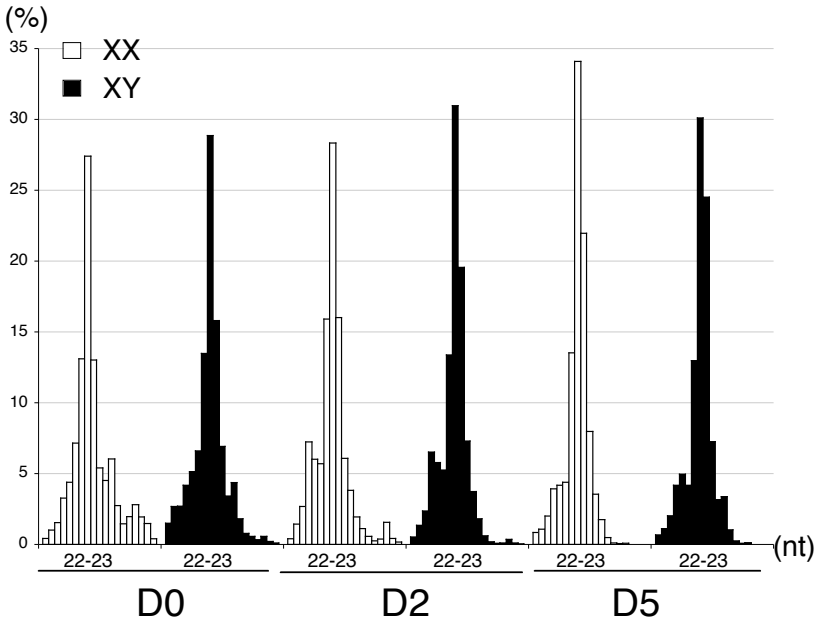
Cell lines used: XX ES cells / XY ES cells

- Cells growing without feeders layer have been used
- Differentiation induced by LIF withdrawal (no RA)

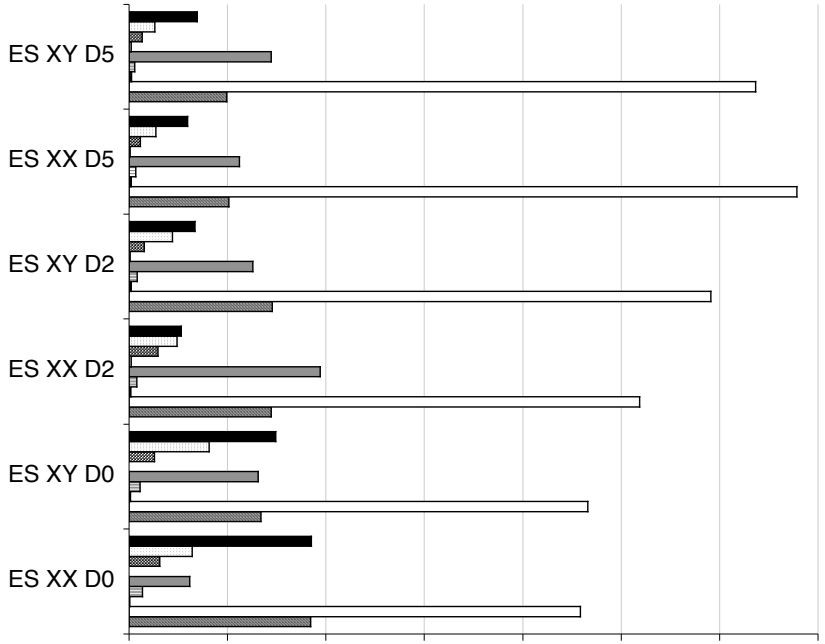
(Ciaudo et al., 2009)

Dynamics of small RNAs during ES cell differentiation

Size distribution



Annotation of small RNAs

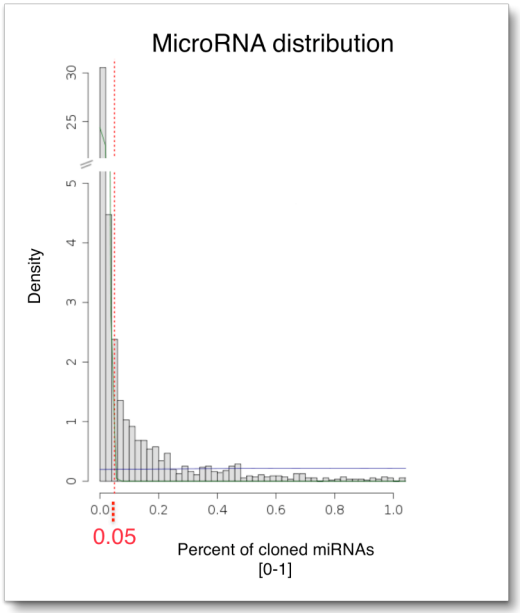


Genome microRNA Mitochondria Repeat rRNA scRNA sn-snoRNA tRNA Not annotated

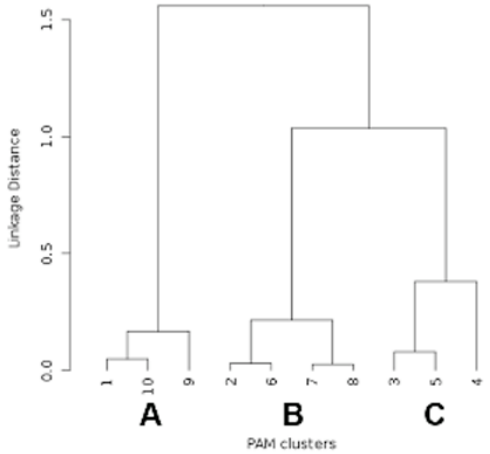
50% of small RNAs are microRNAs in ES cells

(Ciaudo et al., 2009)

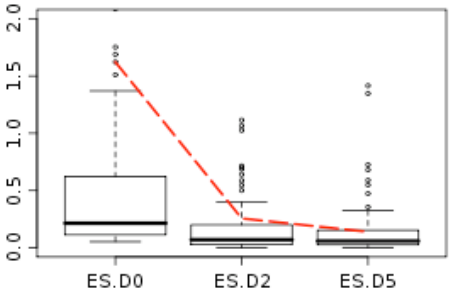
Statistical Analysis of microRNA expression



Hierarchical Clustering of PAM Classes

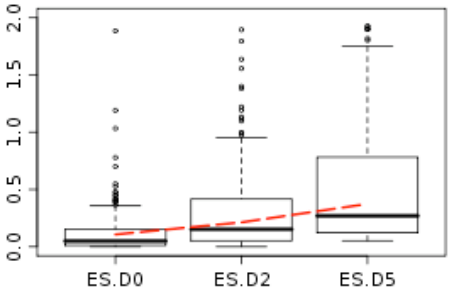


Class A



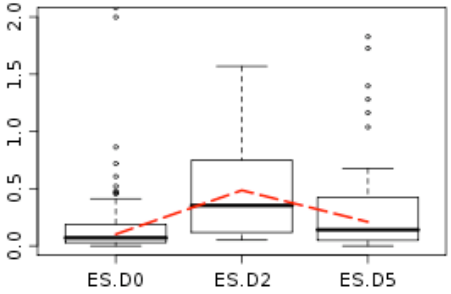
Decrease during differentiation

Class B



Increase during differentiation

Class C

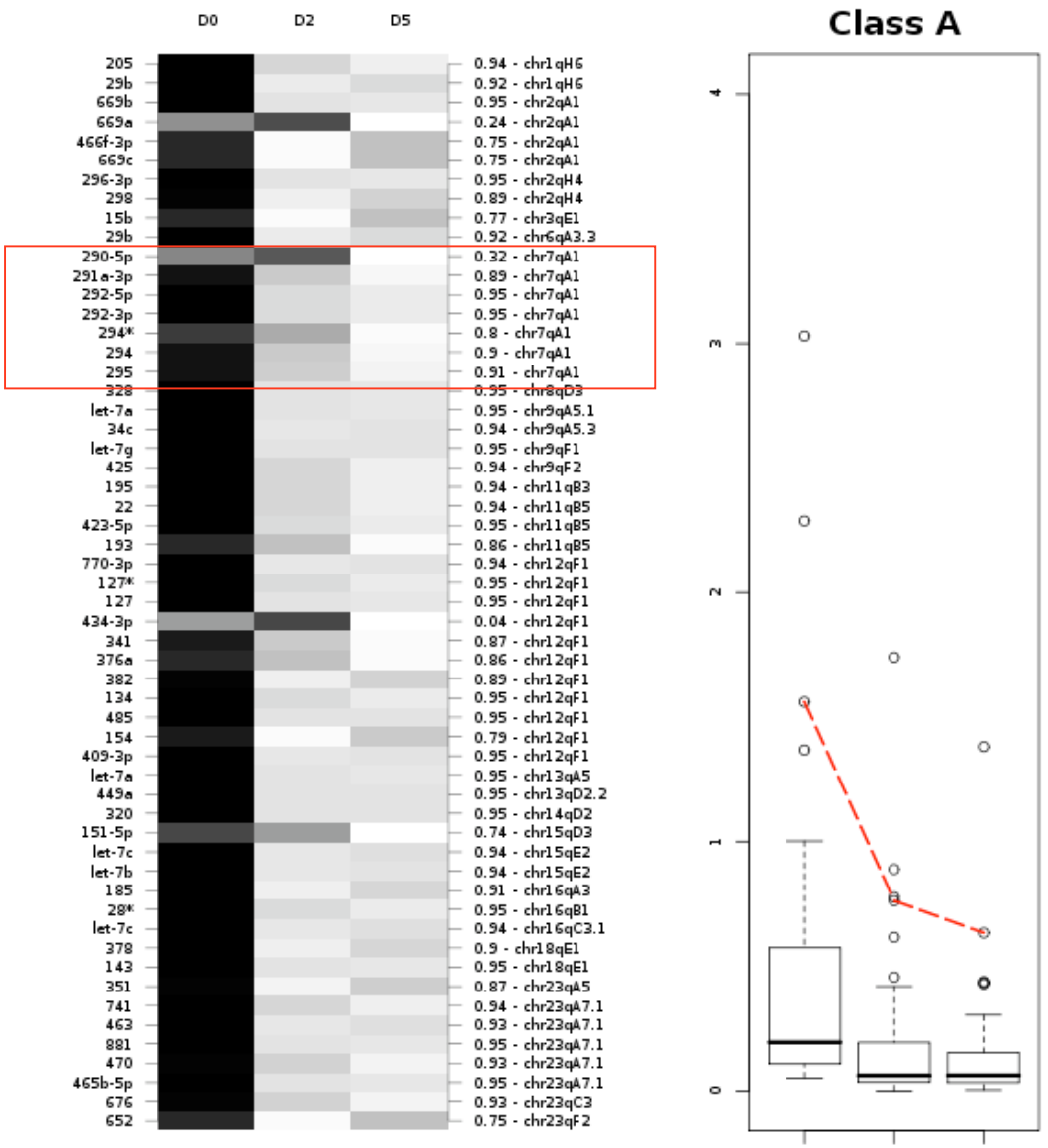


Transient increase during differentiation

Three expression classes of microRNAs have been defined

(Ciaudo et al., 2009)

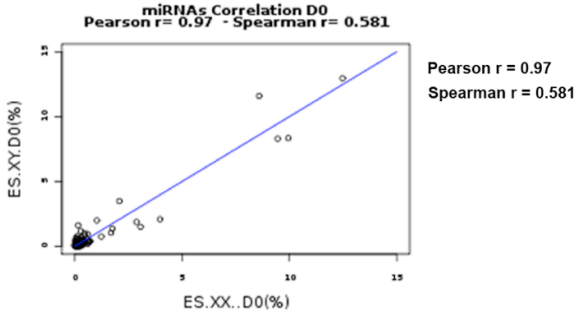
Hierarchical Clustering of PAM Classes



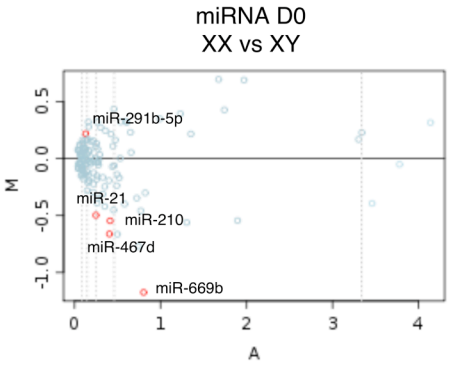
In most of cases, miRNAs known to be genomically clustered were found to be grouped together within the same PAM classes

(Ciaudo et al., 2009)

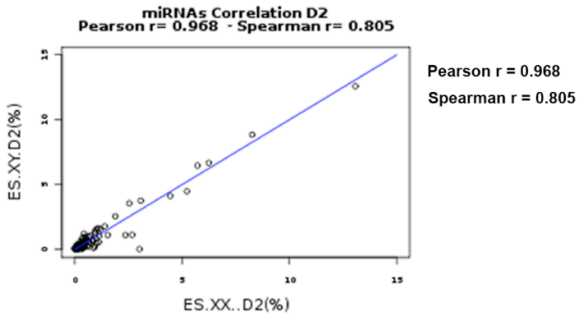
Sex-specific microRNAs



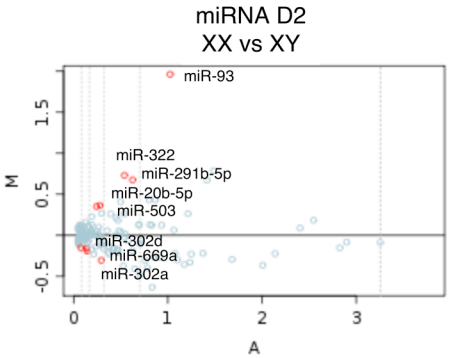
Pearson r = 0.97
Spearman r = 0.581



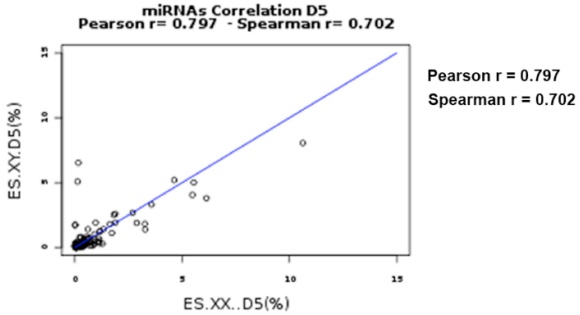
D0	Enriched in	pvalue
miR-21	XY	0.003
miR-669b	XY	0.004
miR-467d	XY	0.01
miR-291b-5p	XX	0.028
miR-210	XY	0.039



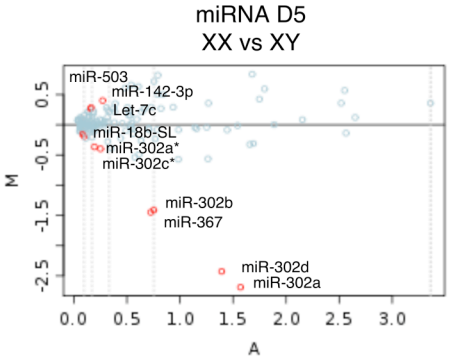
Pearson r = 0.968
Spearman r = 0.805



D2	Enriched in	pvalue
miR-93	XX	0
miR-322	XX	0.004
miR-291b-5p	XX	0.009
miR-20b-5p	XX	0.01
miR-503	XX	0.014
miR-302a	XY	0.014
miR-669a	XY	0.036
miR-409-5p	XY	0.038
miR-467c	XY	0.044
miR-302d	XY	0.05



Pearson r = 0.797
Spearman r = 0.702

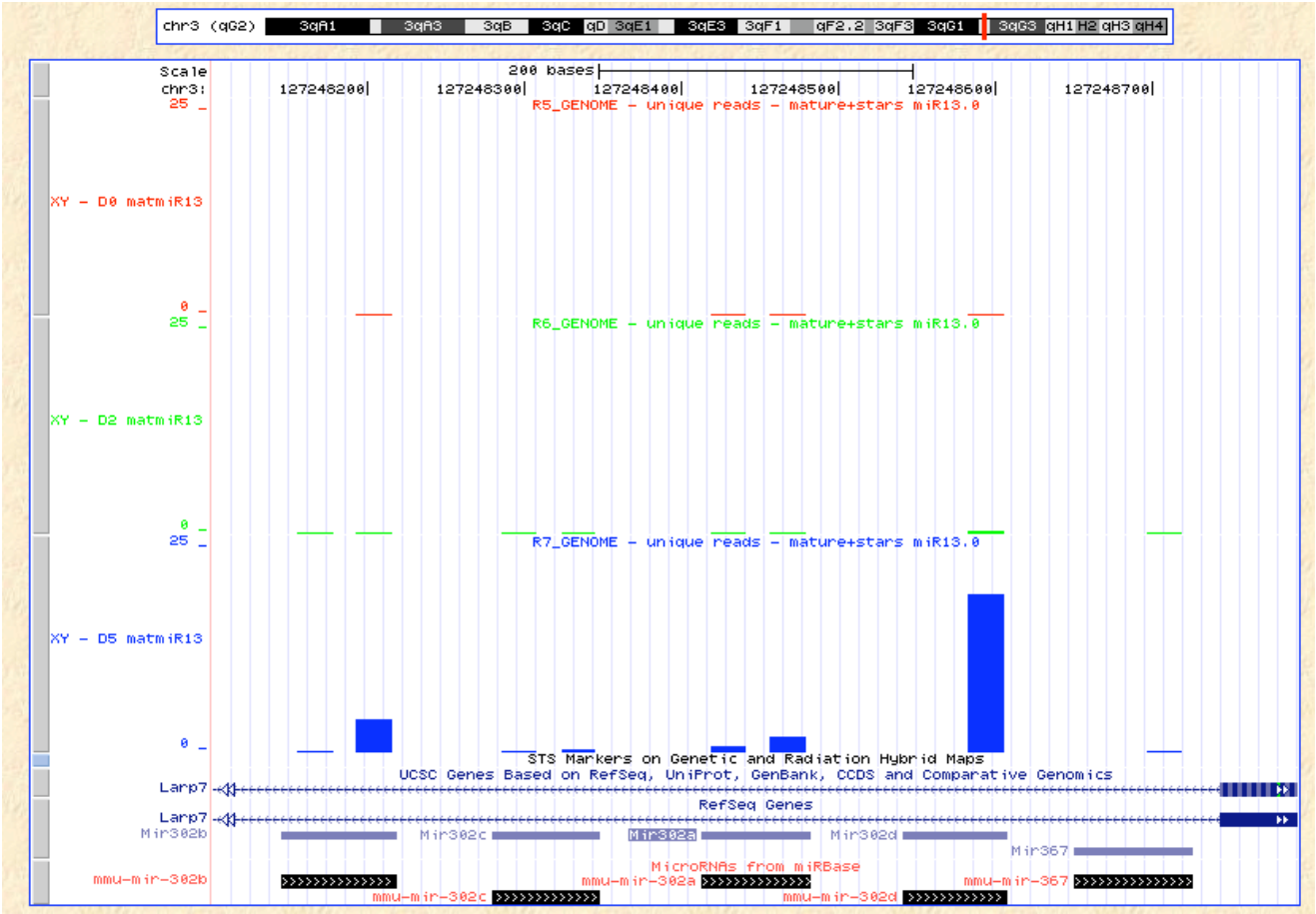


D5	Enriched in	pvalue
miR-302a	XY	0
miR-302d	XY	0
miR-302b	XY	0
miR-367	XY	0.001
miR-302c	XY	0.009
mir-18b-SL	XY	0.01
miR-503	XX	0.011
let-7c(0)	XX	0.014
miR-302c*	XY	0.018
miR-142-3p	XX	0.023
miR-302a*	XY	0.046

Some microRNAs are differentially regulated between males and females during the differentiation process

(Ciaudo et al., 2009)

The miR-302 cluster



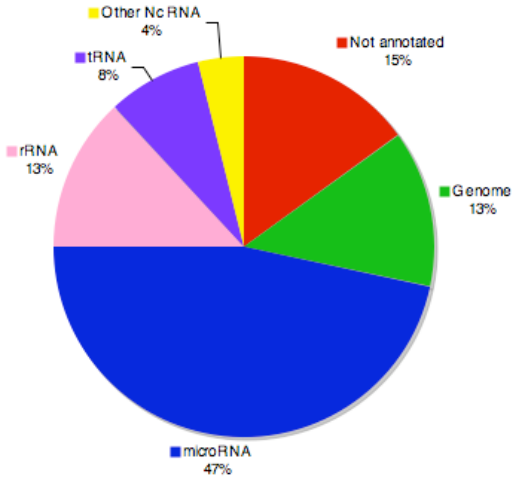
In Vivo validation of miR-302 cluster over-expression in male germline

Looking for targets (deletion of the cluster and over-expression followed by microarray expression experiment)

(Ciaudo et al., 2009)

4. Small RNAs involved in X inactivation process?

Analysis of other types of small RNAs



Collaboration
Dr Jennifer CHOW (Dr Edith Heard)

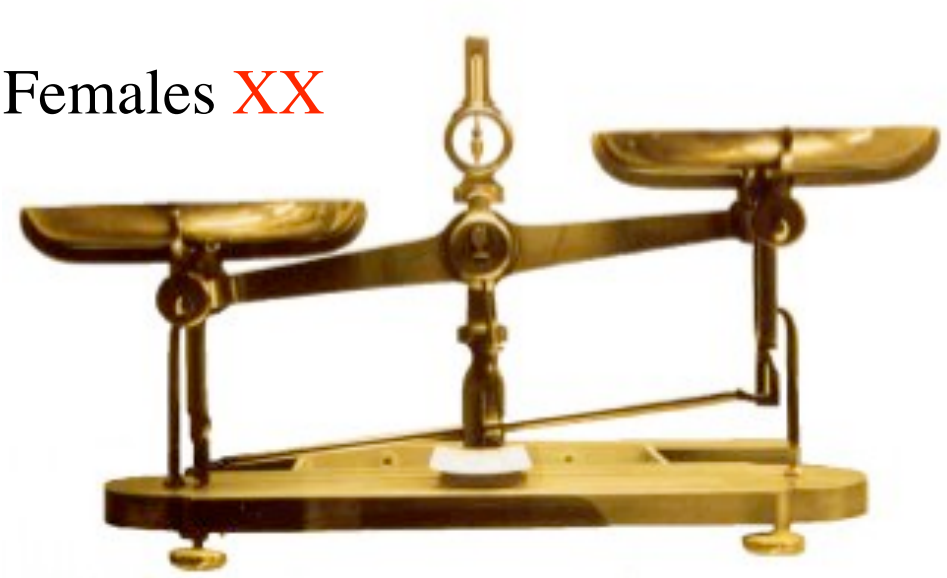
institutCurie
Ensemble, prenons le cancer de vitesse.

Small non-coding RNAs could be involved in X inactivation process?

Heterogametic XY Species

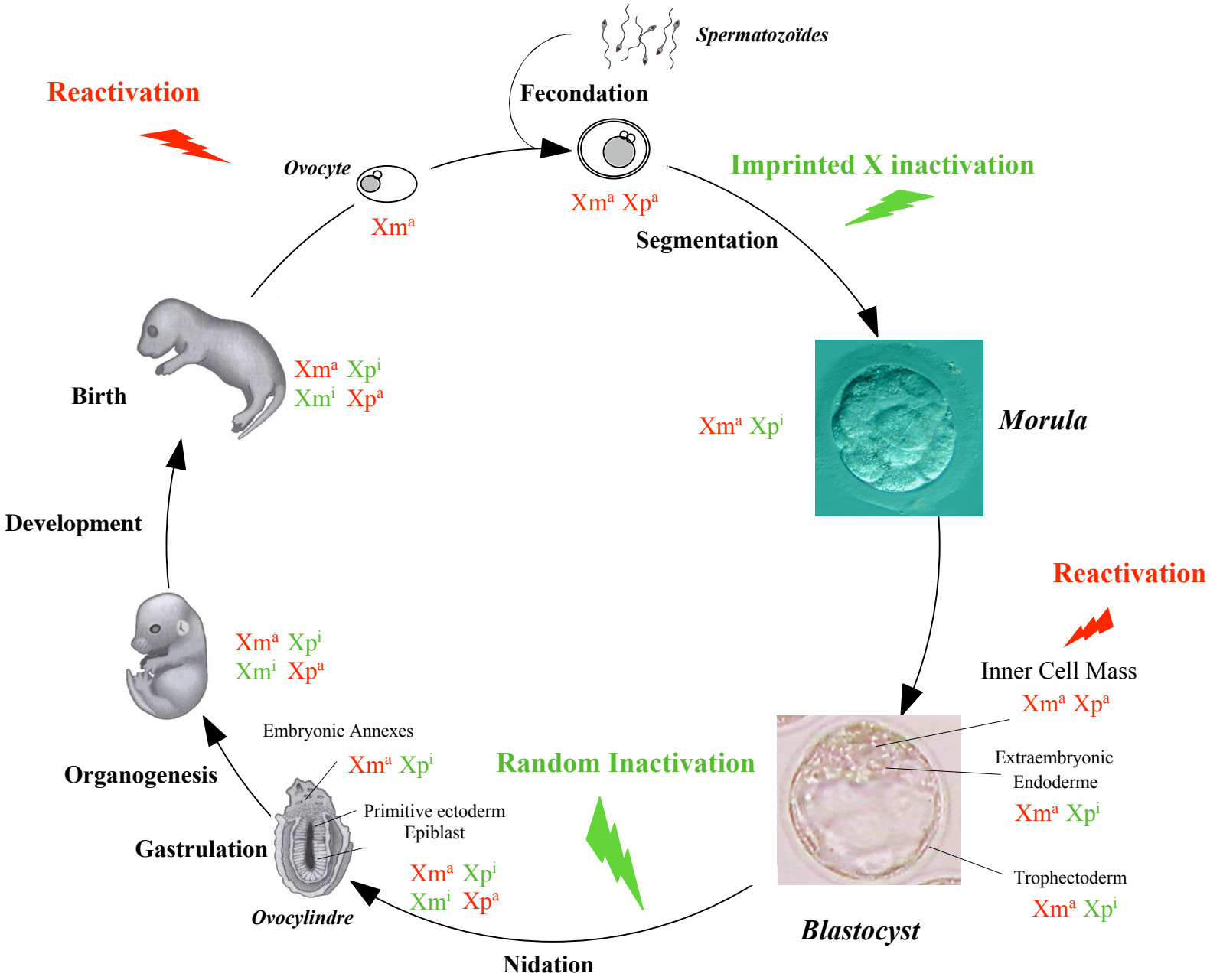
Males XY or XO

Females XX

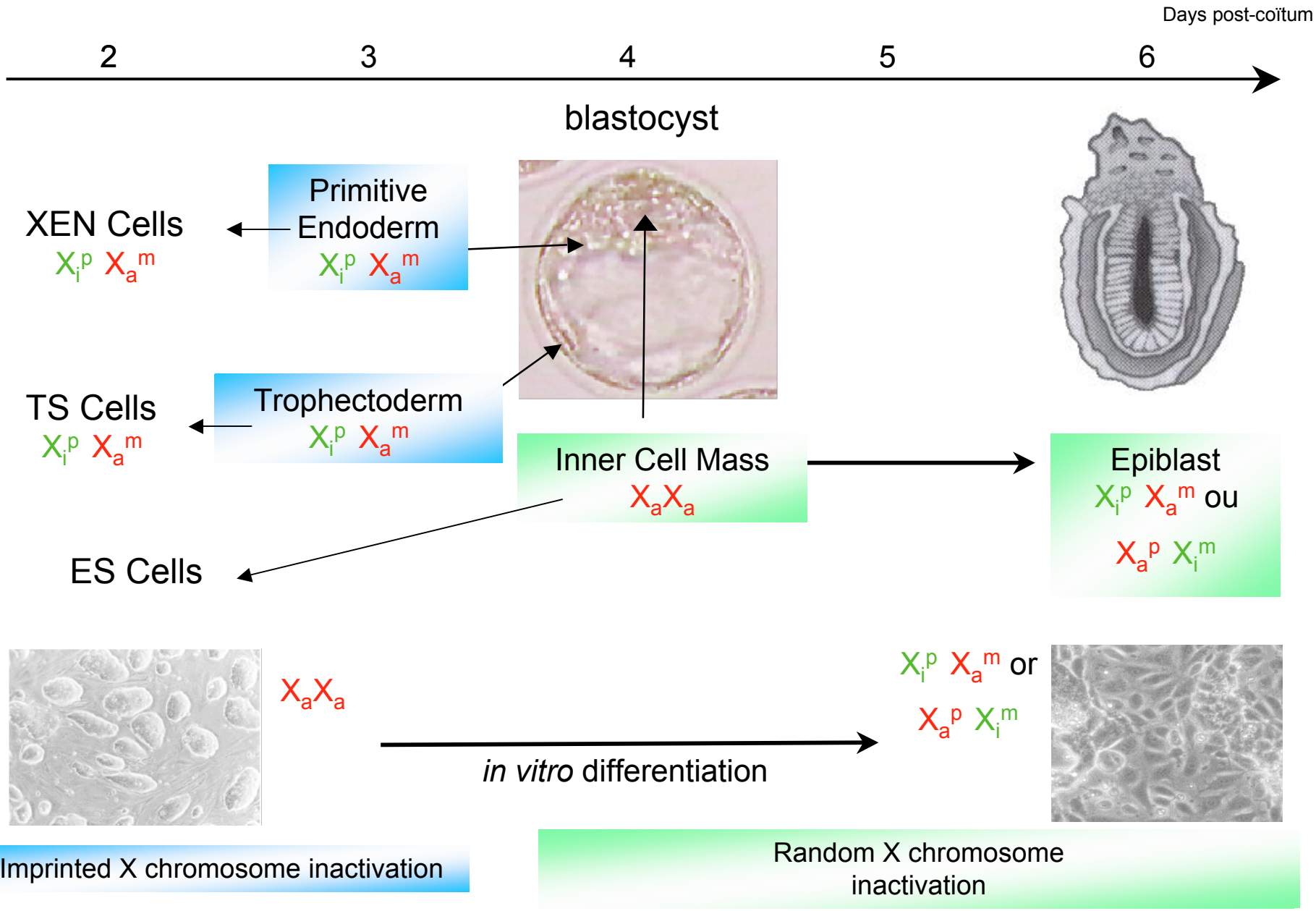


How to equalize X-linked gene expression between sexes?

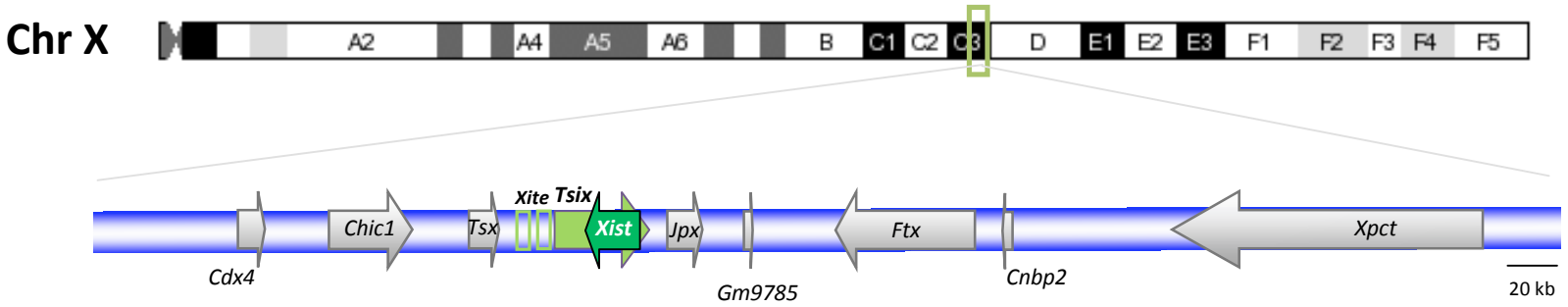
X inactivation during female mice life

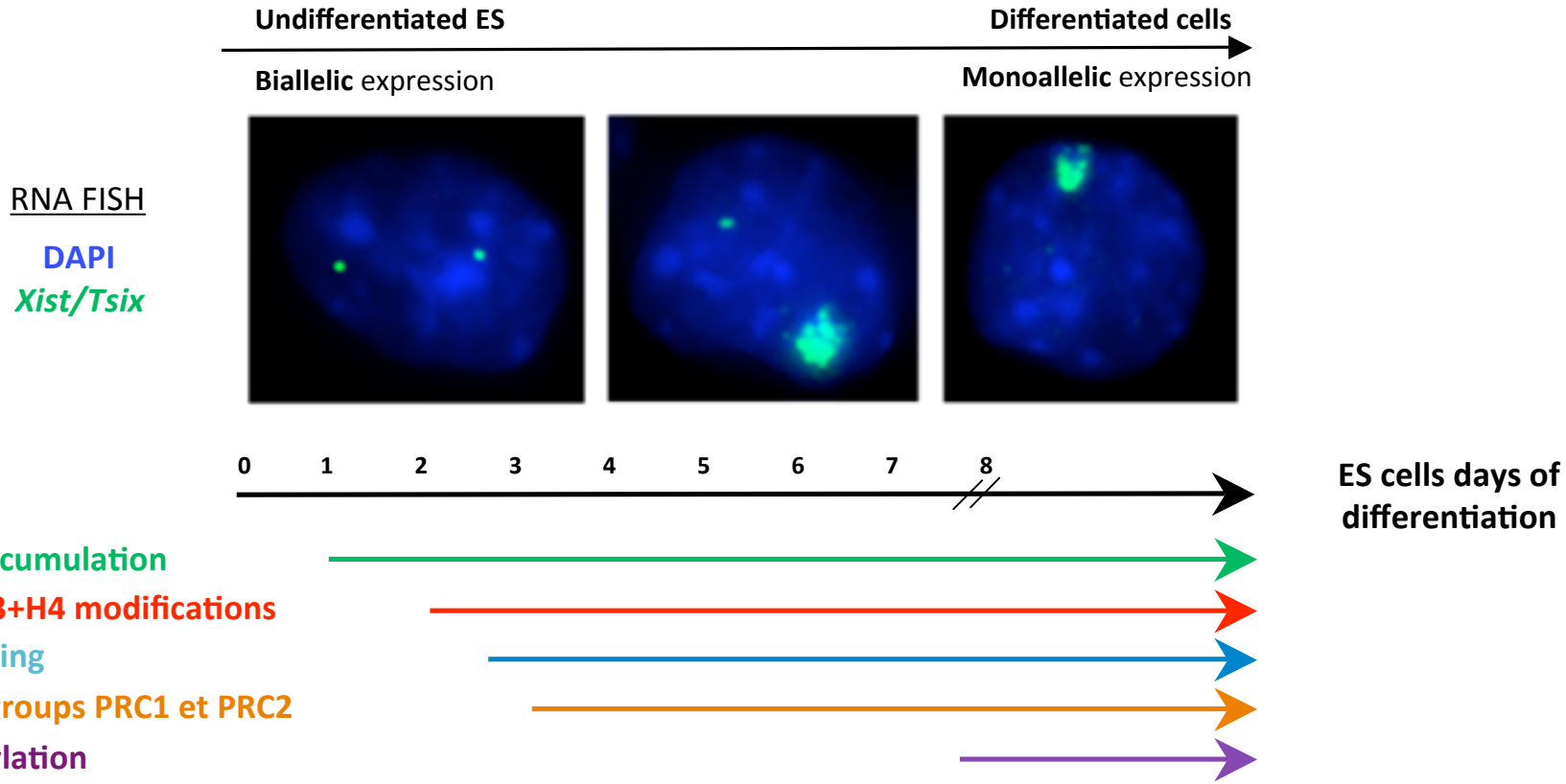


Dynamics of XCI during early development

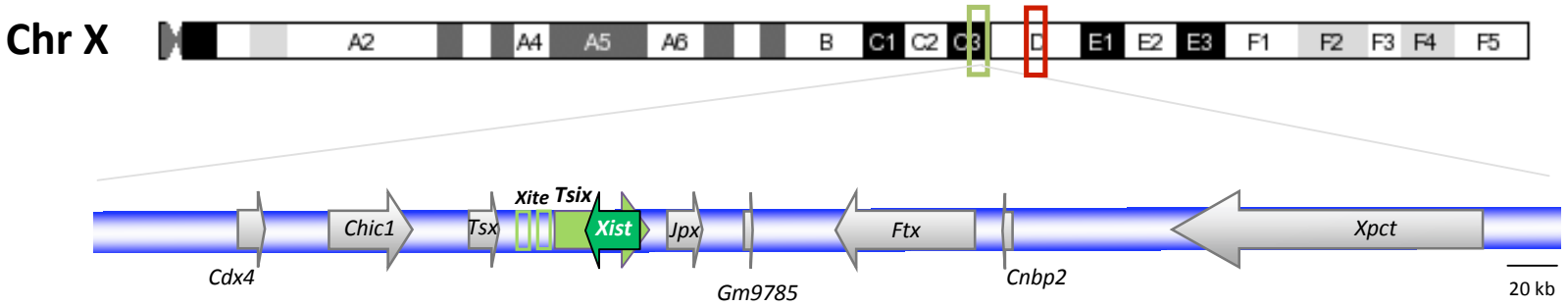


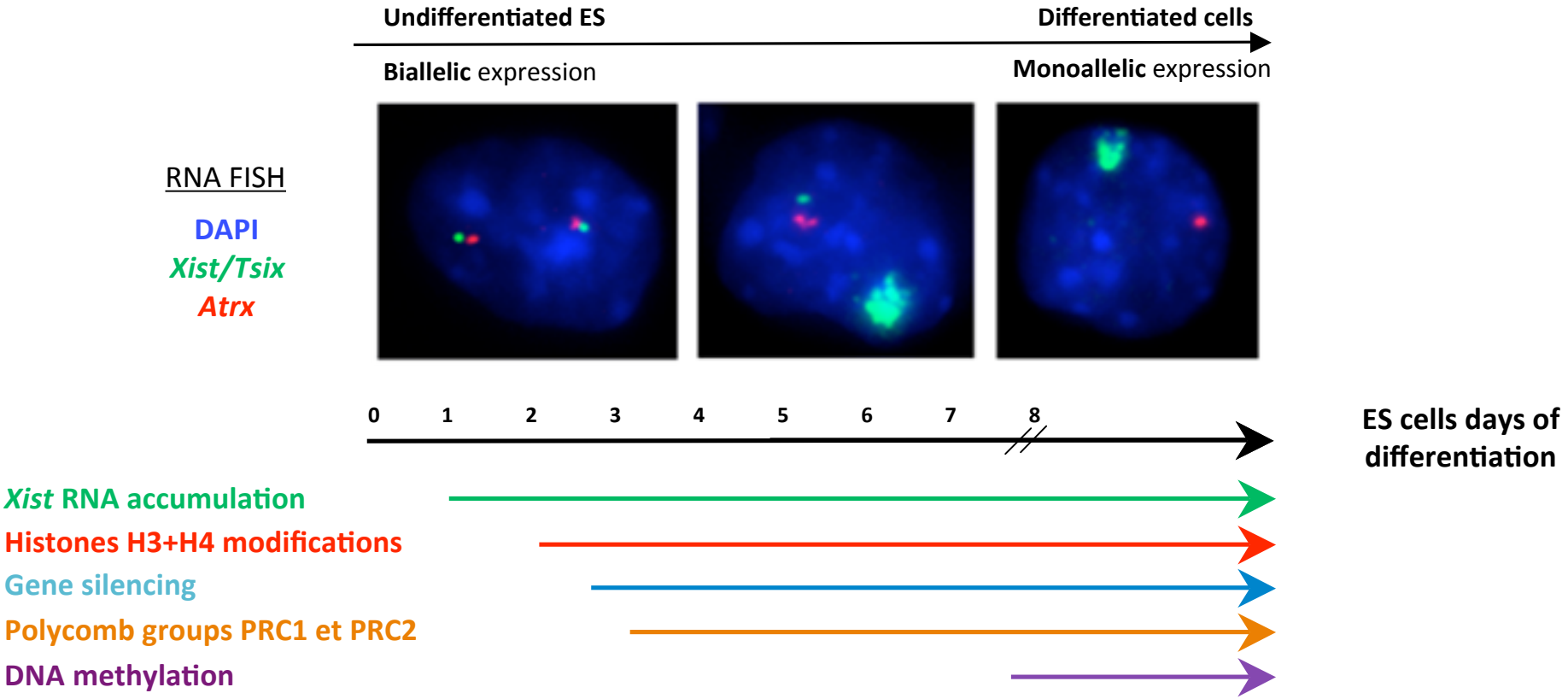
The X inactivation centre (*Xic*)





The X inactivation centre (*Xic*)





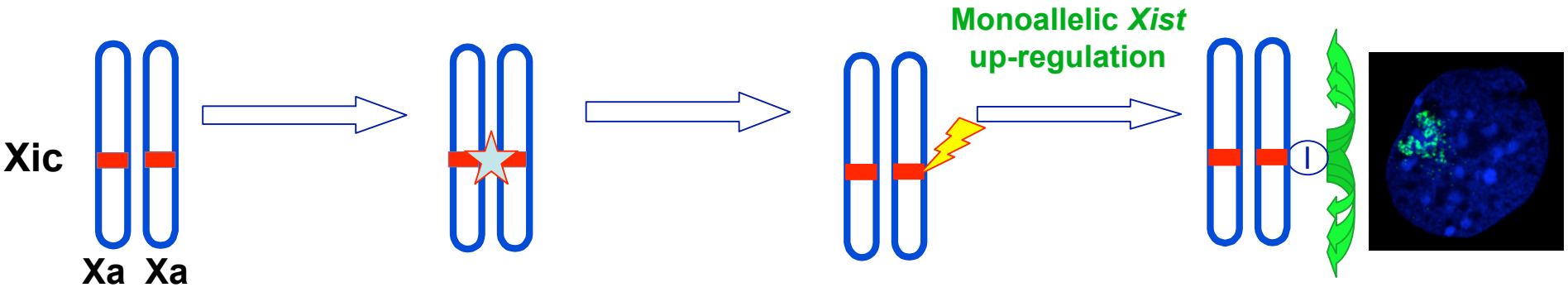
X inactivation :
Multiple steps process

INITIATION

Sensing, counting, choosing…

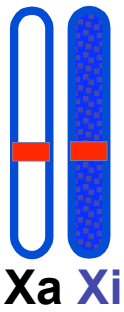
SPREADING

Xist RNA "coats" in cis, silences genes



MAINTENANCE

- histone modifications
- Polycomb group proteins
- late replication
- histone variants
- DNA methylation

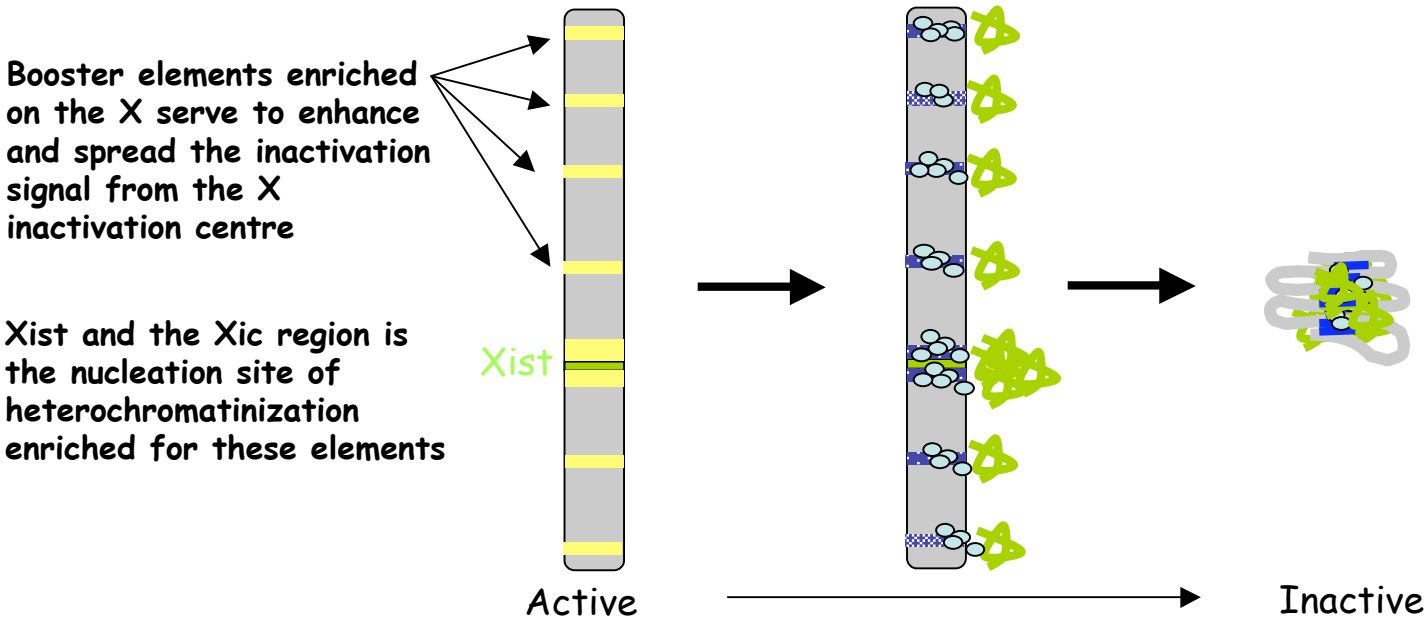


Spread of Heterochromatin?

Booster / relay elements?

The LINE hypothesis (Lyon, 1998)

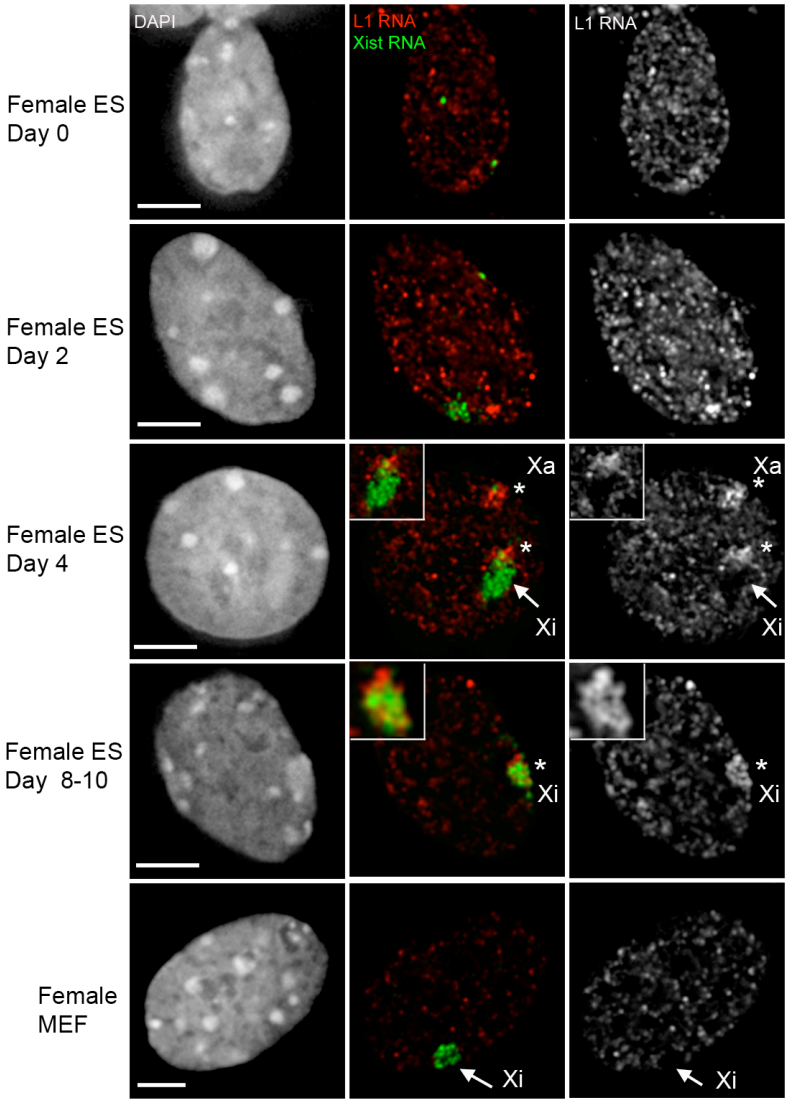
Mary Lyon proposed LINE retro-elements as possible candidates for "booster elements" that spread inactivation across the X chromosome :



(Gartler and Riggs 1983)

What types of repeats are enriched on X chromosome during establishment of X inactivation ?

LINE-1 expression patterns in differentiating female ES cells

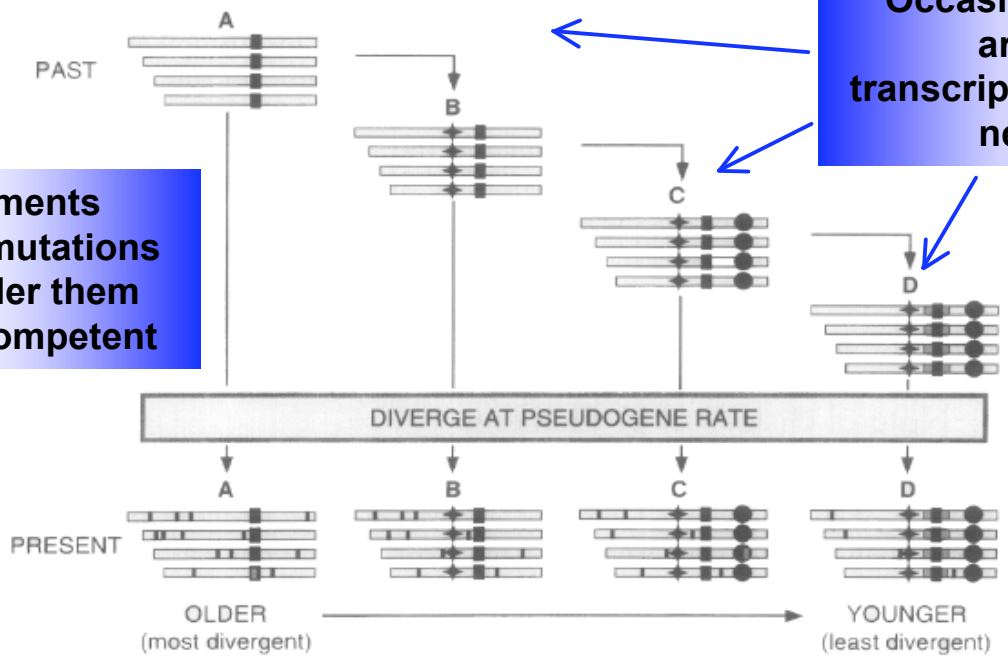


L1 elements (LINEs)

- **At day 4, an accumulation of LINE transcripts appears on both the inactive and active X chromosomes**
- **On the inactive X the LINE 1 transcripts appear adjacent to the Xist RNA domain**
- **At later differentiation stages (day 8-10) the LINE1 transcripts become intermingled with the Xist RNA domain and are only found on the inactive X**

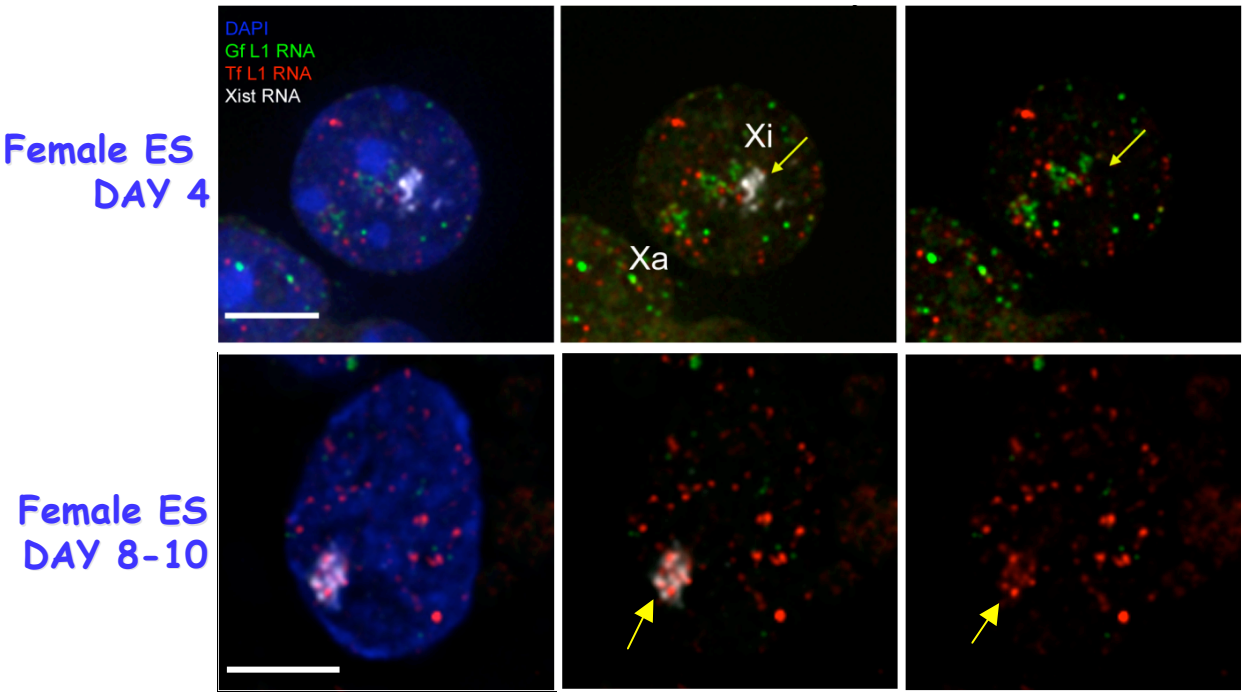
(Chow et al., in revision)

Mouse LINE-1 subfamilies



- **Older L1 elements will have had more time to accumulate mutations and members will be more divergent and largely transcriptionally incompetent.**
- **Younger subfamilies will be the least divergent and will contain the most transcriptionally competent elements.**
- **In mouse there are 3 transcriptionally active "young" L1 subfamilies:**
  - **Tf = most active with 1800 active full length elements**
  - **Gf = 400 active elements**
  - **A = 900 active elements**
  - **The young, transcriptionally active LINEs are the subfamilies that show enrichment on the Xi**

L1 accumulations can be detected with probes specific to the promoter regions of young, transcriptionally active L1



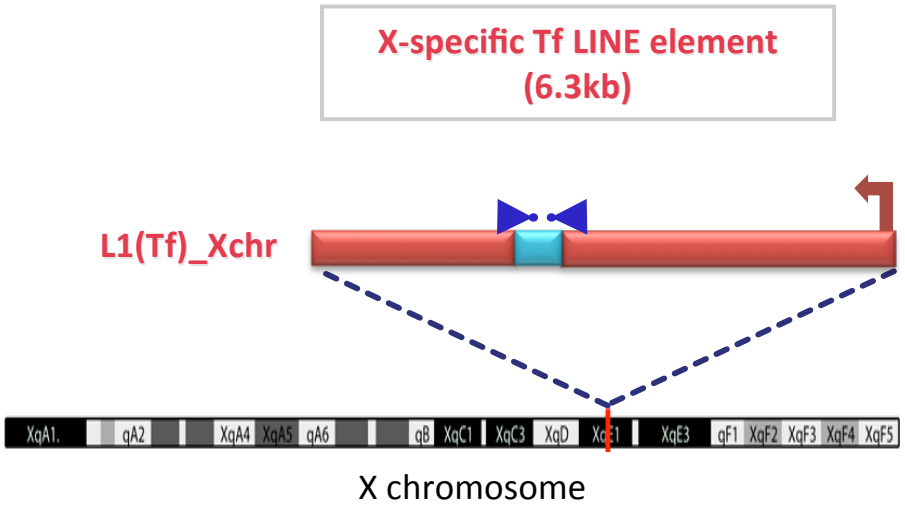
Both Gf and Tf elements are expressed and form an accumulation at the Xi and Xa at day 4

At day 10, mainly Tf transcripts show accumulation at the Xi

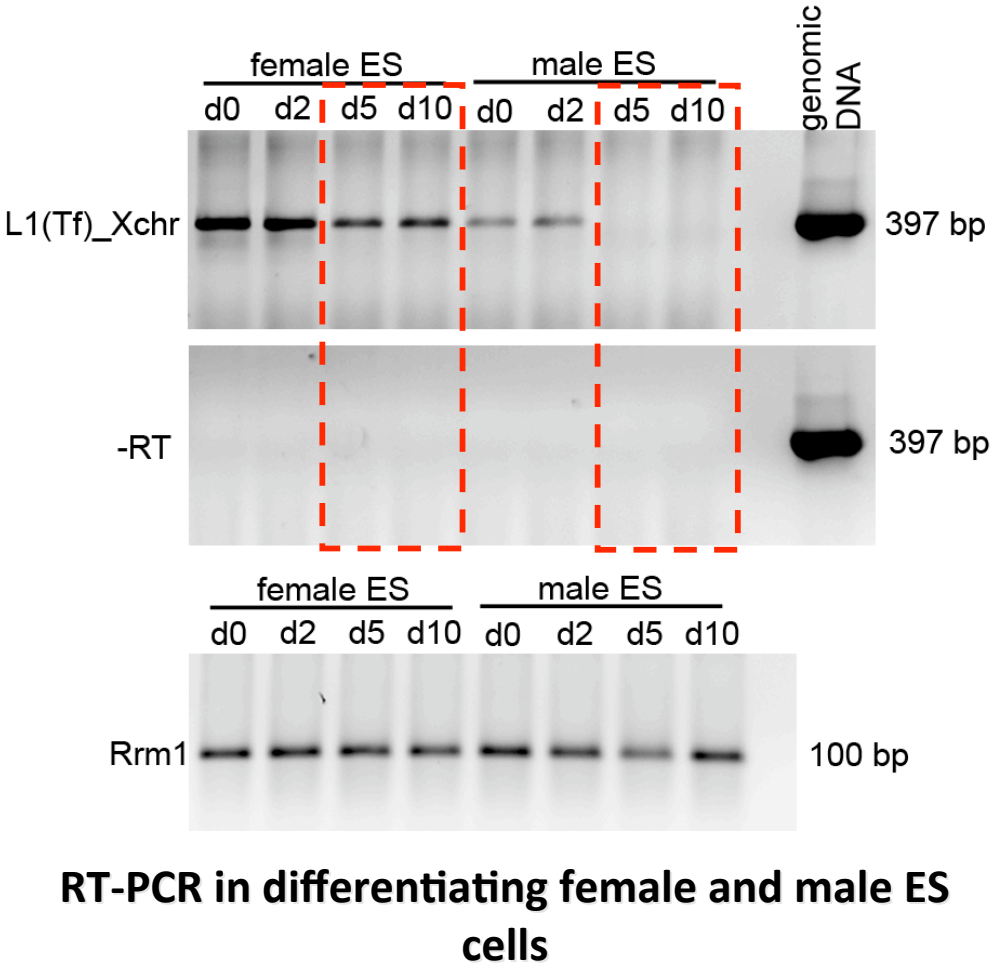
L1 accumulations likely derive from young, transcriptionally active elements
•DAY 4 – Gf and Tf element families
•DAY 8-10 – mainly Tf elements

(Chow et al., in revision)

Expression of an X-specific LINE element persists only in females



The expression of a specific X-linked LINE element persists until later stages of differentiation in female but not male ES cells

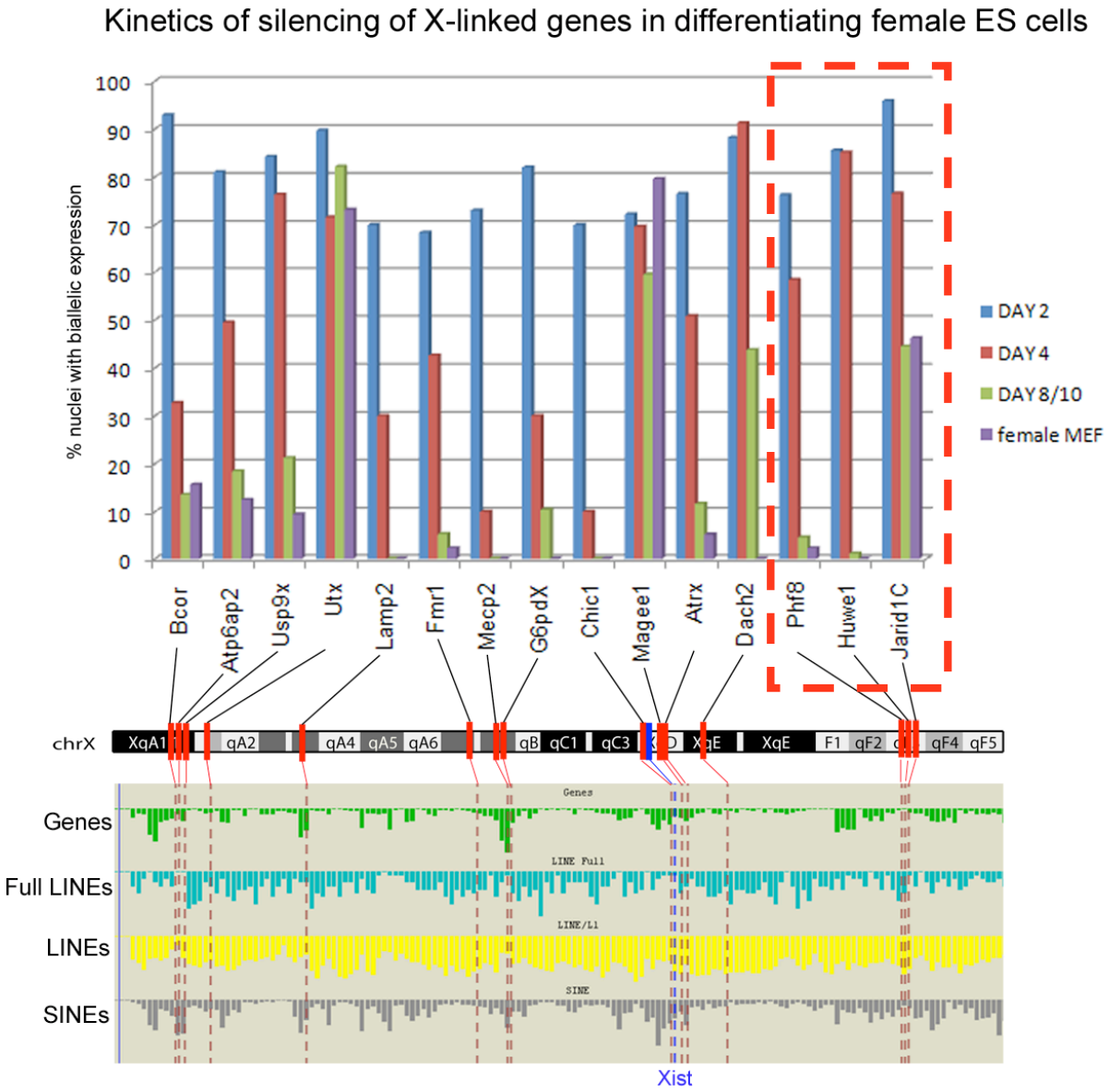


RT-PCR in differentiating female and male ES cells

Differential expression of LINE1 elements depending on whether they are on the active or the inactive X chromosome.

(Chow et al., in revision)

Does the presence/expression of specific X-linked LINEs have an effect on nearby gene expression kinetics?

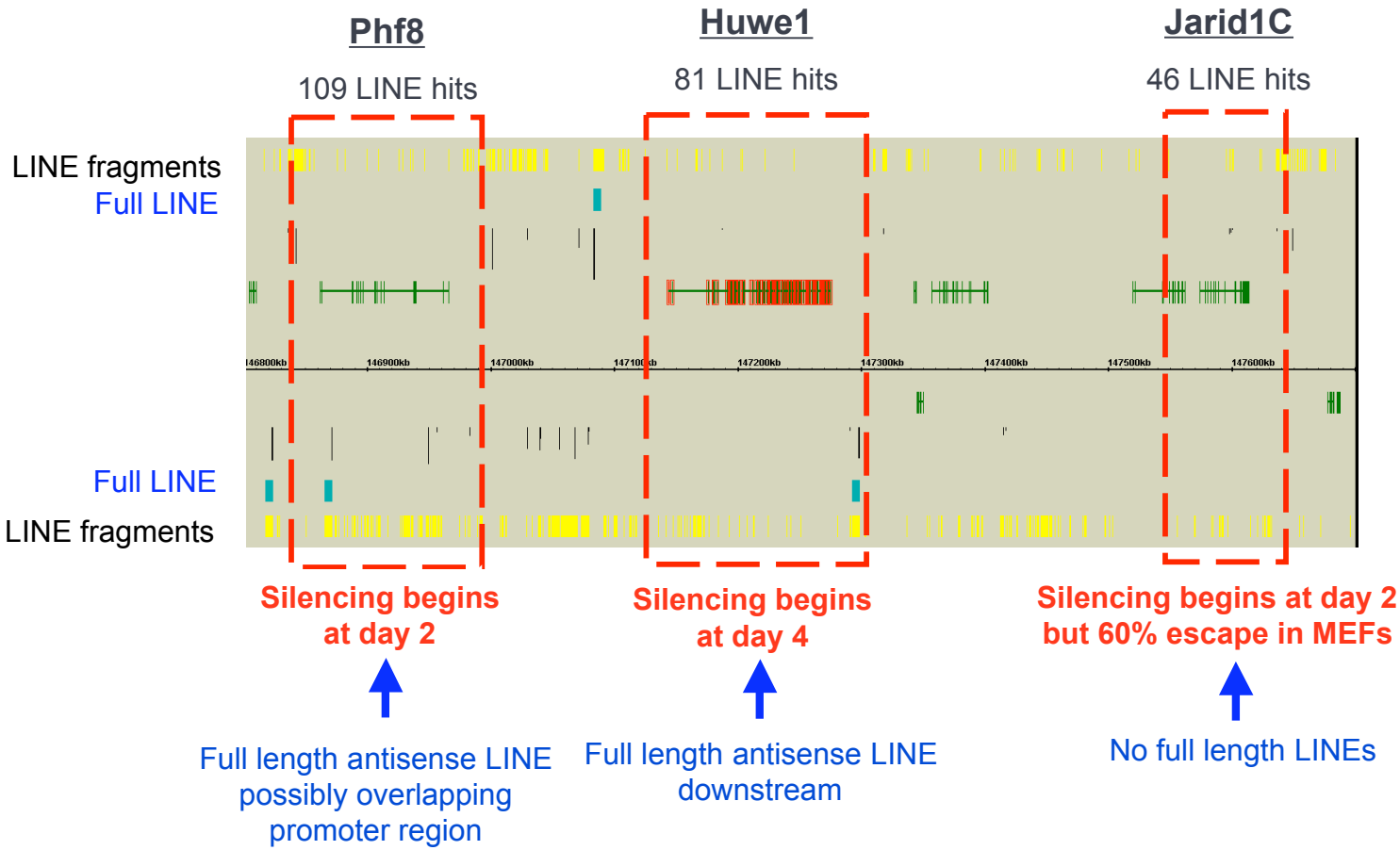


Great deal of variability in kinetics of silencing during XCI

Does XCI efficiency correlate with LINEs or other repeats ?

(Chow et al., in revision)

Does the presence/expression of specific X-linked LINEs have an effect on nearby gene expression kinetics?



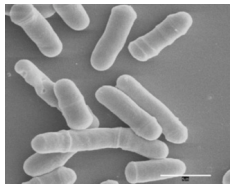
Gene silencing efficiency correlates with LINE density
And even more so with the presence of full length LINEs

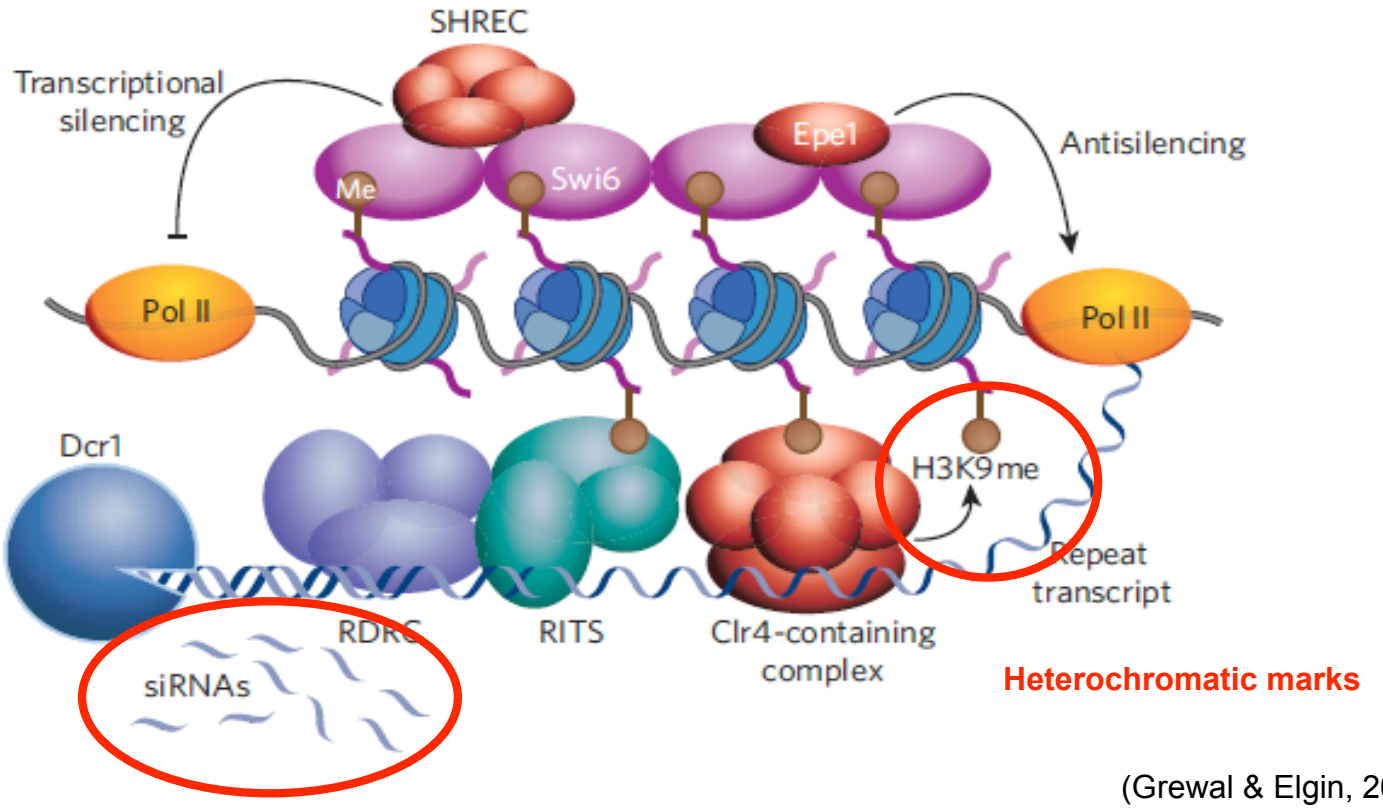
(Chow et al., in revision)

How might the expression of young full length L1 elements help in silencing?

siRNAs have been involved in heterochromatin formation and maintenance

S. Pombe





(Grewal & Elgin, 2007)

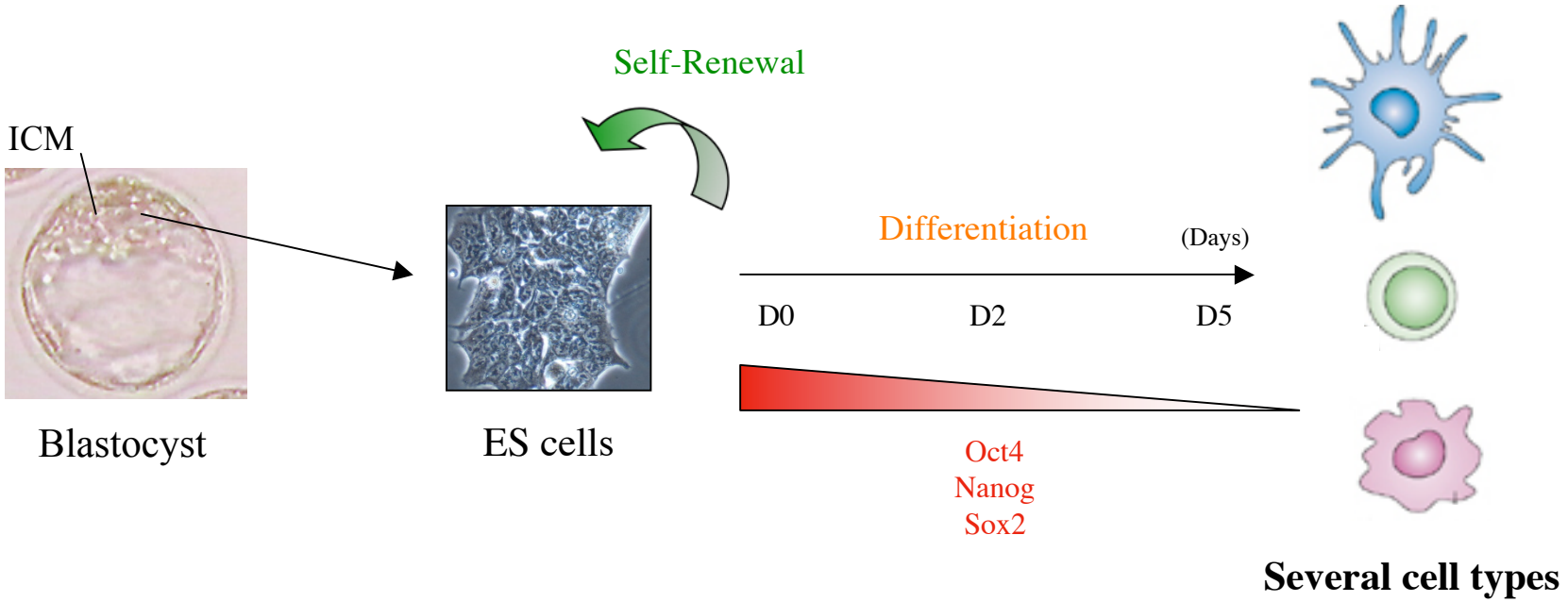
Endogenous-siRNA from repeat elements have been found in mouse oocytes and ES cells

(Tam et al., 2008)

(Watanabe et al., 2008)

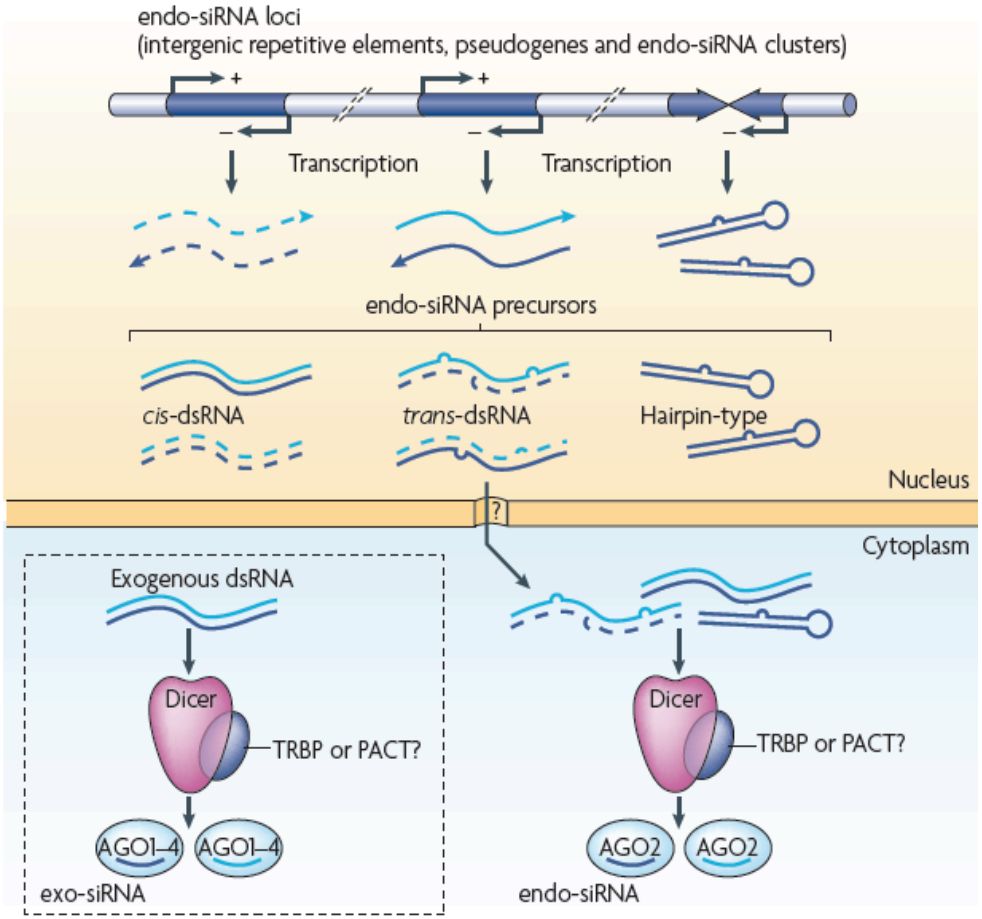
(Calabrese et al., 2006)

Dynamics of small RNAs during ES cell differentiation



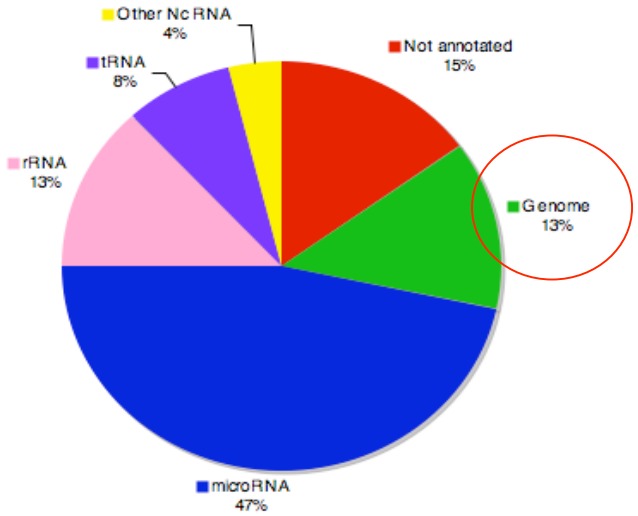
Cell lines used: XX ES cells / XY ES cells

Endogenous-siRNA have been found in mouse oocytes and ES cells



Analysis of non annotated small RNAs

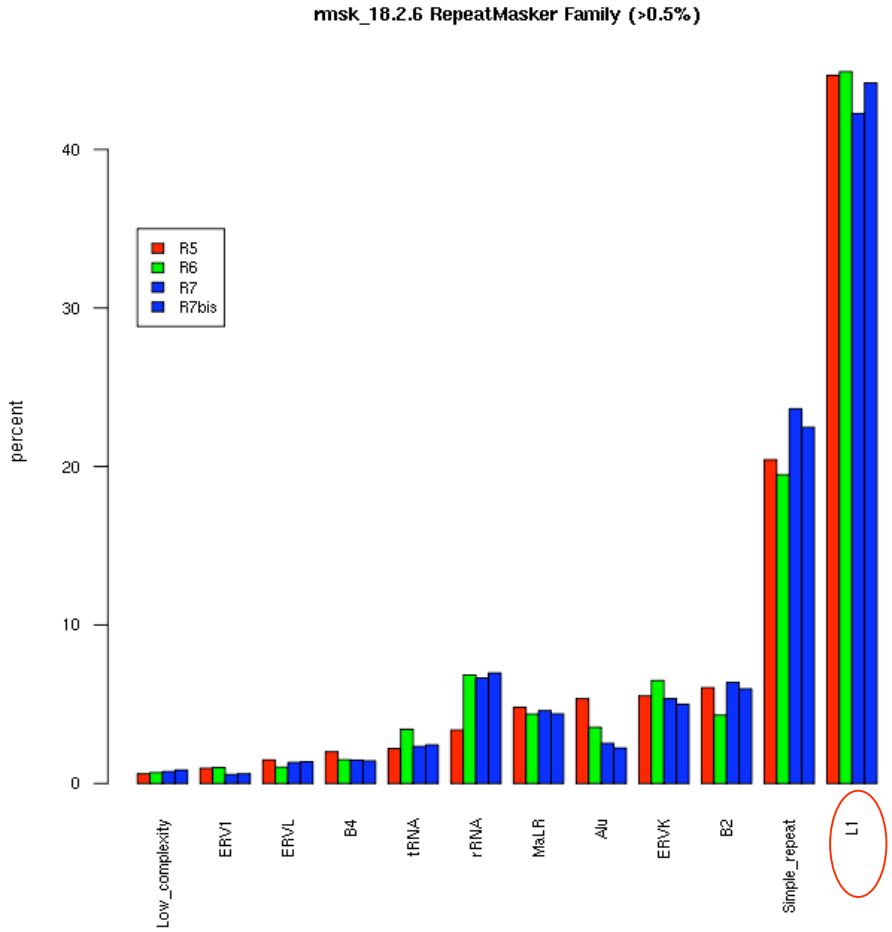
- Derived from repeat elements
- Putative endo-siRNAs

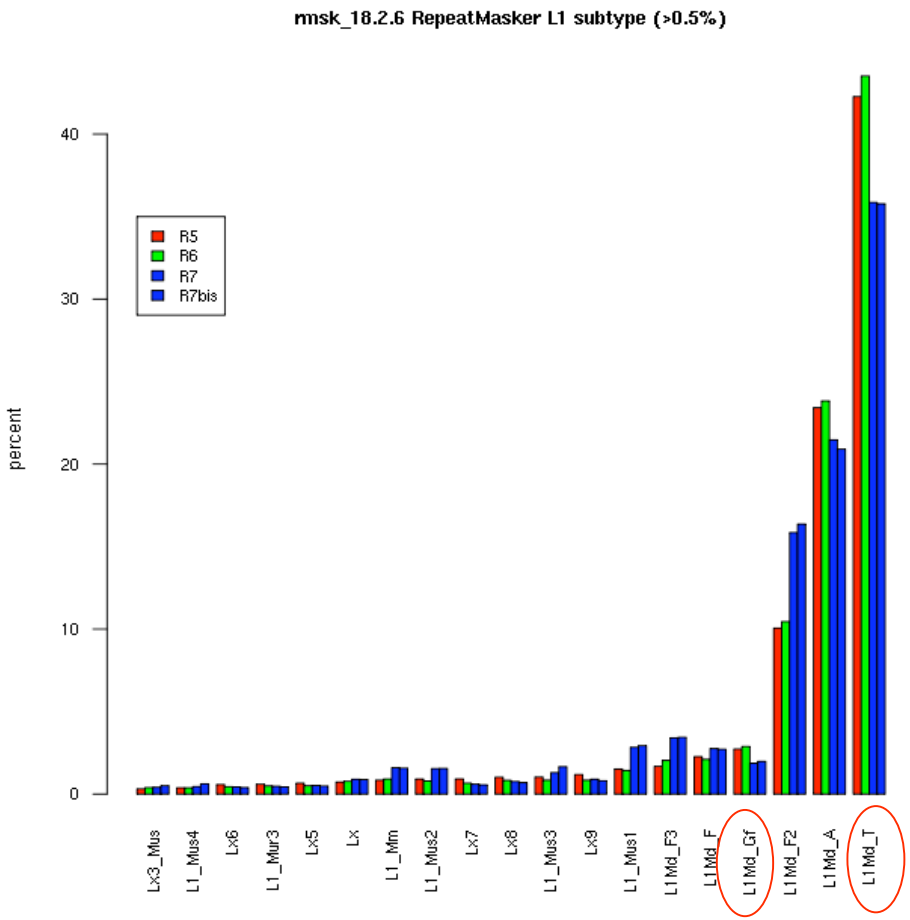


Endo-siRNAs are unique sequence, process “randomly” from a double strand RNA precursor

(Kim et al., 2009)
(Tam et al., 2008)
(Watanabe et al., 2008)
(Calabrese et al., 2006)

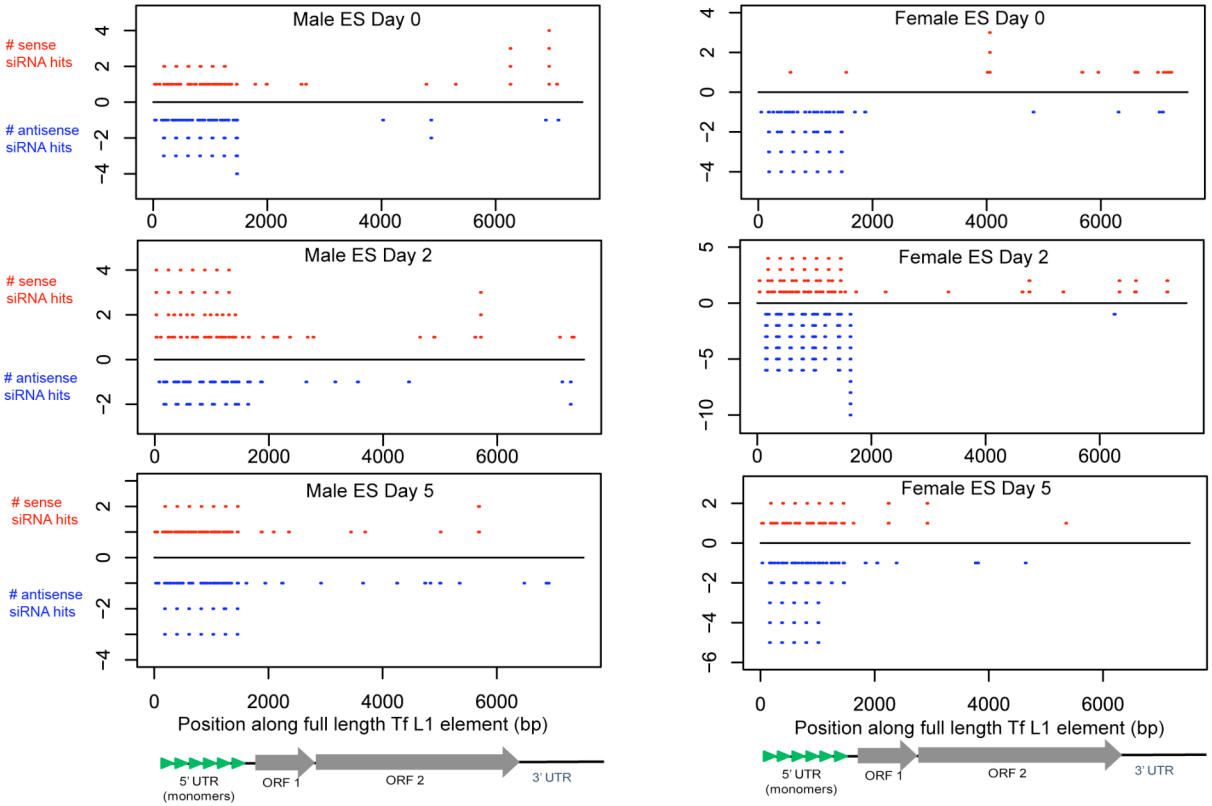
Small RNAs mapped several type of repeats in ES cells





Is there any link between the LINE1 transcription and small RNAs?

Are there small RNAs associated with specific LINE elements and what is their distribution?



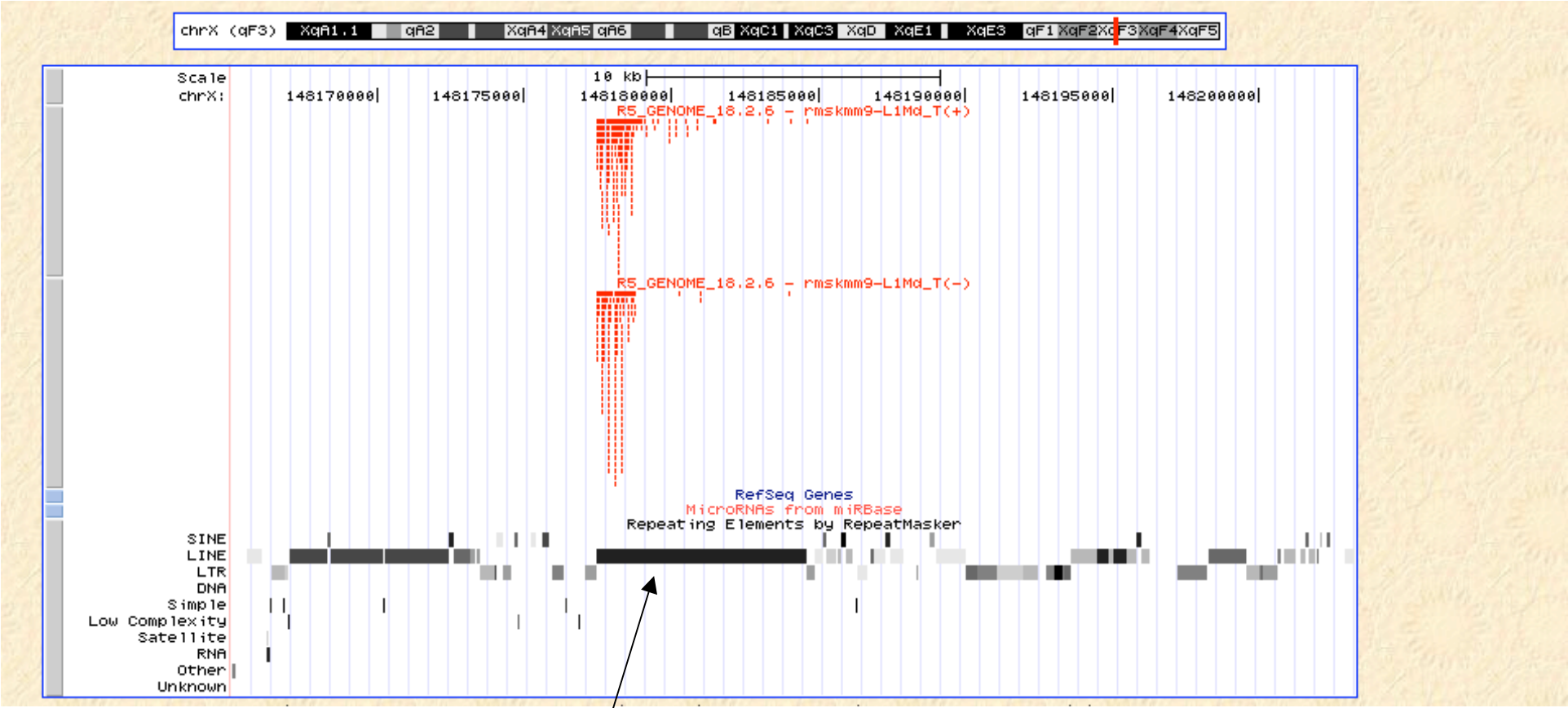
- Small RNAs derived from LINE1 elements can be detected in differentiating male and female ES cells.

- Transcriptionally competent young, full length LINE elements (Tf elements) had the greatest number of hits.

454 sequencing of 19 - 30 nt small RNAs in male and female ES cells (Ciaudo et al., 2009)

(Chow et al., in revision)

SiRNAs derived from LINE1 promoter



Tf element

From SoliD deep sequencing of small RNAs in ES cells (unpublished data)

Genome wide bioinformatics approach

We looking for small RNA in 100 kb region surrounding Tf elements

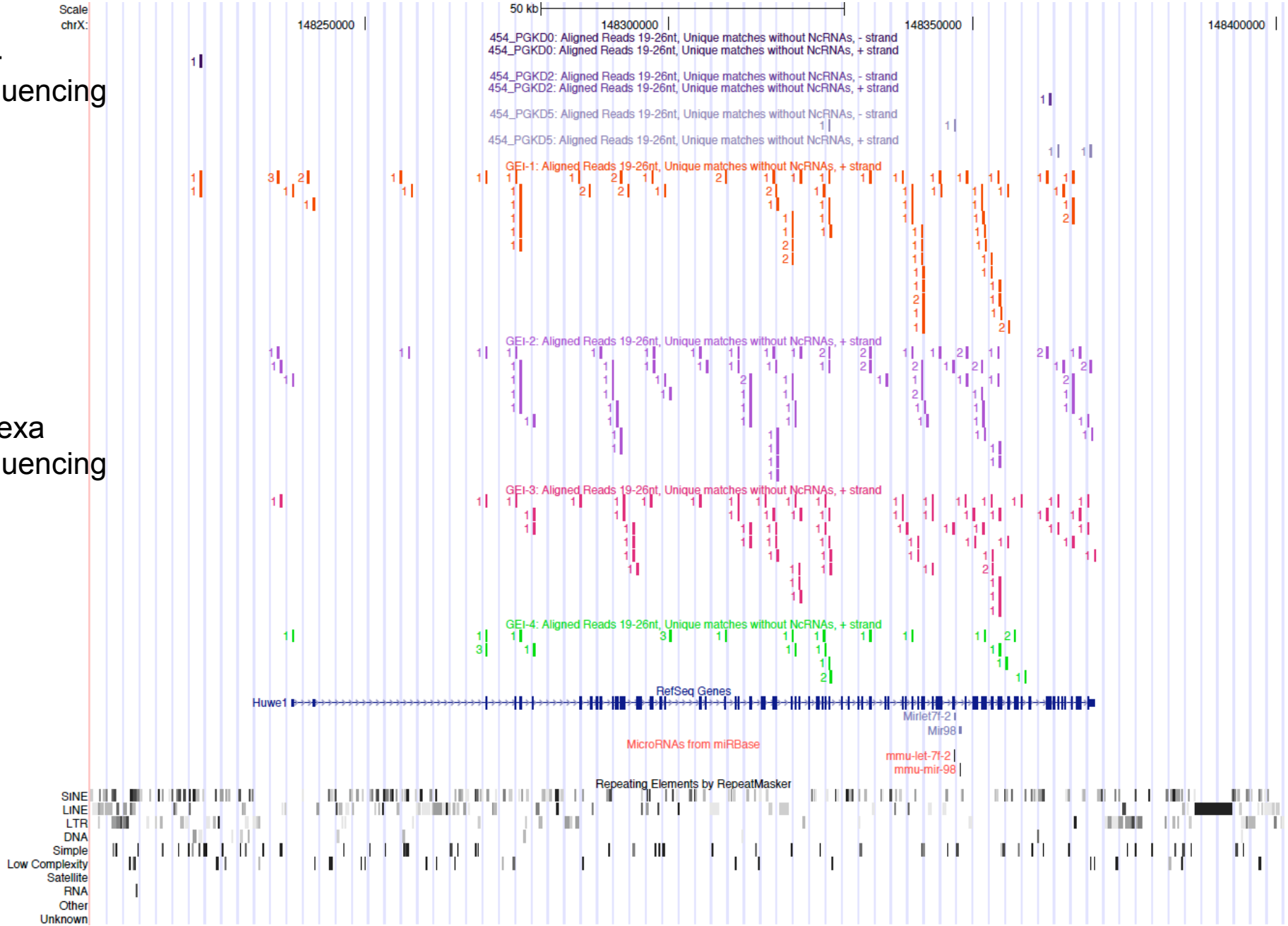
At D5 of the differentiation we identified only on region in female ES cells

(Chow et al., in revision)

Validation by Solexa deep sequencing

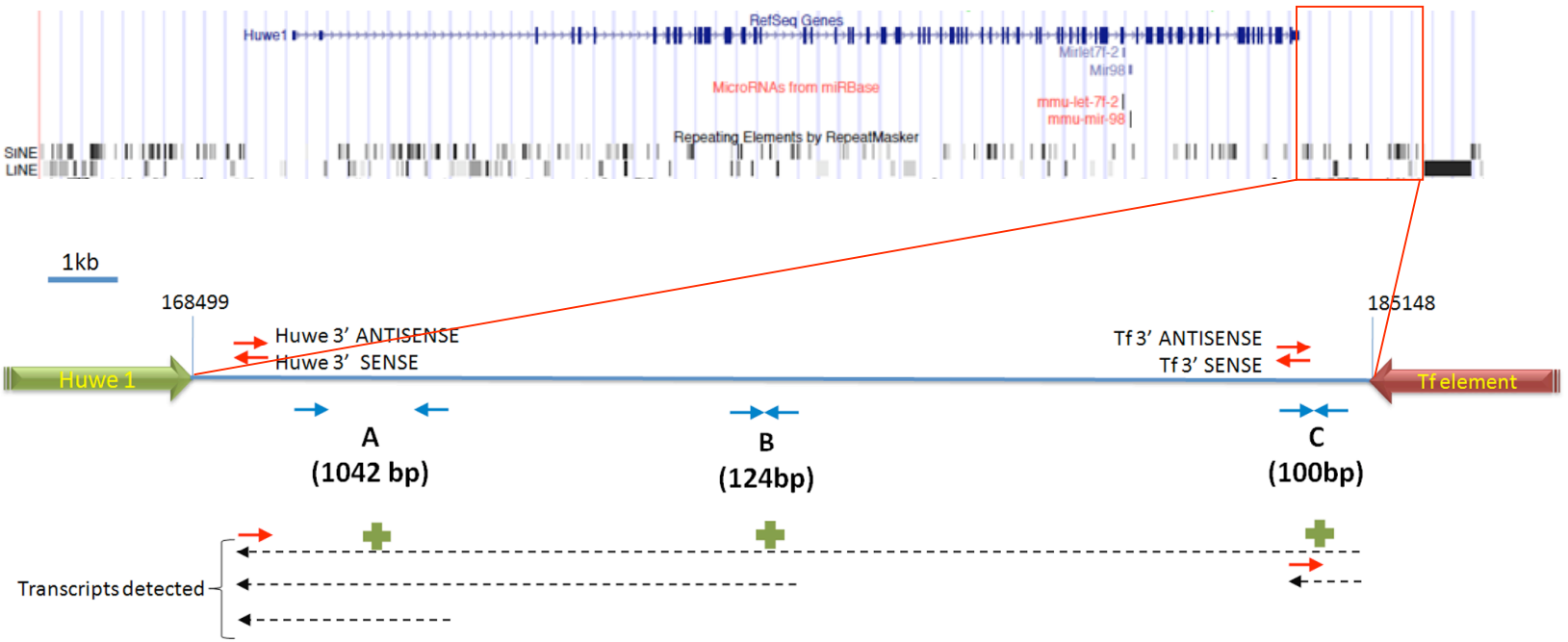
454 sequencing

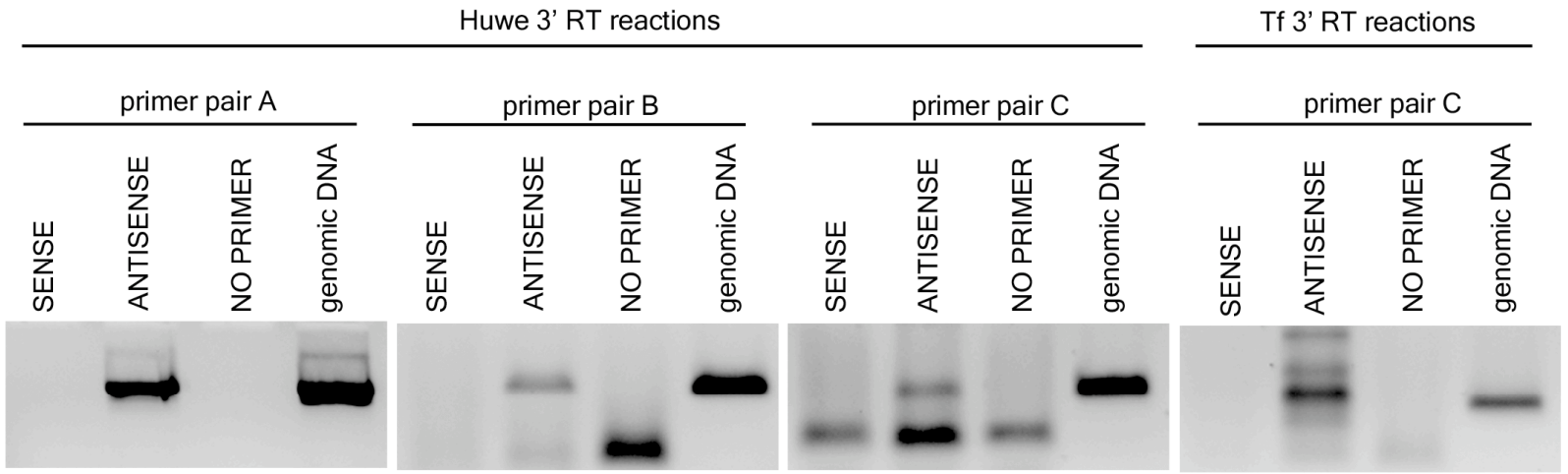
Solexa sequencing



How could these small RNAs be produce only in females?

Antisense transcription running through from the LINE right up to Huwe1



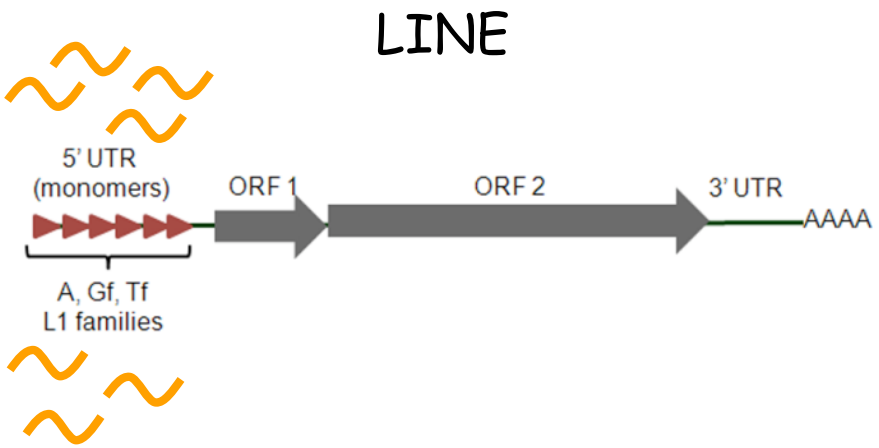


Undifferentiated female ES cells

(Chow et al., in revision)

Identification of endo-siRNAs

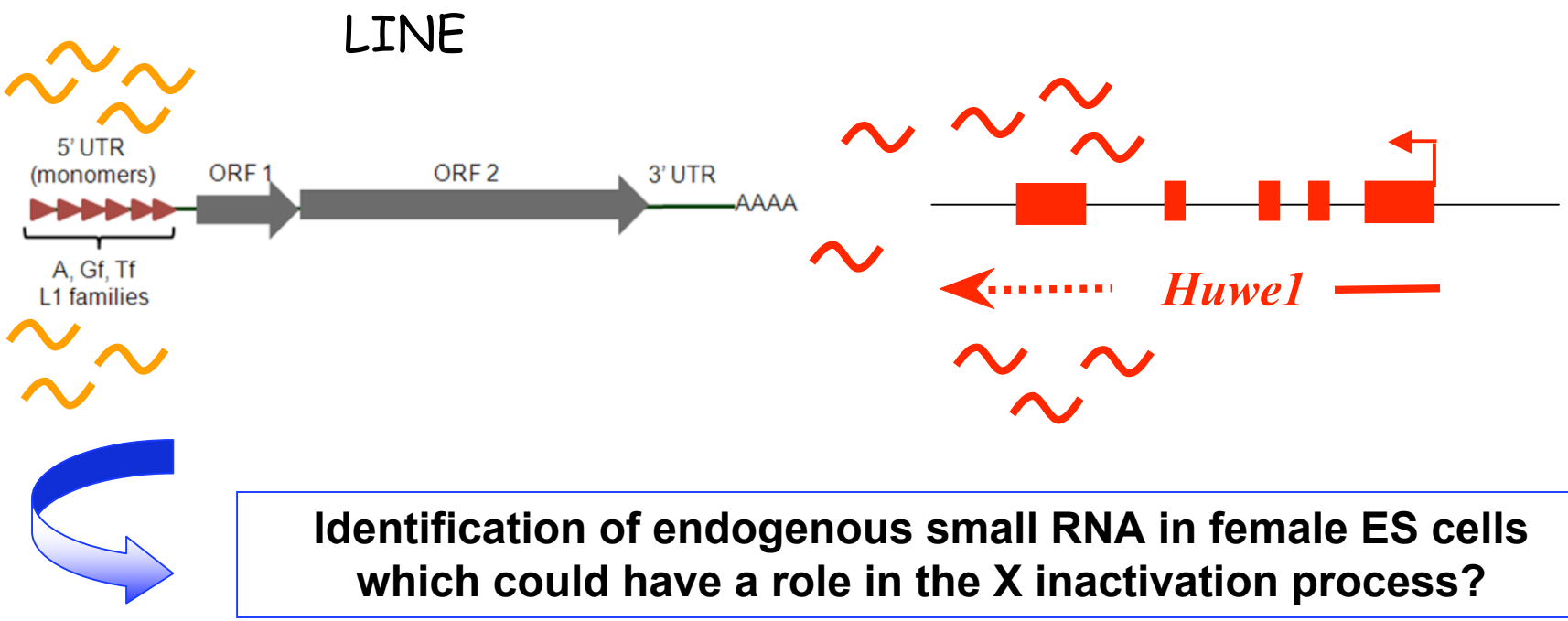
1. From LINEs promoters



Identification of endo-siRNAs

2. From the Huwe1 gene

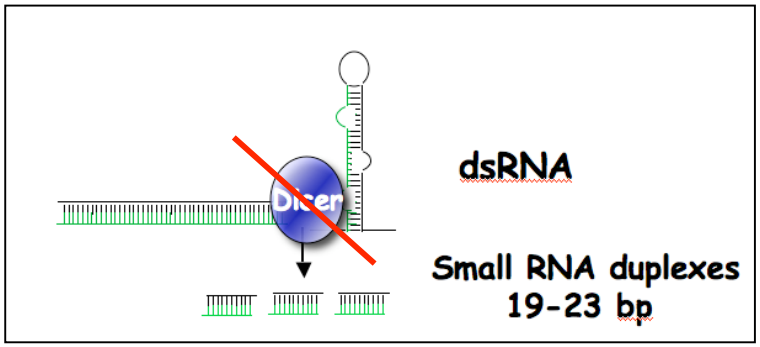
ES at D5



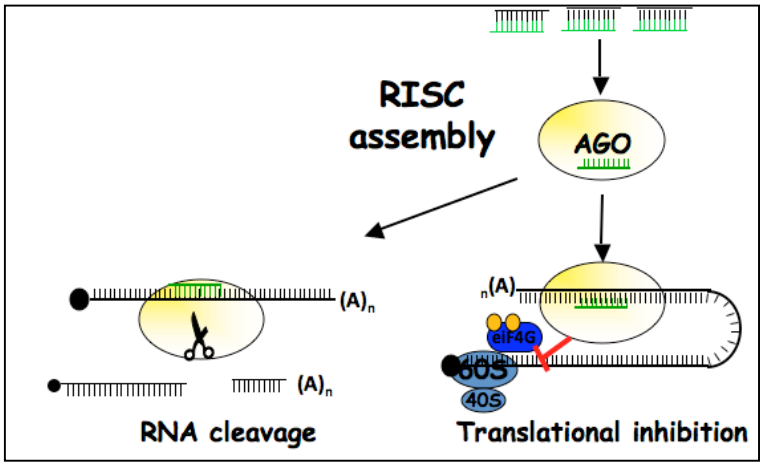
Identification of endogenous small RNA in female ES cells which could have a role in the X inactivation process?

What are the nature and the role of these small RNAs?

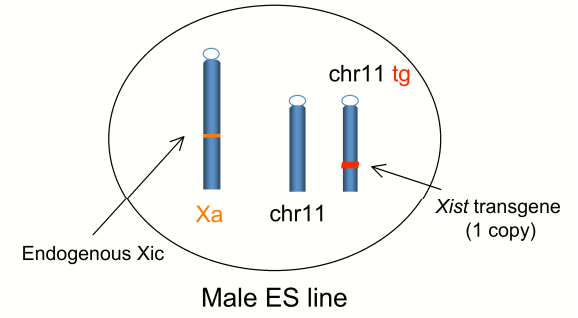
1. Generation of Dicer mutant ES cell lines



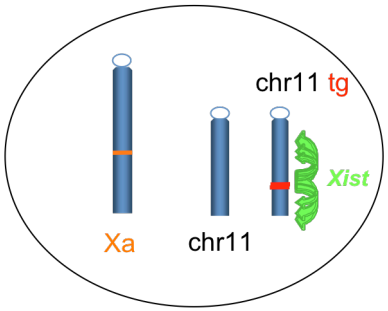
2. Generation of Ago transgenic ES cell lines



3. Used *Xist* inducible transgenic cell lines to try to identify heterochromatic small RNAs in mammalian ES cells



Doxycycline
+



Acknowledgments

Mammalian development and Epigenetics Unit, Curie Institute, Paris

Dr Edith HEARD
Dr Jennifer CHOW
Dr Ikuhiro OKAMOTO
Tim POLLEX

Mechanisms and roles of RNA silencing, IBMP, Strasbourg
Dr Olivier VOINNET
Valérie COGNAT

Collaborators

Cancer et génome : bioinformatique, biostatistiques et épidémiologie d'un système complexe - Unité 900 Inserm/Ecole des Mines/Institut Curie, Paris
Dr Emmanuel BARILLOT
Nicolas SERVANT
Dr Joern TOEDLING

Patrick WINCKER
Julie POULAIN

Solid platform, Curie Institute
Patricia LEGOIX-NE

SIROCCO

SUPPORTED BY ANR

CENTRE NATIONAL
DE LA RECHERCHE
SCIENTIFIQUE

institutCurie
Together, let's beat cancer.

Mammalian Developmental Epigenetics Group

Institut Curie - Centre de Recherche
Bâtiment Biologie du Développement, CNRS UMR 3215 / INSERM U934

HEROIC

FONDATION
RECHERCHE
MÉDICALE