

# Patrocles: a database of polymorphic miR-mediated gene regulation in vertebrates


*Denis Baurain*

*Samuel Hiard*

*Wouter Coppieters*

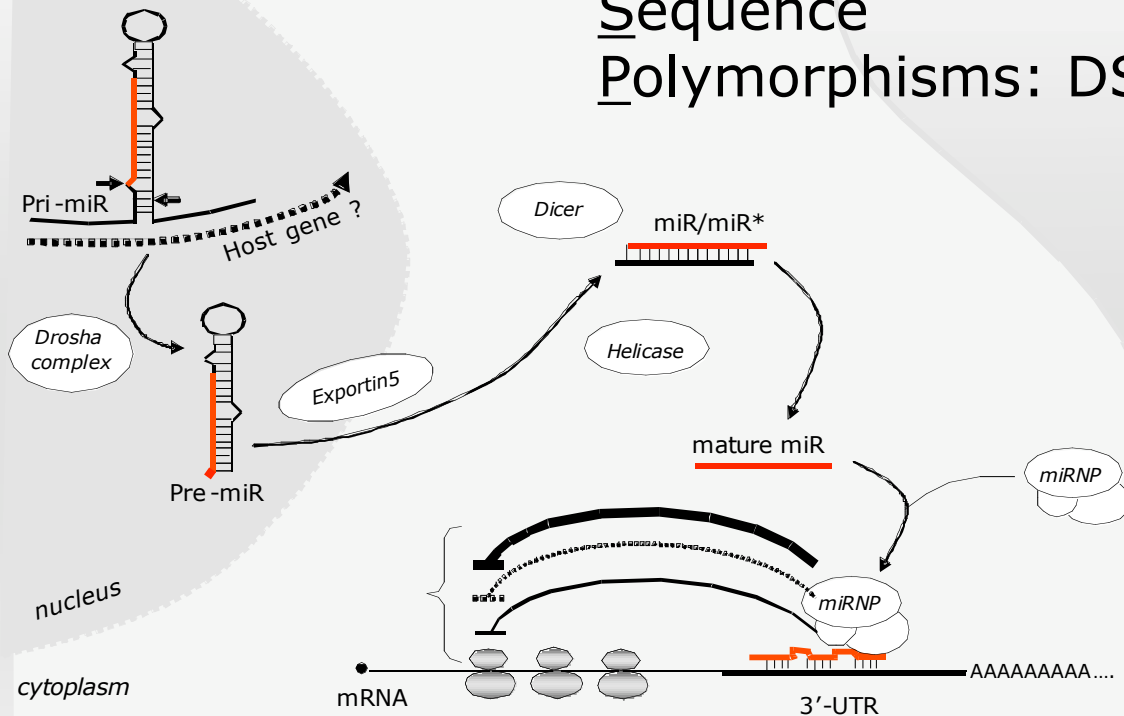
*Carole Charlier*

*Michel Georges*

Unit Animal Genomics	<h2>Patrocles</h2>
Background	<p>The database of polymorphic miRNA-target interactions</p>
Polymorphic targets	
Polymorphic miRNAs	
Polymorphic silencing machinery	
Patrocles finder	
Useful links	
Patrocles Home	<p>Why the name "PATROCLES"?</p> <p>To a large extent, Patrocles was killed by Hector because he went to battle wearing the armour of his friend Achilles. Likewise, the mutant Texel MSTN mRNA has mistakenly become the target of miRNAs because of its disguise using a target octamer motif borrowed from genuine target genes.</p>

# Polymorphic miR-mediated gene regulation

DNA  
Sequence  
Polymorphisms: DSPs



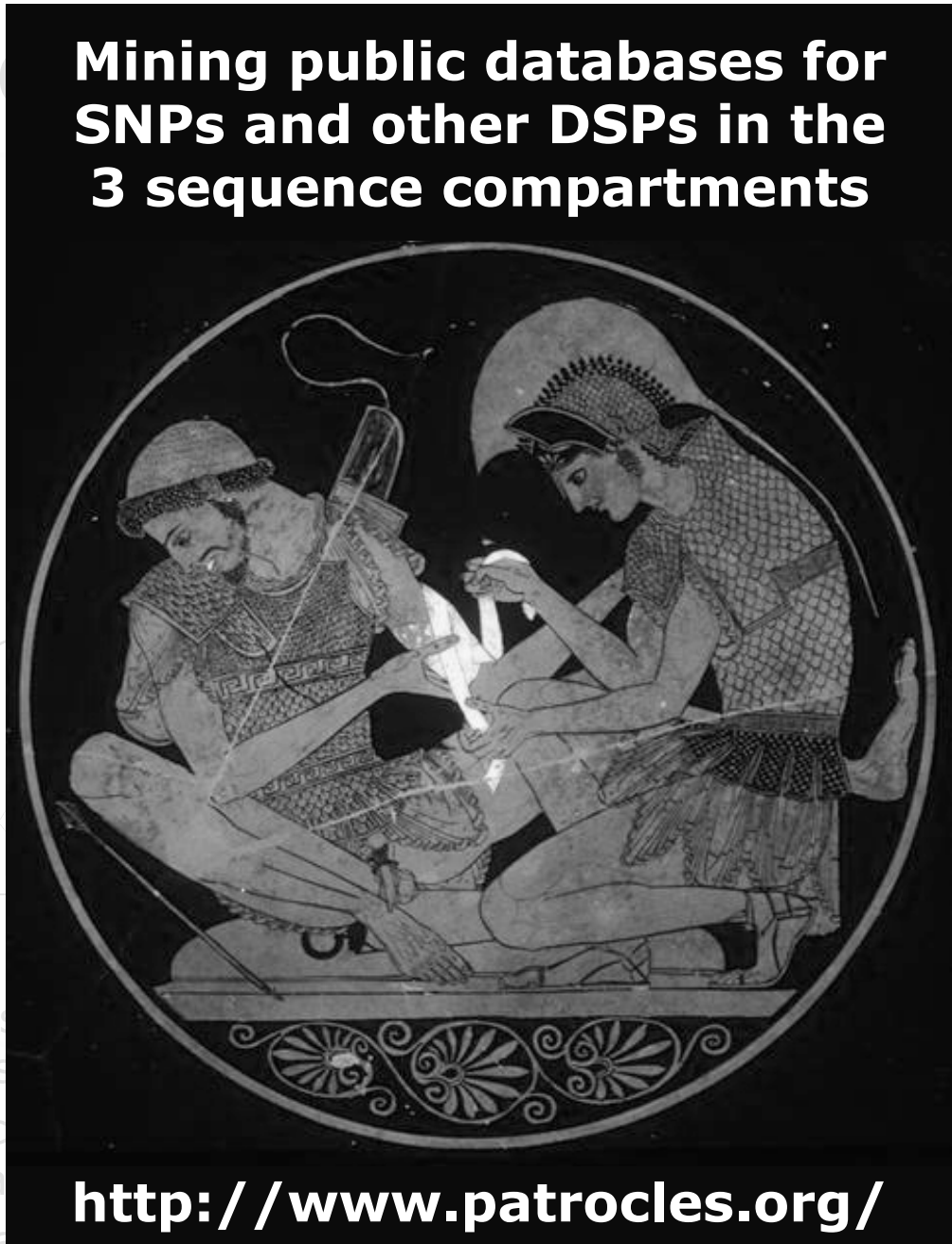
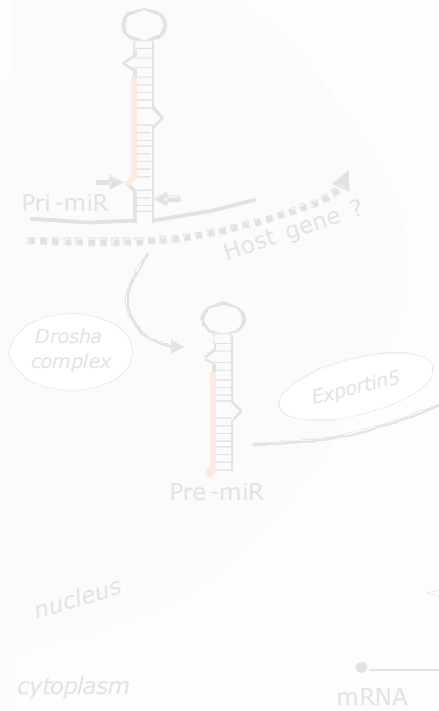
- Targets (1)  
    ∨
- miRs (100s)  
    ∨
- Silencing machinery (overall effect)

- Considerable sequence space is devoted to miR-mediated gene regulation (targets, miRs, silencing machinery)
- DSPs in silencing components are likely to contribute to (complex) phenotypic variation including disease
- Proof in animals: Texel sheep, Clop *et al.* (2006) *Nat. Genet.* 38:813-818
- Suggestions in humans: Sethupathy & Collins (2008) *TIG* 24:489-497

P

# Mining public databases for SNPs and other DSPs in the 3 sequence compartments

mediated  
ulation



ets (1)

(100s)

icing machinery  
rall effect)

gene regulation

omplex)

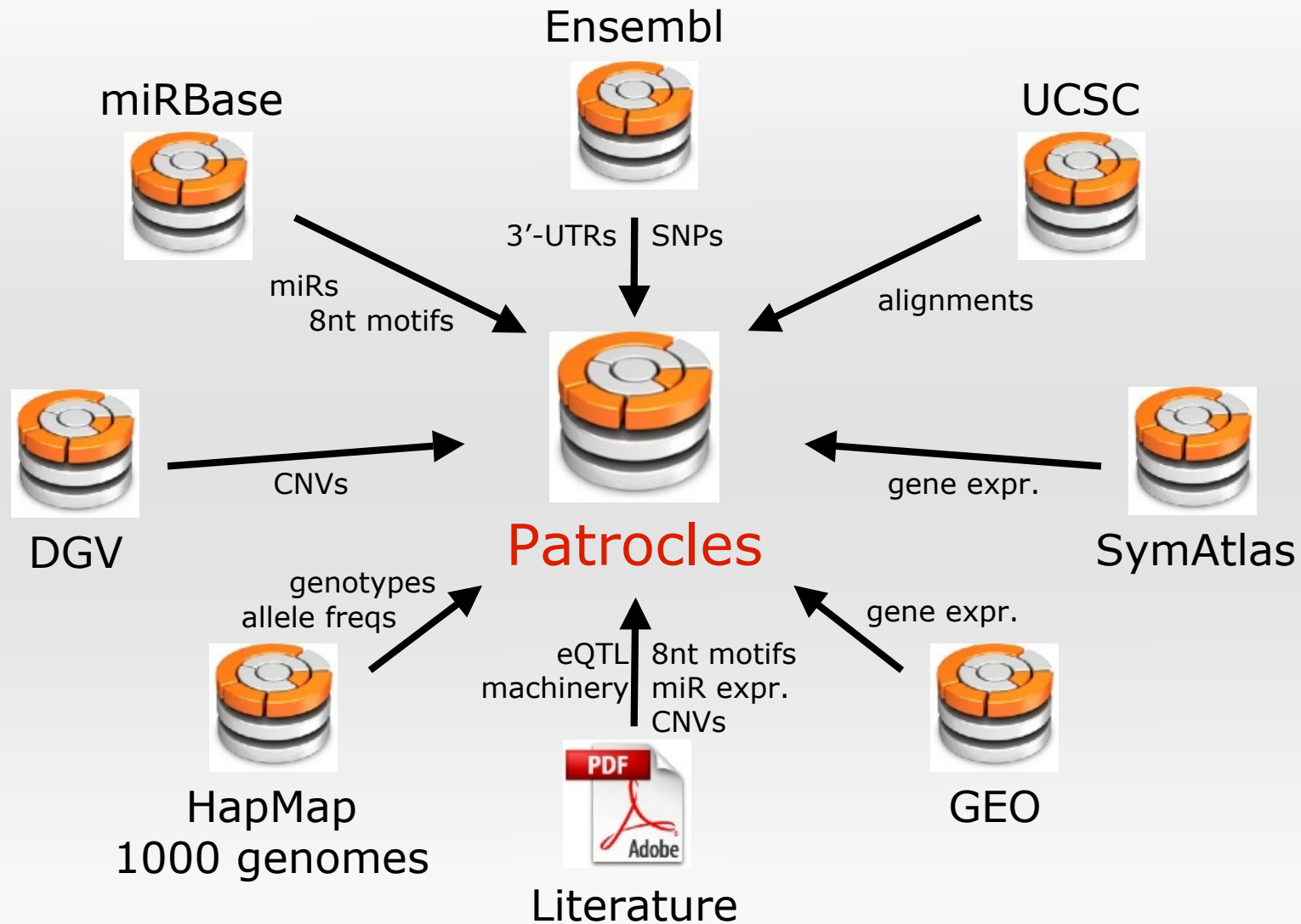
38:813-818

24:489-497

- Considerable (targets, miRs)
- DSPs in silencing phenotypic variation
- Proof in animals: Rexer sheep, Glop et al. (2006) *Nat. Genet.* 38:813-818
- Suggestions in humans: Sethupathy & Collins (2008) *TIG* 24:489-497

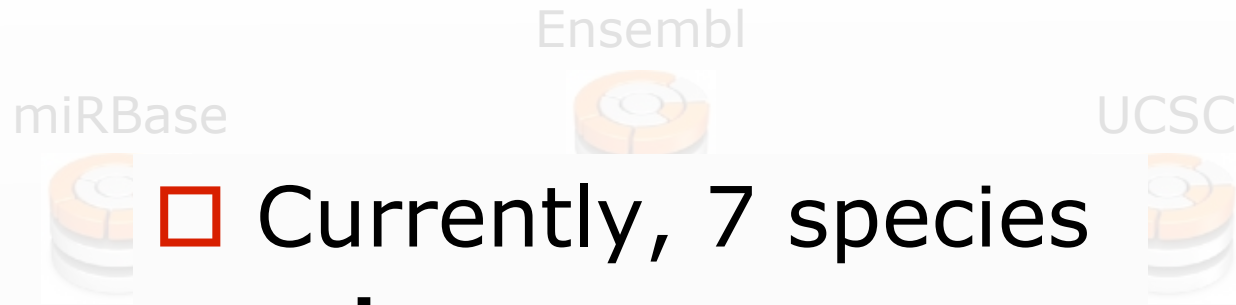
<http://www.patrocles.org/>

# Patrocles - Overview



**Updates and Synchronization...**

# Patrocles - Overview



□ Currently, 7 species

- **human**
- chimp
- **mouse**
- rat
- dog
- cow
- chicken



DGV



SymAtlas

1000 genomes

Literature

Updates and Synchronization...

		<b>human</b>	<b>mouse</b>
3'-UTRs	genes	24,319	21,911
	sequence space	26,261,732	21,634,548
SNPs in 3'-UTRs	total	136,159	126,589
	known ancestral allele	114,305 (83.9%)	111,178 (87.8%)
target site motifs	X-octamers	540	540
	miR	676	484
	miR*	170	117
	L-octamers	683	466
	X OR L-octamers	1,164	948
	X AND L-octamers	59	58
target sites in 3'-UTRs	X-targets	323,833	267,644
	L-targets	375,054	219,392
	X OR L-targets	661,187	455,620
	X AND L-targets	37,700	31,416
	conserved X AND L-targets	10,425 (27.7%)	9,436
	conserved X NOT L-targets	64,010 (22.4%)	57,154
	conserved L NOT X-targets	30,290 (9.0%)	19,595
	sequence space	4,072,176 (15.5%)	2,674,395 (12.4%)

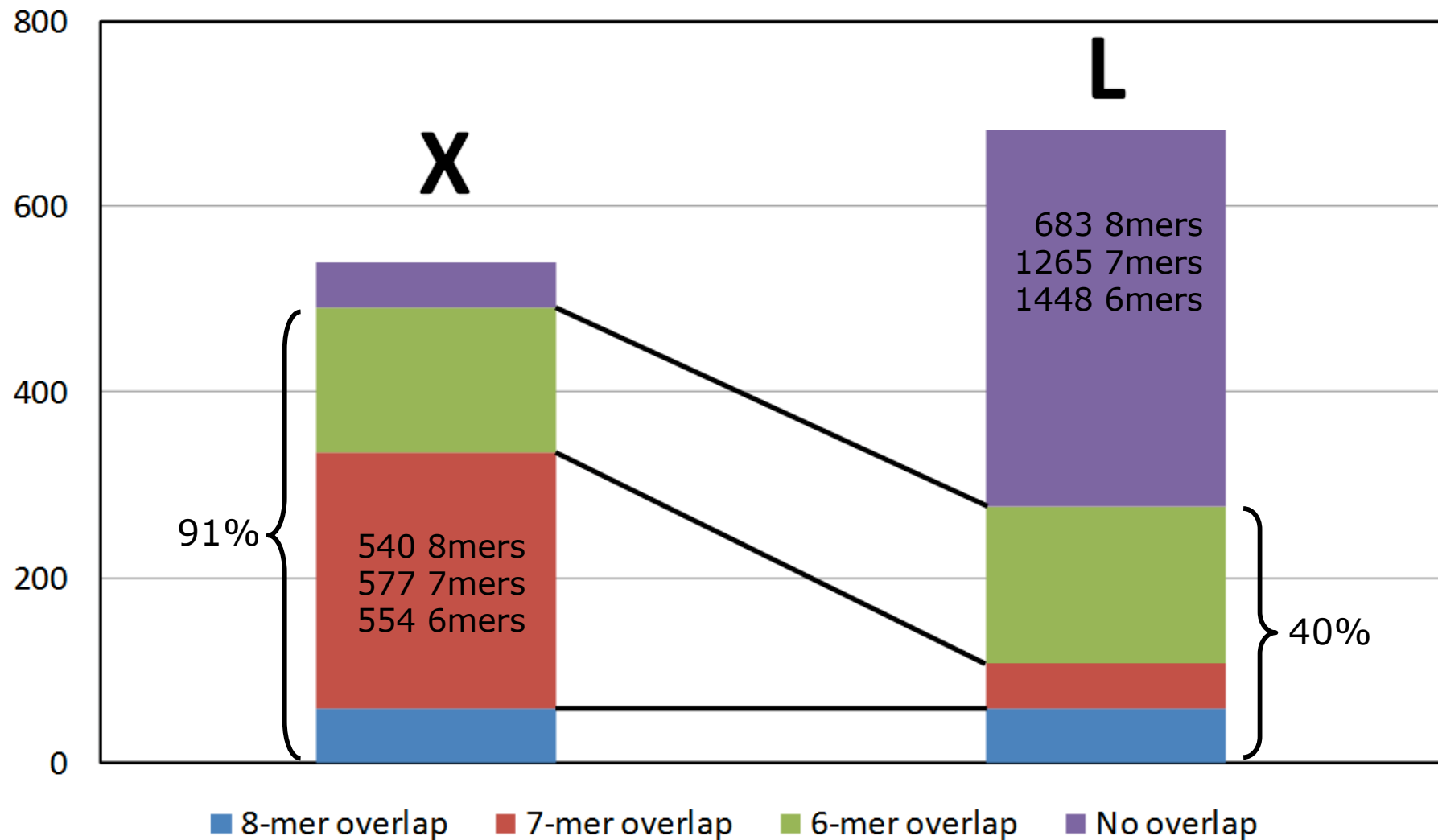
		<b>human</b>	<b>mouse</b>
3'-UTRs	genes	24,319	21,911
	sequence space	26,261,732	21,634,548
SNPs in 3'-UTRs	total	136,159	126,589
	known ancestral allele	114,305 (83.9%)	111,178 (87.8%)
target site motifs	X-octamers	540	540
	miR	676	484
	miR*	170	117
	L-octamers	683	466
	X OR L-octamers	1,164	948
	X AND L-octamers	59	58
target sites in 3'-UTRs	X-targets	323,833	267,644
	L-targets	375,054	219,392
	X OR L-targets	661,187	455,620
	X AND L-targets	37,700	31,416
	conserved X AND L-targets	10,425 (27.7%)	9,436
	conserved X NOT L-targets	64,010 (22.4%)	57,154
	conserved L NOT X-targets	30,290 (9.0%)	19,595
	sequence space	4,072,176 (15.5%)	2,674,395 (12.4%)



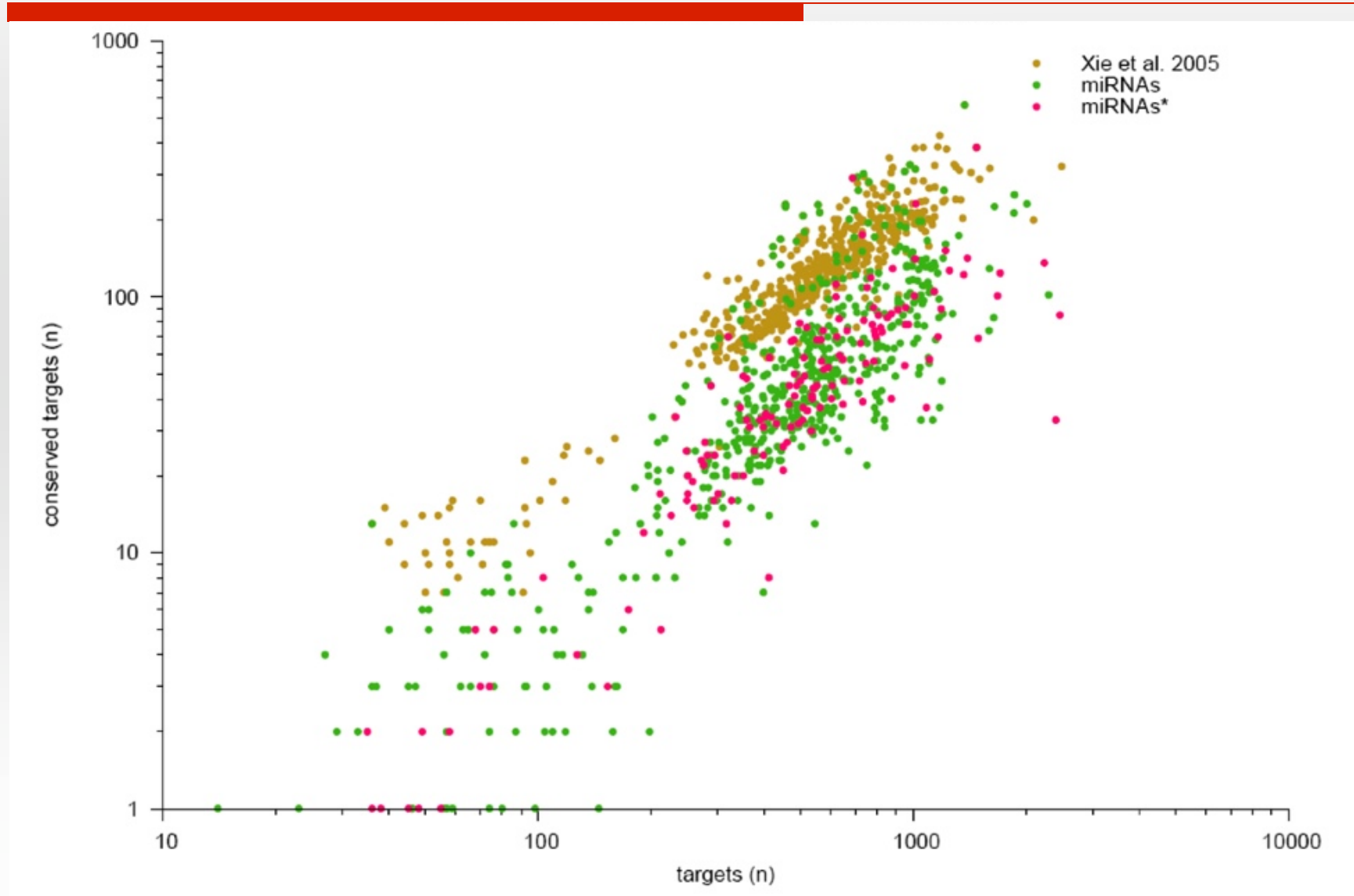
		<b>human</b>	<b>mouse</b>
3'-UTRs	genes	24,319	21,911
	sequence space	26,261,732	21,634,548
SNPs in 3'-UTRs	total	136,159	126,589
	known ancestral allele	114,305 (83.9%)	111,178 (87.8%)
target site motifs	X-octamers	540	540
	miR	676	484
	miR*	170	117
	L-octamers	683	466
	X OR L-octamers	1,164	948
	X AND L-octamers	59	58
target sites in 3'-UTRs	X-targets	323,833	267,644
	L-targets	375,054	219,392
	X OR L-targets	661,187	455,620
	X AND L-targets	37,700	31,416
	conserved X AND L-targets	10,425 (27.7%)	9,436
	conserved X NOT L-targets	64,010 (22.4%)	57,154
	conserved L NOT X-targets	30,290 (9.0%)	19,595
	sequence space	4,072,176 (15.5%)	2,674,395 (12.4%)

		<b>human</b>	<b>mouse</b>
3'-UTRs	genes	24,319	21,911
	sequence space	26,261,732	21,634,548
SNPs in 3'-UTRs	total	136,159	126,589
	known ancestral allele	114,305 (83.9%)	111,178 (87.8%)
target site motifs	X-octamers	540	540
	miR	676	484
	miR*	170	117
	L-octamers	683	466
	X OR L-octamers	1,164	948
	X AND L-octamers	59	58
target sites in 3'-UTRs	X-targets	323,833	267,644
	L-targets	375,054	219,392
	X OR L-targets	661,187	455,620
	X AND L-targets	37,700	31,416
	conserved X AND L-targets	10,425 (27.7%)	9,436
	conserved X NOT L-targets	64,010 (22.4%)	57,154
	conserved L NOT X-targets	30,290 (9.0%)	19,595
	sequence space	4,072,176 (15.5%)	2,674,395 (12.4%)

# Targets - Concordance between X and L target site motifs



# Targets - Conserved vs. total numbers of target sites



# Targets

## Patrocles SNPs - Methods

---

1.	human: A	...TTTGGTGAACCAAC...	=> ancestral allele
	human: G	...TTTGGTGAACCAAC...	=> derived allele
	chimp	...TTTGGTGAACCAAC...	=> sibling species
2.	rat	...TTTGGTGAACAAAC...	
	mouse	...CTTGGTGAACAAAC...	
3.	dog	...TTTGGTGAACTAAC...	
	cow	...TTTGGTGAACTAAC...	
(3/3)		TTTGGTGA	
(3/3)		TTGGTGA	
(3/3)		TGGTGA	
(3/3)		GGTGA	
(2/3)	not in dog/cow	gtgaacc	
(2/3)	not in dog/cow	tgaacca	
(2/3)	not in dog/cow	gaaccaa	=> hsa-miR-29b-2*
(2/3)	not in dog/cow	aaaccaac	

---

# Targets

## Patrocles SNPs - Methods

1. human: A ...TTTGGTGAACCAAC... => ancestral allele

|||||||

3'-GAUUCGGUGGUACACUUUGGUC-5' => hsa-miR-29b-2\*

|||.||||

human: G ...TTTGGTGAACCAAC... => derived allele

chimp ...TTTGGTGAACCAAC... => sibling species

2. rat ...TTTGGTGAACAAAC...

mouse ...CTTGGTGAACAAAC...

3. dog ...TTTGGTGAACTAAC...

cow ...TTTGGTGAACTAAC...

(3/3) TTTGGTGA

.....

.....

(2/3) not in dog/cow gaaccaa => hsa-miR-29b-2\*

(2/3) not in dog/cow aaccaac

site \ allele	anc	der	?
conserved	DC	(CC)	P
not cons.	DNC	CNC	

+S +W7C / S7C

# Targets

## Patrocles SNPs - Results

---

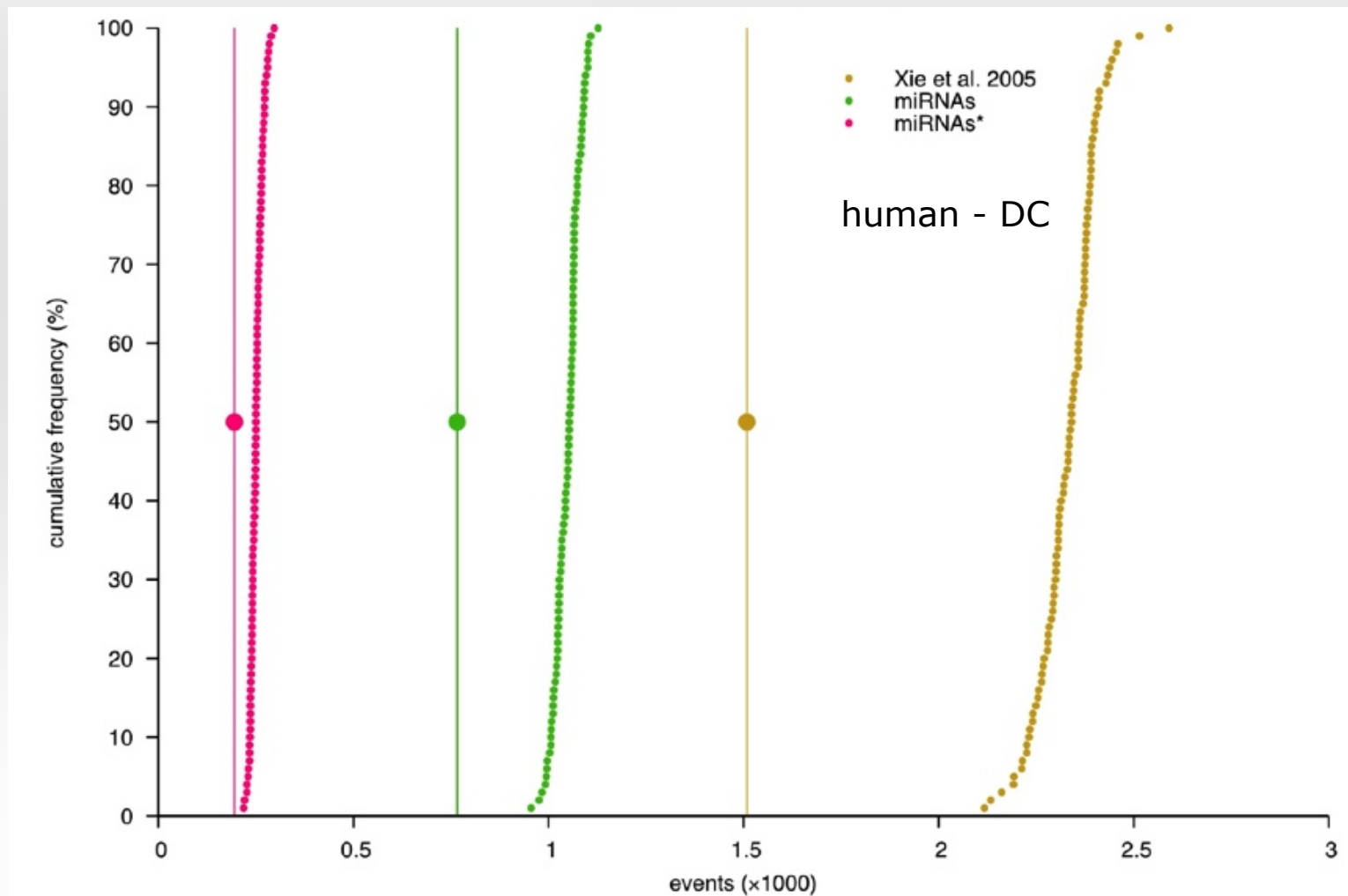
	human		mouse	
pSNP class	Xie	Lewis	Xie	Lewis
total	20,679	26,719	19,657	17,505
DC+CC	1,546+50	959+58	951+102	496+65
DNC	7,392	10,328	7,732	7,250
CNC	9,006	11,244	8,545	7,573
P	1,944	3,295	2,290	2,065
S	741	837	37	56

} # destructions  
=  
} # creations

# Targets - Patrocles SNPs

## Evidence for purifying selection

SNP shuffling in 3'-UTR sequence space with preservation of trinucleotide context



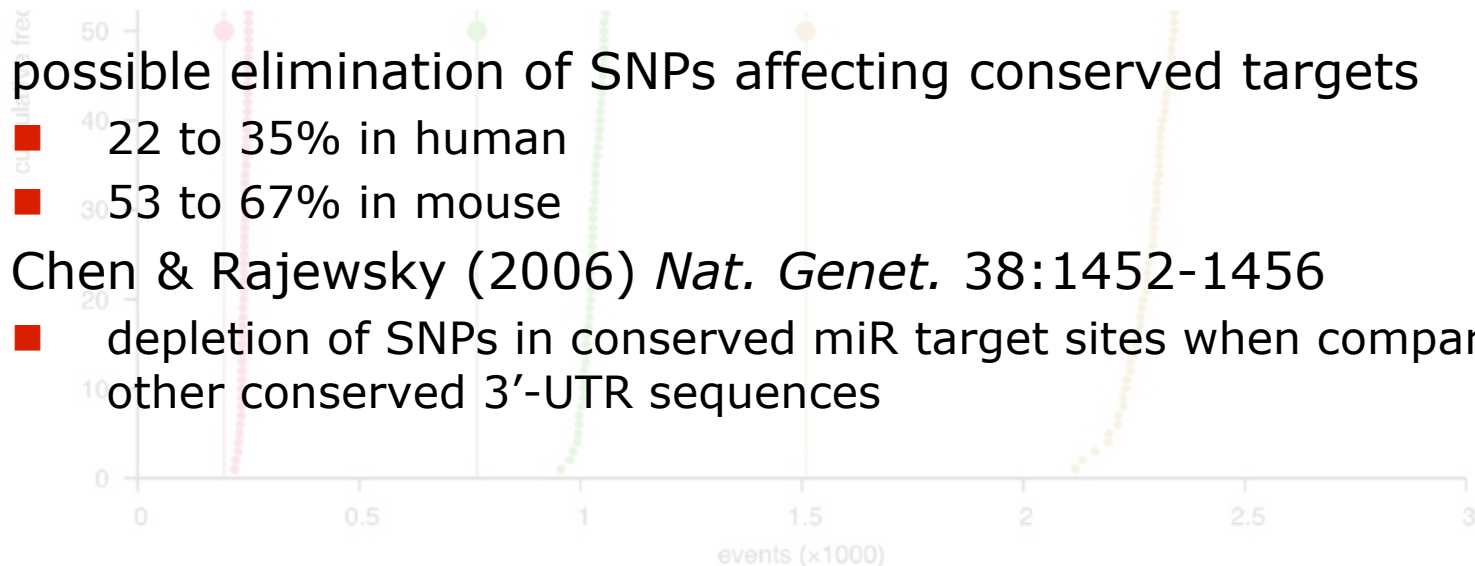
# Targets - Patrocles SNPs

## Evidence for purifying selection

CND shuffling in 3' UTR sequence space with preservation of trinucleotide context

	DC			DNC			CNC			
	X	L	L*	X	L	L*	X	L	L*	
Human	1509	766	195	7592	8160	2530	9202	9005	2618	OBS
	0.647	0.730	0.775	0.968	0.968	1.009	0.916	0.953	0.987	[OBS/SIM]
	-10.832	-8.843	-3.766	-2.249	-3.302	0.503	-5.743	-4.489	-0.559	[OBS-SIM]/SD_SIM
Mouse	951	410	94	8850	6604	1759	9598	6711	1933	OBS
	0.324	0.390	0.467	0.987	0.997	1.070	0.890	0.959	1.014	[OBS/SIM]
	-23.933	-18.267	-8.320	-0.814	-0.240	2.706	-8.061	-2.949	0.627	[OBS-SIM]/SD_SIM

- possible elimination of SNPs affecting conserved targets
  - 22 to 35% in human
  - 53 to 67% in mouse
- Chen & Rajewsky (2006) *Nat. Genet.* 38:1452-1456
  - depletion of SNPs in conserved miR target sites when compared to other conserved 3'-UTR sequences

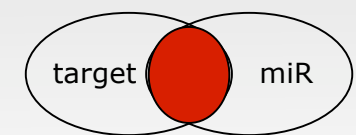


# Targets

## Prioritization for lab validation

---

- most interesting pSNPs are
  - pSNPs destroying conserved target sites
  - pSNPs creating target sites in anti-targets
- to yield a phenotype, target and miR have to be expressed in the same tissue (at the same time)
- co-expression plots for human and mouse
  - target genes: SymAtlas
  - miRs: Landgraf *et al.* (2007) *Cell* 129:1401-1414
- two different kinds of plots
  - comparing miR and target
  - comparing miR host gene (if any) and target



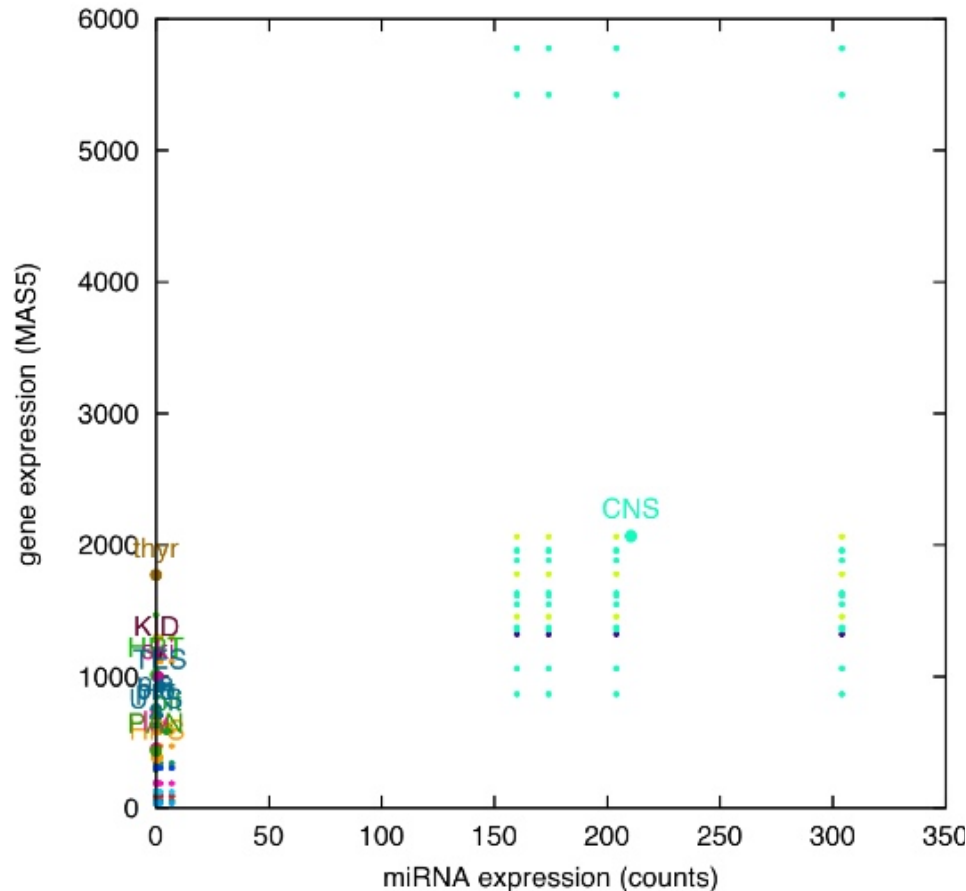
tissue mapping?

---

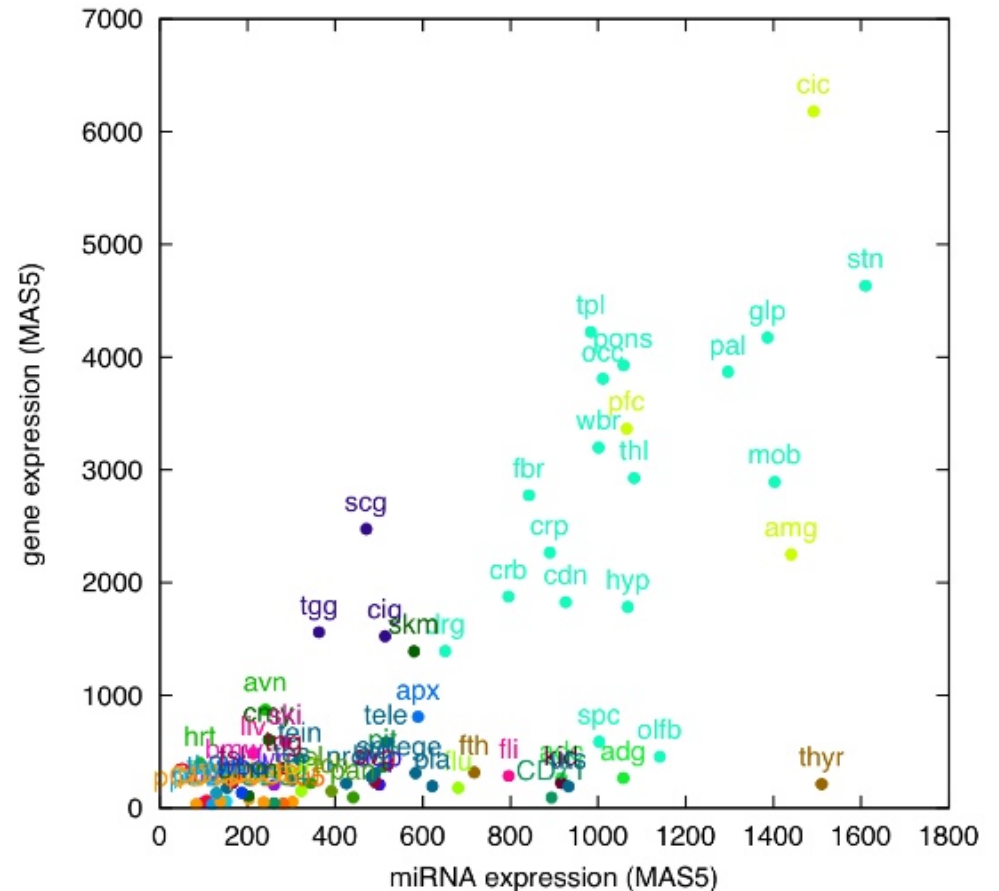
# pSNPs - Co-expression plots

rs34542287 A/G [0.985/0.015]  
 Destroyed Conserved target site  
 miR-9 vs. actin-binding LIM protein 1  
 ACCA[A]AGA

rs28399411 G/A [0.994/0.006]  
 Destroyed Conserved target site  
 miR-32 vs. Axonal membrane protein GAP-43  
 TGTGC[A]AT



mature miR counts  
 [+] direct evidence  
 [-] gross tissue mapping



host gene expression  
 [+] perfect matching of tissues  
 [-] indirect evidence

# Polymorphic miRs - Methods

□ pri-miRs (stem-loops) from miRBase

□ pDSPs altering miR **sequence**

■ SNPs (de-)stabilizing interaction (seed, mature non-seed)

□ pDSPs altering miR **concentration**

■ SNPs altering processing efficiency (anywhere in stem-loop)

■ CNVs encompassing miR genes

□ **human:** <http://projects.tcag.ca/variation/>

□ **mouse:** She *et al.* (2008) *Nat. Genet.* 40:909-914

□ **rat:** Guryev *et al.* (2008) *Nat. Genet.* 40:538-545

■ eQTL (or allelic imbalance) corresponding to host genes (only **human**)

□ Morley *et al.* (2004) *Nature* 430:743-747

□ Cheung *et al.* (2005) *Nature* 437:1365-1369

□ Ge *et al.* (2005) *Genome Res.* 15:1584-1591

□ Stranger *et al.* (2005) *PLoS Genet.* 1:e78

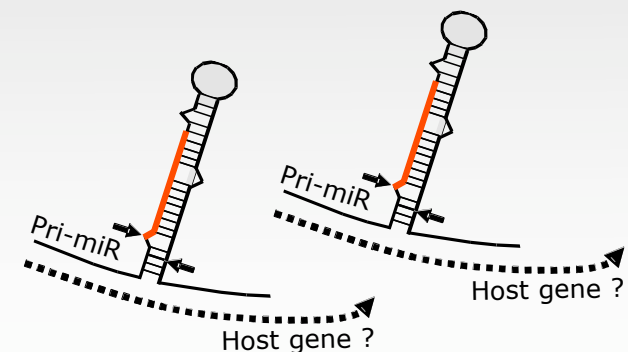
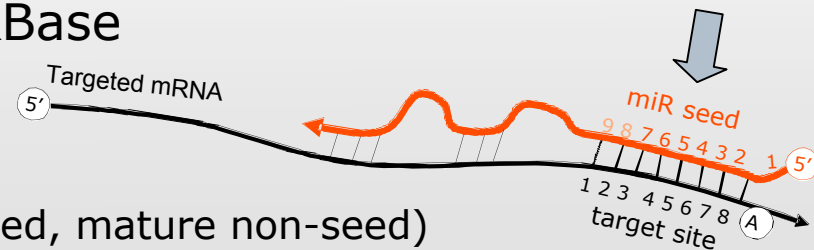
□ Pant *et al.* (2006) *Genome Res.* 16:331-339

□ Dixon *et al.* (2007) *Nat. Genet.* 39:1202-1207

□ Goring *et al.* (2007) *Nat. Genet.* 39:1208-1216

□ Spielman *et al.* (2007) *Nat. Genet.* 39:226-231

□ Stranger *et al.* (2007) *Nat. Genet.* 39:1217-1224



# Polymorphic miRs – Results

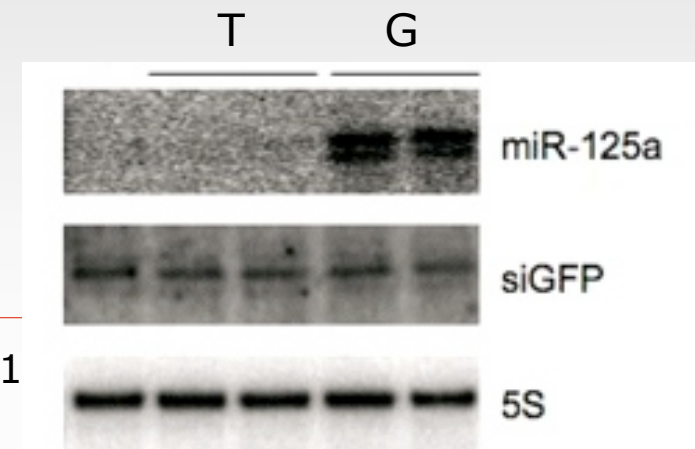
		<b>human</b>	<b>mouse</b>
	pre-miRs	676	466
SNPs in pre-miRs	affected miRs	136	71
	total	184	89
	seed	12	4
	mature non-seed	26	6
	other	146	79
miRs in CNVs	CNVs	158	0
	affected miRs	256	0
miRs hosted in eQTL genes	eQTL	78	n.d.
	affected miRs	85	n.d.

# Polymorphic miRs – Results

SNP					
SNP Id	SNP valid	SNP site	Patrocles alleles	DbSnp alleles	Sequence
<a href="#">rs12975333</a>	?	seed	G/T	G/T	<pre>           U   U U   UC  UG  C   UA   ----- A G (WT) GCCAG C CUAGG CC  AGA CCUU ACCUGUGA GG C         CCGUC G GGUCC GG  UCU GGAG UGGACACU CC A           C   U C   GA  GU  U   --   GGGA U           -----           U   U U   UC  UG  UAC  UA   ----- A T       GCCAG C CUAGG CC  A  CCUU ACCUGUGA GG C         CCGUC G GGUCC GG  U  GGAG UGGACACU CC A           C   U C   GA  GU  CUU  --   GGGA U           </pre>

e.g., hsa-miR-125a

SNP in seed (+8) blocks processing of pri-miR to pre-miR

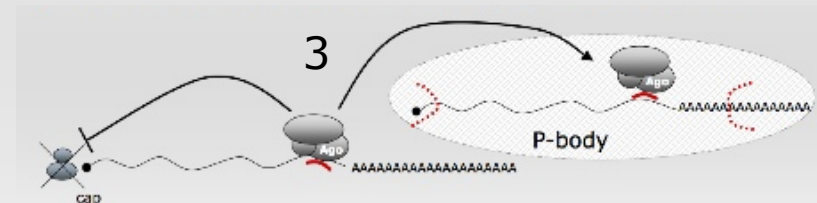
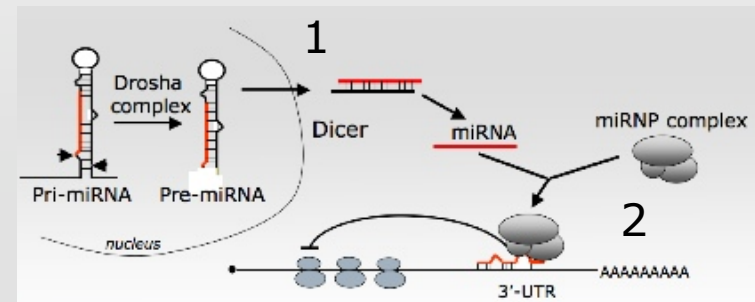


# Silencing machinery – Methods

- manually curated list of 52 gene products involved in RNA-mediated gene silencing

- 3 broad compartments

- 1. miR biogenesis: 4 (+4)
- 2. RISC/mRNP: 12 (+2)
- 3. P-bodies: 27 (+3)



- pDSPs altering machinery gene **sequence**
  - SNPs (non-synonymous, stop/frameshift, splicing site)
- pDSPs altering machinery gene product **concentration**
  - CNVs encompassing machinery genes (human, mouse, rat)
  - eQTL corresponding to machinery genes (human)

# Silencing machinery – Results

---

		<b>human</b>	<b>mouse</b>
	genes	52	51
SNPs in machinery genes	affected genes	49	35
	total	237	127
	non-synonymous	151	73
	stops / frameshifts	45	2
	splicing sites	42	52
machinery genes in CNVs	CNVs	17	0
	affected genes	17	0
machinery genes identified as eQTL		21	n.d.

---

# Conclusions

---

- other features of Patrocles
  - allelic imbalance plots (HapMap)
  - reported associations between pSNPs (or SNPs in miR genes) and phenotypes
  - Patrocles Finder for custom sequences
- most pSNPs are likely false positives due to poor specificity of target site predictions
  - Patrocles still contains some interesting biology (e.g., purifying selection on non-conserved sites)
  - systematic validation of pSNPs become possible (e.g., AgoIP + estimation of allelic imbalance in RISC-bound mRNAs; HITS-CLIP)