Exchange Among Sex Chromosomes and Autosomes - Human and Mouse

Yan Zhang

7/24/2013

Some summary points from my previous analyses

- In both human and mouse, co-residence is not random.
  - In human, Y is exceptionally significant.
- In human, pseudogene type is an important contributor determining co-residence.
  - DUP pgenes tend to reside on the same chromosome with the parents, compared to PSSD pgenes.
  - Transcribed/Not plays a weaker role.
- Exchange between sex chromosomes and autosomes
  - In both human and mouse, Y chromosome very seldom exports to other chromosomes.

Human: Exchange among X, Y and Autosomes (Hyp6)

As seen before:

Pseudogene (destination)

	Parent (source)		
	Autosomes	**X**	**Y**
Autosomes	7729	306	4
X	549	81	1
Y	105	33	87

- 3by3: Pearson's Chi-squared test & Fisher's Exact Test: p-value < 2.2e-16.
- 2by2 (X, Y only): Fisher's Exact Test: p-value < 2.2e-16.
  Sample estimat: odds ratio ≈ 208.
- Y chromosome very seldom exports to other chromosomes.

Mouse: Exchange among X, Y and Autosomes (Hyp6)

As seen before:

Pseudogene (destination)

	Parent (source)		
	Autosomes	**X**	**Y**
Autosomes	6798	219	1
X	622	168	1
Y	21	0	1

- 3by3: Pearson's Chi-squared test & Fisher's Exact Test: p-value < 2.2e-16.

- Y chromosome very seldom exports to other chromosomes.

Some observations in human

- 14 unique human parent TranscriptIDs on Y, trace back to genes, among which:
  - CDY1, USP9Y, RBMY1A1 and TSPY1 have high number of pseudogenes.
  - All four (highly) expressed in testis, i.e. germline tissue (based on RNA-seq Body Map).

➔Correlation between parent gene expression level and number of pseudogenes.

More detailed investigation:

- Other genes with low Freq.ParentTranscriptID have not been checked.
- PSSD and DUP are not separated

Some observations in human (ctd)

- Comparison with literature [1]:
  - "For any newly arisen pseudogene to be heritable, **it has to originate in the germline, or embryonic stem cells that will give rise to the germline**. Thus, genes that are expressed only in somatic cells do not have processed pseudogenes."
    - Consistency. We observed genes with high number of pgenes highly expressed in germline.
  - "Highly expressed genes, such as house-keeping genes, have a higher probability of retroposition."
    - GO term investigated. May go on to look at whether there are house-keeping genes in the list.
  - "The observation that **genes expressed at high levels in the germline generate the majority of processed pseudogenes** prompted the hypothesis that the level of retroposition correlates positively with the amount of expression in the germline. Only recently has this correlation been verified (Podlaha and Zhang, 2009, Trends in Genetics)"
    - Worth looking at DUP and PSSD separately.

Step aside, read and think

- Extensive Gene Traffic on the Mammalian X Chromosome.

  J. J. Emerson et al. Science. 2004.

Rough summary points from their work

- "The Y chromosome has been shown to recruit male-specific ganes, **whereas a few individual X-linked genes have male specific duplicate counterparts on autosomes**." "**Some male-specific genes appear to be enriched on the X chromosomes.**"
  - Surprising. Q: GO terms of parents and pgenes on X?
- Traffic between X and autosomes. Y seems not a focus in this paper.

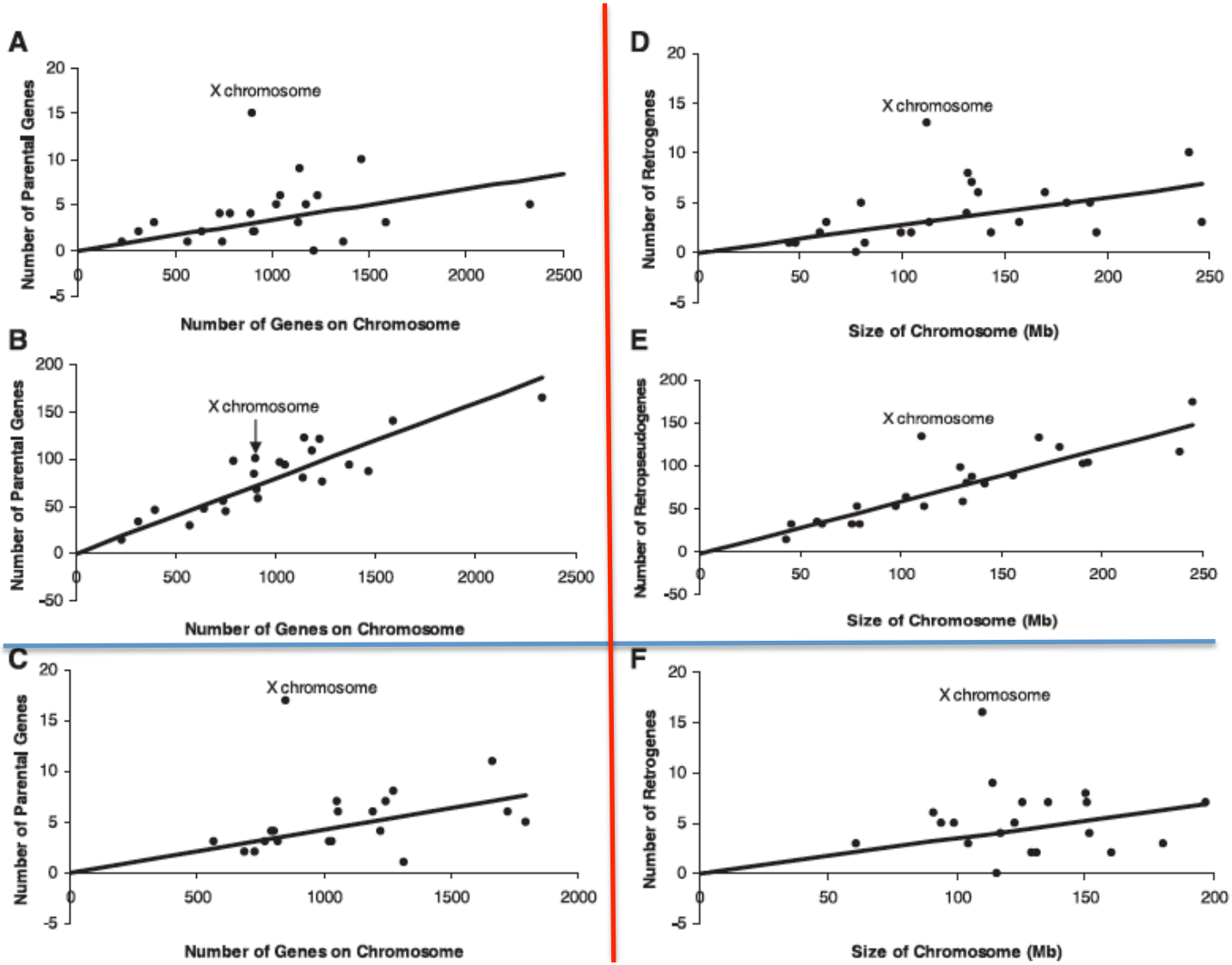
Rough summary points from their work (ctd)

- **Exclude retroposed pairs that share chromosome-linkage.**
  - Thus, they did not look at co-residence on chromosomes or co-localization in bins (threshold D).
  - Datasets generated:
    - 94 retrogene-parent pairs in human
    - 105 retrogene-parent pairs in mouse
    - 1859 retropseudogene(PSSD)-parent pairs in human
    - DUP-parent pairs not included. Their insertion pattern might not be random.
- Analysis methods:
  - Linear regression
  - Multinomial Monte Carlo resampling test

Rough summary from their work (ctd)

- Linear regression results (Y was not included):
  - #Genes or size of chromosomes vs. #Parents/Retroposed

Fig. 1. Regressions for the parental genes of retrogenes in (A) human and (C) mouse and for the parental genes of retropseudogenes (B) in human. Regressions for the size of a chromosome in (D) human and (F) mouse and for the retropseudogenes (E) in human. In the plots, X is shown as 75% of its size as predicted by the model (*15*), although allowing X to assume 100% of its size does not change the results. Probabilities for the hypothesis that the chromosome with the highest observed/expected ratio [where the expected number is calculated as in (*16*)] is an outlier are calculated using Grubbs and Dixon outlier tests (*16*). For every distribution [except (B)], the X has the largest ratio and is an outlier with $P < 0.005$ and $P < 0.01$ for the Grubbs and Dixon tests, respectively; (B) shows no such outliers.



Rough summary from their work (ctd)

- Multinomial Monte Carlo resampling test
  - Null assumptions:
    - #retrogenes
      proportional to #genes on a chromosome
    - Chromosome accept #retroposed copied
      proportional to size of the chromomosome
  - Then, random sample same total number of pairs
    ➔Expected number of retrocopies
  - Finally, significance tests.

Rough summary from their work (ctd)

- Multinomial Monte Carlo resampling test results:

Only between X and autosome, and exclude co-residence.

Table 1. Multinomial Monte Carlo resampling test of the difference between observation and the expected frequencies. X→ and →X refer to the predictions and observations for genes leaving and entering the X, respectively; A→ and →A predictions and observations for genes leaving and entering an autosome, respectively. The *P* values were calculated from Monte Carlo simulations (*16*) by analyzing the movements involving both autosomes and the X chromosome. Excess = (Observed – Expected)/Expected.

Direction of retroposition	Expected	Observed	Excess (%)	*P*
Human				
X→	3.76	15	299	0.00012
A→	90.24	79	–12	
→X	3.61	13	260	0.00244
→A	90.39	81	–10	
Mouse				
X→	4.16	17	309	0.00011
A→	100.84	88	–13	
→X	4.62	16	246	0.00015
→A	100.38	89	–11	