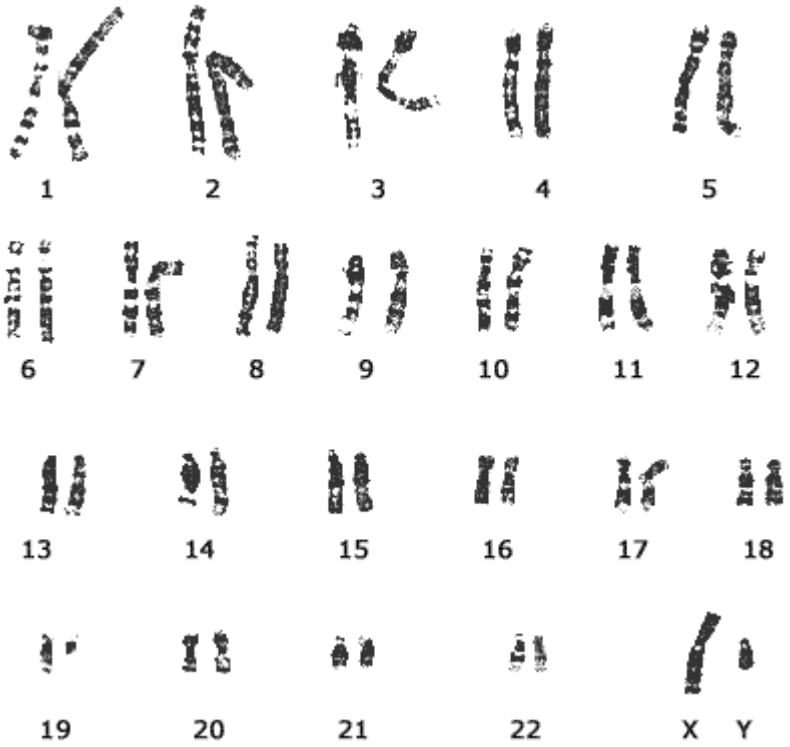
Robert Yu
Statistical Analyst

Analysis of Genetic Data on Human Chromosome X

1. Biology of Chromosomes
    1) 23 pairs of chromosomes and sex chromosome
    2) Cell division and chromosome split
    3) DNA on chromosome
    4) Gene expression
2. Chromosomal Data for Statistical Genetic Study
    1) Genetic markers and their info
    2) Statistical Analyses of the genetic information
3. The Case of Chromosome X
    1) Purpose of studying chromosome X
    2) Difference in males and females
    3) Dosage compensation
    4) Challenges

Successful statistical modeling relies on understanding underlying events.

Chromosome – the basis for DNA

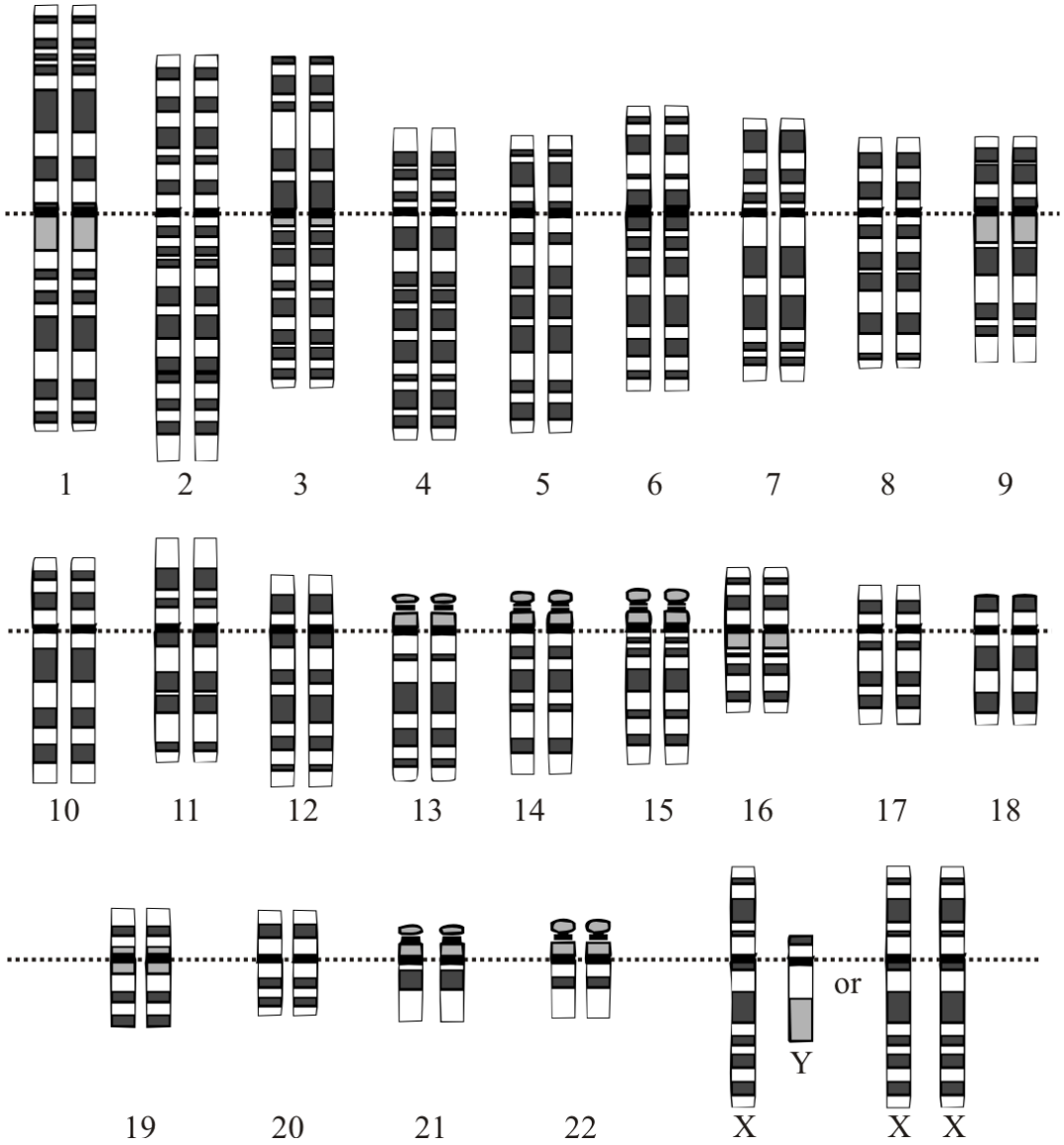


In humans, a cell nucleus contains 46 individual chromosomes in 23 pairs.

Half of these chromosomes come from one father and another half come from the mother.

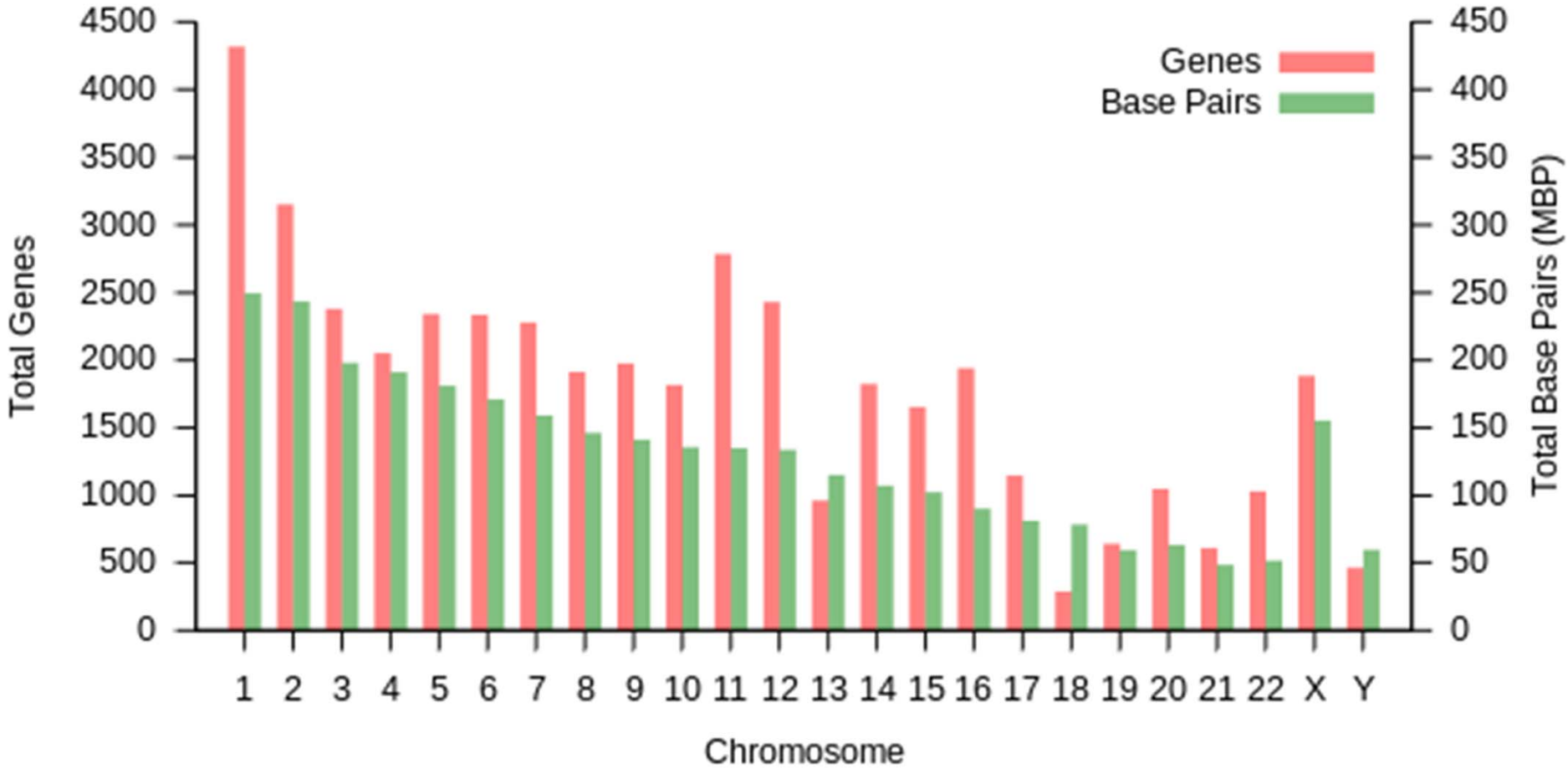
Source: “Genes and How They Work” @ http://www.dnaletters.com/bgenes.html

Schematic Representation

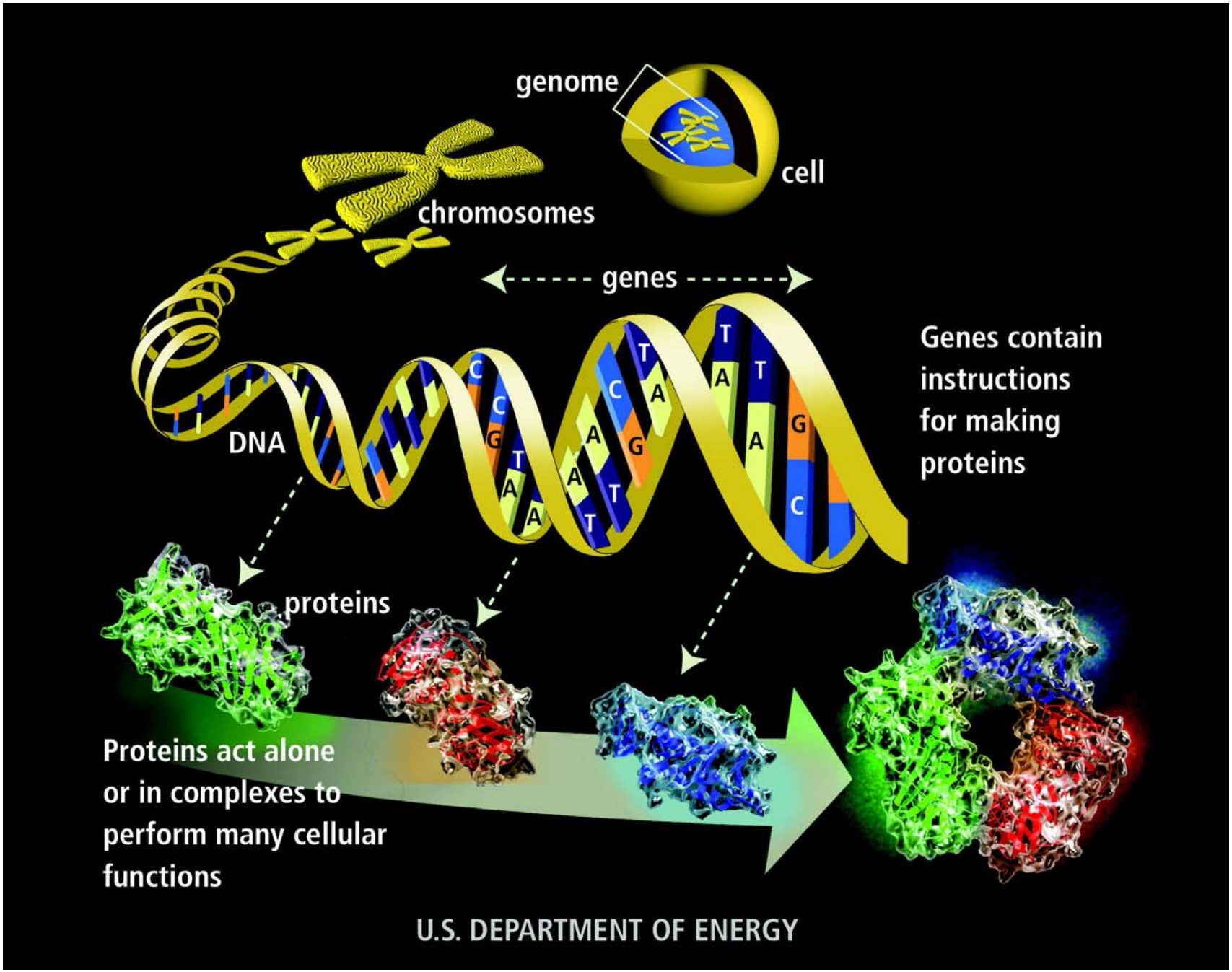


Source: “Human Karyotype” @ http://www.biologycorner.com/APbiology/inheritance/10-1_meiosis.html

Distribution of Genes and DNA Base Pairs

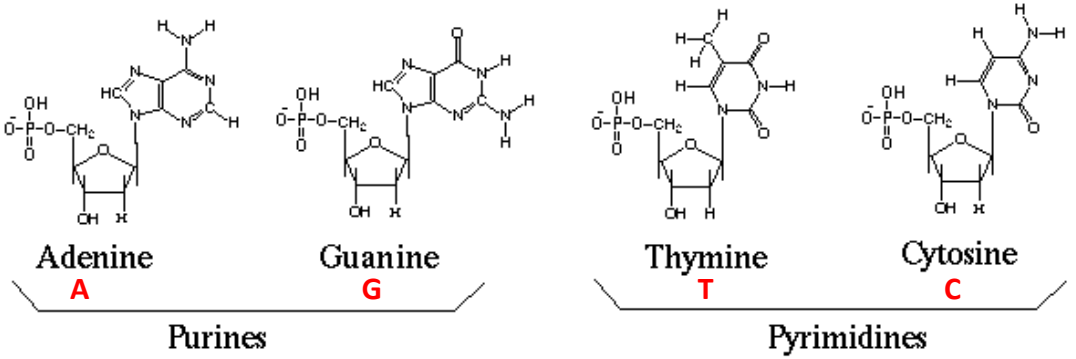


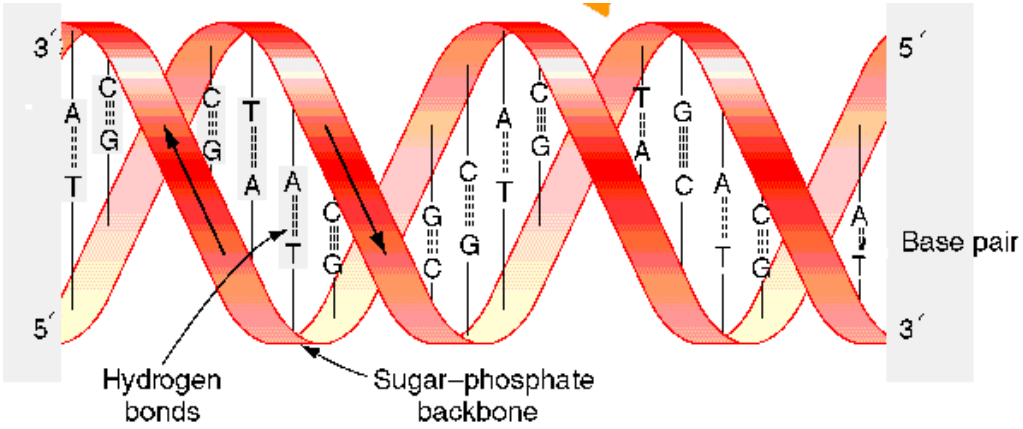
Source: “Chromosome” – Wikipedia @ http://en.wikipedia.org/wiki/Chromosome

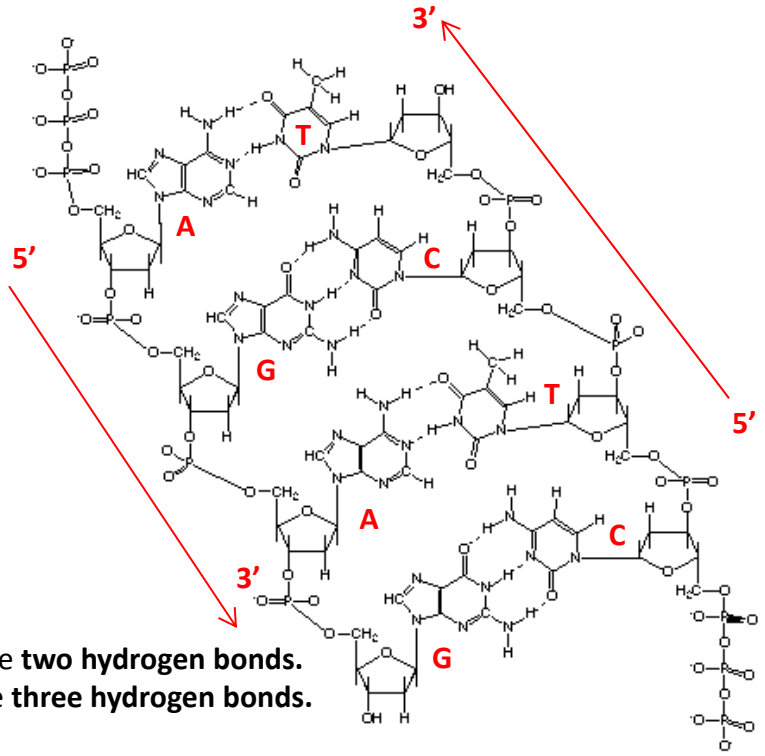

genome
cell
chromosomes
genes
DNA
Genes contain instructions for making proteins
proteins
Proteins act alone or in complexes to perform many cellular functions
U.S. DEPARTMENT OF ENERGY

The Nucleotides of DNA

a nitrogenous base, a sugar and a phosphate







Adenine pairs with **Thymine** because they make **two hydrogen bonds.**
Guanine pairs with **Cytosine** because they make **three hydrogen bonds.**

Genome Variations – Mutation or Polymorphism

- **Polymorphism** is a DNA variation in which each possible sequence is present in at least 1 percent of people.

  For example, a place in the genome where 93 percent of people have a T and the remaining 7 percent have an A is a polymorphism.
- If one of the possible sequences is present in less than 1 percent of people (*99.9 percent of people have a G and 0.1 percent have a C*), then the variation is called a **mutation**.
- Many polymorphisms actually do affect a person's characteristics, though in more complex and sometimes unexpected ways.
- About 90 percent of human genome variation comes in the form of **single nucleotide polymorphisms**, or **SNP**s
- The human genome contains more than 2 million SNPs (some estimate over 10 million SNPs, i.e. a SNP/300 bps).

Source: "**SNPs: VARIATIONS ON A THEME**" – NCBI: A Science Primer @ http://www.ncbi.nlm.nih.gov/About/primer/snps.html

Chromosome	Genes	Total base pairs	Sequenced base pairs
1	4,220	247,199,719	224,999,719
2	1,491	242,751,149	237,712,649
3	1,550	199,446,827	194,704,827
4	446	191,263,063	187,297,063
5	609	180,837,866	177,702,766
6	2,281	170,896,993	167,273,993
7	2,135	158,821,424	154,952,424
8	1,106	146,274,826	142,612,826
9	1,920	140,442,298	120,312,298
10	1,793	135,374,737	131,624,737
11	379	134,452,384	131,130,853
12	1,430	132,289,534	130,303,534
13	924	114,127,980	95,559,980
14	1,347	106,360,585	88,290,585
15	921	100,338,915	81,341,915
16	909	88,822,254	78,884,754
17	1,672	78,654,742	77,800,220
18	519	76,117,153	74,656,155
19	1,555	63,806,651	55,785,651
20	1,008	62,435,965	59,505,254
21	578	46,944,323	34,171,998
22	1,092	49,528,953	34,893,953
X (sex chromosome)	1,846	154,913,754	151,058,754
Y (sex chromosome)	454	57,741,652	25,121,652
Total	32,185	3,079,843,747	2,857,698,560

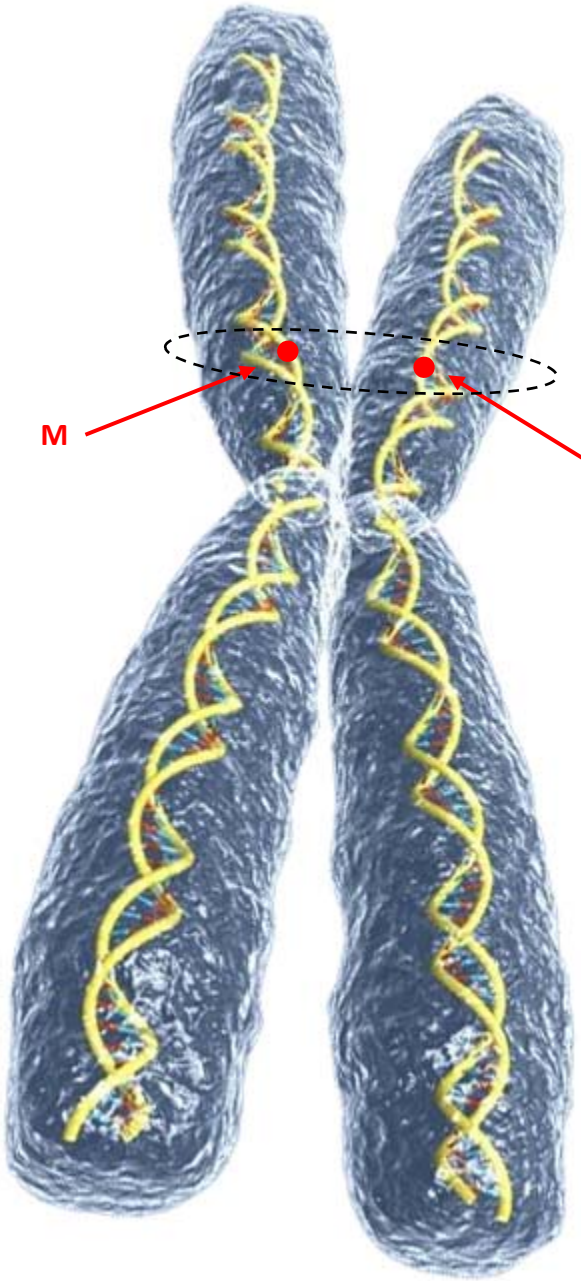
Source: “Chromosome” – Wikipedia @ http://en.wikipedia.org/wiki/Chromosome

- ~ 0.2-1% base pairs are SNPs

SNPs and Disease

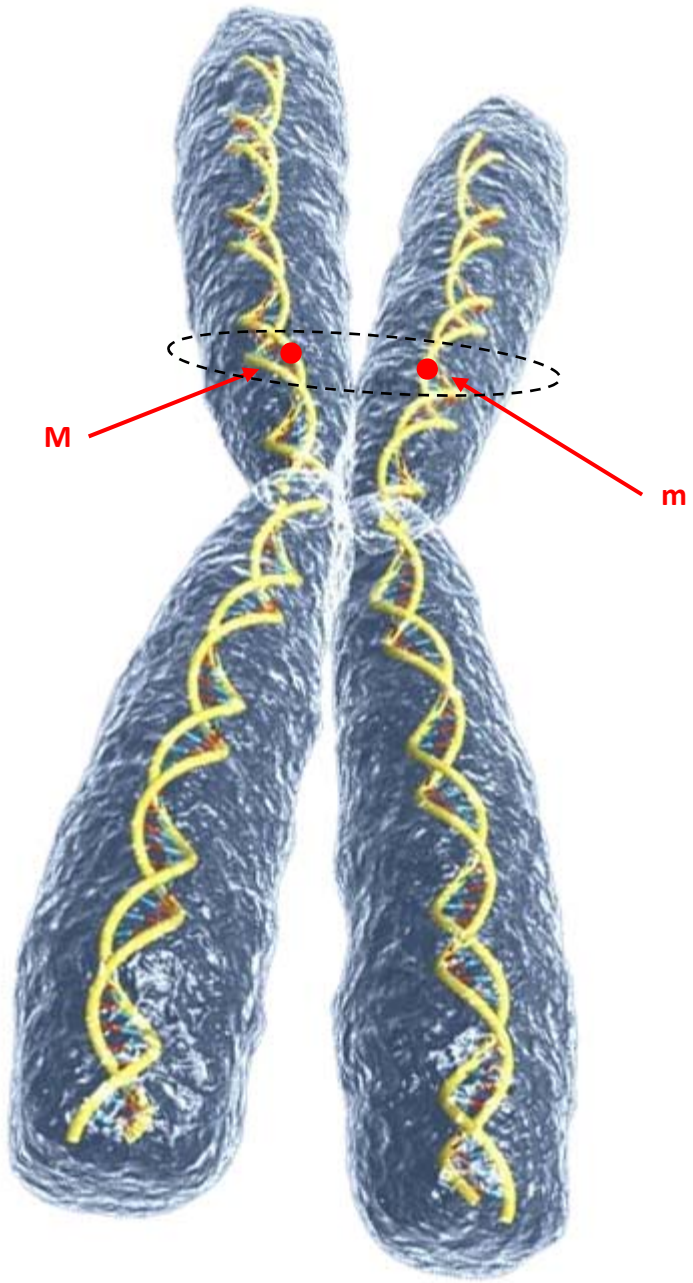
- Many common diseases in humans are not caused by a genetic variation within a single gene but are influenced by complex interactions among multiple genes as well as environmental and lifestyle factors.
- a person's **genetic predisposition**, or the potential of an individual to develop a disease based on genes and hereditary factors.
- genetic factors also affect a person's response to drug therapy, **DNA polymorphisms** such as SNPs will be useful in helping researchers determine and understand why individuals differ in their abilities to absorb or clear certain drugs, etc.

Source: "**SNPs: VARIATIONS ON A THEME**" – NCBI: A Science Primer @ http://www.ncbi.nlm.nih.gov/About/primer/snps.html



a locus with 2 alleles, *M* and/or *m*

- A pattern of the 2 alleles of the same locus is called genotype
- In the case of *MM* or *mm*, the genotype is termed homozygous, and *Mm* heterozygous
- The genotype is often used to test the correlation with phenotype



- An implicit assumption is that
  - The version of a gene represented by the one allele on one chromosome is as active as the other, i.e. *M* and *m* have the same biological activity.
- In the case of *MM* or *mm*, the genes are expressing in the *same* amount from each chromosome, *double dosage*.
- In the case of *Mm*, both *M* and *m* contribute the *same* amount of its protein to the cell.
- If the *influence* of *M* overplays the one from *m*, *Mm* is considered showing the similar biology effect as *MM*. Thus, *M* is seen as dominant (over *m*).
- If *MM* produces double dosage comparing with *Mm* does, then *M* is considered as behaving in additive mode.
- If an effect of *M* can only be shown when in *MM* genotype, then *M* is seen as recessive.

Genotype Coding for Statistical Analyses is based on biological phenomena

Genotype	Dominant	Recessive	Additive
	(assuming *M* is the affecting allele)		
MM	1	1	2
Mm	1	0	1
Mm	0	0	0

Note: *Additive* model reflects a dosage effect of the underlying gene and sometimes also termed as *codominant* model.

Comment: if two alleles behave differently, i.e. having unequal activity, from the two chromosomes, the resulting biological effect will be different, and thus the respective statistical model should be adjusted accordingly. Later, this will be discussed in the case of X chromosome.

Examples of Statistical Analysis Methods

	Cases	Controls	Total
MM	n_{2A}	n_{2U}	$n_{2\cdot}$
Mm	n_{1A}	n_{1U}	$n_{1\cdot}$
mm	n_{0A}	n_{0U}	$n_{0\cdot}$
Total	$n_{\cdot A}$	$n_{\cdot U}$	$n_{\cdot\cdot}$

Pearson's χ^2 Test (2-df)

$$X^2 = \sum_{i=0,1,2} \sum_{j=A,U} \frac{(n_{ij} - E[n_{ij}])^2}{E[n_{ij}]}$$

where

$$E[n_{ij}] = \frac{n_{i\cdot} n_{\cdot j}}{n_{\cdot\cdot}}$$

X^2 has χ^2 distribution with 2 degrees of freedom under null hypothesis.

Cochran-Armitage Trend Test (1-df)

$$X^2 = \frac{\left[\left(p_{2A} + \frac{1}{2}p_{1A}\right) - \left(p_{2U} + \frac{1}{2}p_{1U}\right)\right]^2}{\left(\frac{1}{n_{\cdot A}} + \frac{1}{n_{\cdot U}}\right)\left(\frac{1}{n_{\cdot\cdot}^2}\right)\left[n_{\cdot\cdot}\left(\frac{1}{4}n_{1\cdot} + n_{2\cdot}\right) - \left(\frac{1}{2}n_{1\cdot} + n_{2\cdot}\right)^2\right]}$$

where

$$p_{ij} = \frac{n_{ij}}{n_{\cdot j}}$$

- X^2 has χ^2 distribution with 1 degree of freedom under null hypothesis.

A Case of Association Studies of X Chromosome

Genetic variation in *PCDH11X* is associated with susceptibility to late-onset Alzheimer's disease

Minerva M Carrasquillo[1], Fanggeng Zou[1], V Shane Pankratz[2], Samantha L Wilcox[1], Li Ma[1], Louise P Walker[1], Samuel G Younkin[1], Curtis S Younkin[1], Linda H Younkin[1], Gina D Bisceglio[1], Nilufer Ertekin-Taner[1,3], Julia E Crook[4], Dennis W Dickson[1], Ronald C Petersen[3,5], Neill R Graff-Radford[1,3] & Steven G Younkin[1]

By analyzing late-onset Alzheimer's disease (LOAD) in a genome-wide association study (313,504 SNPs, three series, 844 cases and 1,255 controls) and evaluating the 25 SNPs with the most significant allelic association in four additional series (1,547 cases and 1,209 controls), we identified a SNP (rs5984894) on Xq21.3 in *PCDH11X* that is strongly associated with LOAD in individuals of European descent from the United States. Analysis of rs5984894 by multivariable logistic regression adjusted for sex gave global *P* values of 5.7×10^{-5} in stage 1, 4.8×10^{-6} in stage 2 and 3.9×10^{-12} in the combined data. Odds ratios were 1.75 (95% CI = 1.42–2.16) for female homozygotes ($P = 2.0 \times 10^{-7}$) and 1.26 (95% CI = 1.05–1.51) for female heterozygotes ($P = 0.01$) compared to female noncarriers. For male hemizygotes ($P = 0.07$) compared to male noncarriers, the odds ratio was 1.18 (95% CI = 0.99–1.41).

Late-onset Alzheimer's disease (LOAD) is a neurodegenerative disease characterized by large numbers of senile plaques and neurofibrillary tangles in the brain. LOAD is the most common cause of dementia in the elderly, affecting approximately 10% of those aged 65 years or older[1]. Multiple rare mutations in the genes encoding the amyloid β protein precursor, presenilin 1 and presenilin 2 cause an early-onset

353 cases, 331 controls) and Rochester, Minnesota (RS: 245 cases, 701 controls). A third, autopsy-confirmed series (AUT: 246 cases, 223 controls) was assembled from the Mayo brain bank (**Supplementary Methods**). In stage 1, SNPs were tested for allelic association using the χ^2 test implemented in PLINK[6] (**Supplementary Table 3** online). After adjustment for population stratification using EIGENSTRAT[7] and Bonferroni correction for the 313,504 SNPs tested, only six *APOE*-linked SNPs showed genome-wide significance in stage 1 (**Supplementary Tables 3** and **4** and **Supplementary Methods** online).

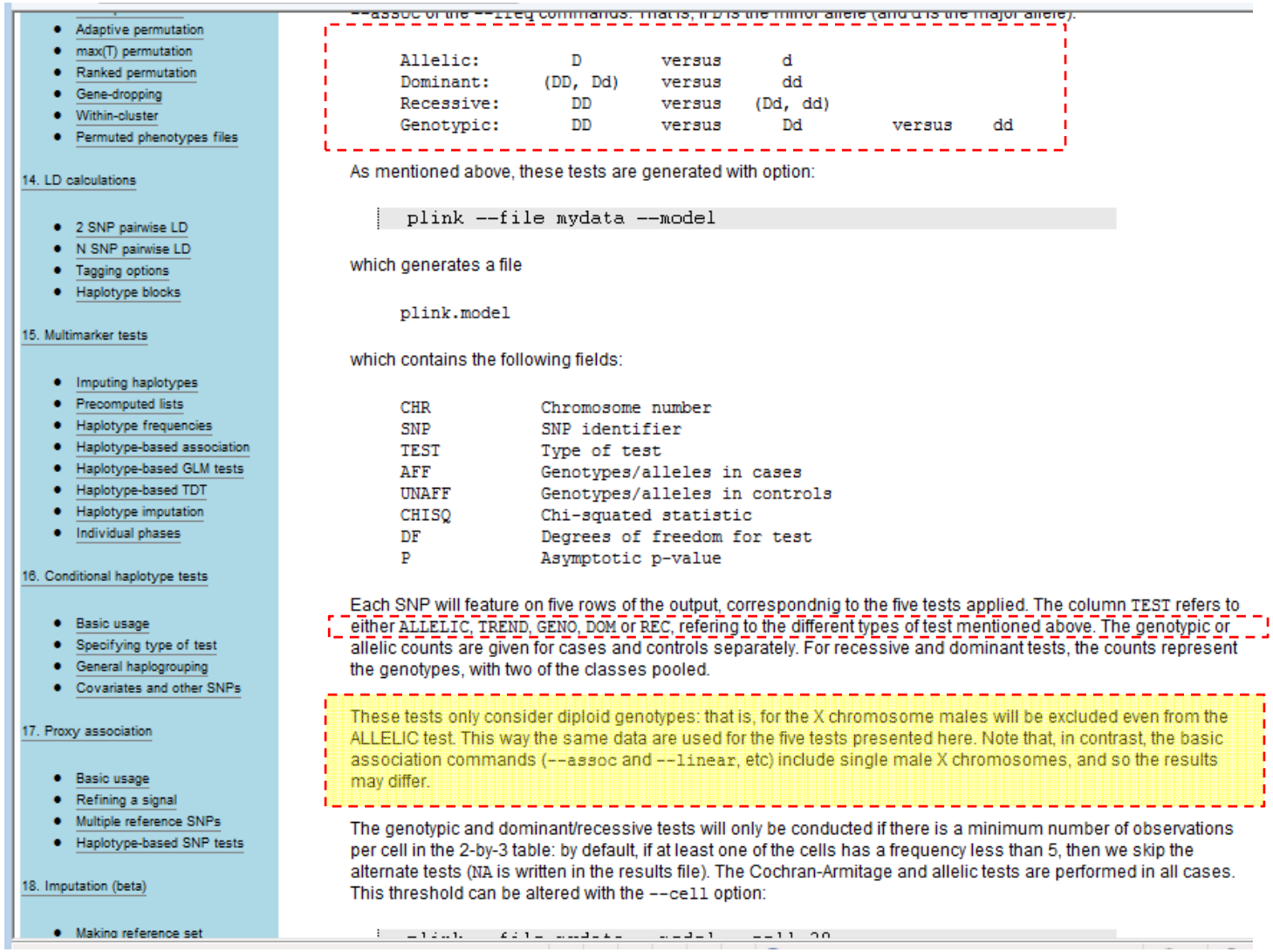
In stage 2, we used SEQUENOM iPLEX technology to genotype the 25 SNPs with the most significant association in stage 1 (**Supplementary Table 3**) in an additional 845 cases and 1,000 controls drawn from the same three series but with ages at diagnosis of over 80 years (JS: 237 cases, 260 controls; RS: 276 cases, 624 controls; AUT: 332 cases, 116 controls) and from a National Cell Repository for Alzheimer's disease series of 702 cases (1 per family) and 209 controls with ages at diagnosis of over 60 years. The top 25 SNPs tested in stage 2 included 10 SNPs in the *APOE* region of chromosome 19 and 15 SNPs on other chromosomes. The allelic association results for these 25 SNPs in stages 1, 2 and 1 + 2 are shown in **Supplementary Table 3**. All ten SNPs in the *APOE* region showed significant association in stage 2,

Statistical analyses. Genotype reports produced by Illumina BeadStudio 2.0 software (stage 1 data) or SEQUENOM Typer 4.0 software (stage 2 data) were used to generate lgen, map and fam files that were imported into PLINK. We analyzed the SNP genotypes in stage 1, stage 2 and the combined datasets for allelic association with Alzheimer's disease using the allelic association χ^2 test implemented in PLINK. With $\alpha = 0.05$ and Bonferroni correction for the 313,504 SNPs tested in stage 1, a P value of 1.6×10^{-7} is required for 'genome-wide' significance. Using this criterion, the only SNPs to achieve genome-wide significance in stage 1 were six *APOE*-linked SNPs. In the combined stage 1 + 2 data, rs5984894 and one additional *APOE*-linked SNP also achieved this level of significance (**Supplementary Table 3**).

As rs5984894 showed highly significant association with LOAD in stages 1, 2 and 1 + 2, and as this previously unknown LOAD SNP was analyzed in seven distinct case-control series, we analyzed it in PLINK using not only the χ^2 test on combined allele counts but also the Mantel-Haenszel method in which the Breslow-Day option was used to test for series to series heterogeneity (**Table 1**). We also analyzed rs5984894 by multivariable logistic regression (**Table 2**), as described in the **Supplementary Methods**.

Evaluation of rs5984894 by multivariable logistic regression. In our initial analyses, we used logistic regression with male sex as a covariate to obtain ORs for male hemizygotes, female heterozygotes, and female homozygotes, and test them for significance. In this model, the ORs for female homozygotes and female heterozygotes compare these groups to female non-carriers and the coefficient for male hemizygotes compares this group to male non-carriers (Supplementary Table 5, model 1). Sex was not a significant covariate in the combined stage I, II, or I + II datasets. Overall in stage I, the male hemizygotes (P=0.04), female heterozygotes (P=0.02), and female homozygotes (P=0.0002) had ORs of 1.33 (95% CI 1.02-1.74), 1.43 (95% CI 1.06-1.92) and 1.92 (95% CI 1.36-2.70) respectively with a global P value of 5.7×10^{-5}. Overall in stage II, the male hemizygotes (P=0.74), female heterozygotes (P=0.15) and female homozygotes (P=0.0002) had ORs of 1.04 (95% CI 0.82-1.33), 1.19 (95% CI 0.94-1.50) and 1.70 (95% CI 1.29-2.24) respectively with a global P value of 4.8×10^{-6}. When all data in stage I and II were combined, the male hemizygotes (P=0.07), female heterozygotes (P=0.01), and female homozygotes ($P=2.0 \times 10^{-7}$) had ORs of 1.18 (95% CI 0.99-1.41), 1.26 (95% CI 1.05-1.51), and 1.75 (95% CI 1.42-2.16) respectively with a global P value of 3.9×10^{-12}.

PLINK approach



- Adaptive permutation
- max(T) permutation
- Ranked permutation
- Gene-dropping
- Within-cluster
- Permuted phenotypes files

14. LD calculations

- 2 SNP pairwise LD
- N SNP pairwise LD
- Tagging options
- Haplotype blocks

15. Multimarker tests

- Imputing haplotypes
- Precomputed lists
- Haplotype frequencies
- Haplotype-based association
- Haplotype-based GLM tests
- Haplotype-based TDT
- Haplotype imputation
- Individual phases

16. Conditional haplotype tests

- Basic usage
- Specifying type of test
- General haplogrouping
- Covariates and other SNPs

17. Proxy association

- Basic usage
- Refining a signal
- Multiple reference SNPs
- Haplotype-based SNP tests

18. Imputation (beta)

- Making reference set

--assoc or the --freq commands. That is, if D is the minor allele (and d is the major allele).

```
Allelic:         D        versus       d
Dominant:     (DD, Dd)    versus       dd
Recessive:       DD       versus    (Dd, dd)
Genotypic:       DD       versus       Dd        versus    dd
```

As mentioned above, these tests are generated with option:

```
plink --file mydata --model
```

which generates a file

```
plink.model
```

which contains the following fields:

```
CHR          Chromosome number
SNP          SNP identifier
TEST         Type of test
AFF          Genotypes/alleles in cases
UNAFF        Genotypes/alleles in controls
CHISQ        Chi-squated statistic
DF           Degrees of freedom for test
P            Asymptotic p-value
```

Each SNP will feature on five rows of the output, correspondnig to the five tests applied. The column `TEST` refers to either `ALLELIC`, `TREND`, `GENO`, `DOM` or `REC`, refering to the different types of test mentioned above. The genotypic or allelic counts are given for cases and controls separately. For recessive and dominant tests, the counts represent the genotypes, with two of the classes pooled.

These tests only consider diploid genotypes: that is, for the X chromosome males will be excluded even from the ALLELIC test. This way the same data are used for the five tests presented here. Note that, in contrast, the basic association commands (`--assoc` and `--linear`, etc) include single male X chromosomes, and so the results may differ.

The genotypic and dominant/recessive tests will only be conducted if there is a minimum number of observations per cell in the 2-by-3 table: by default, if at least one of the cells has a frequency less than 5, then we skip the alternate tests (`NA` is written in the results file). The Cochran-Armitage and allelic tests are performed in all cases. This threshold can be altered with the `--cell` option:

```
plink --file mydata --model --cell 20
```

PLINK approach

How does PLINK handle the X chromosome in association tests?

By default, in the linear and logistic (`--linear`, `--logistic`) models, for alleles `A` and `B`, males are coded

```
A   ->   0
B   ->   1
```

and females are coded

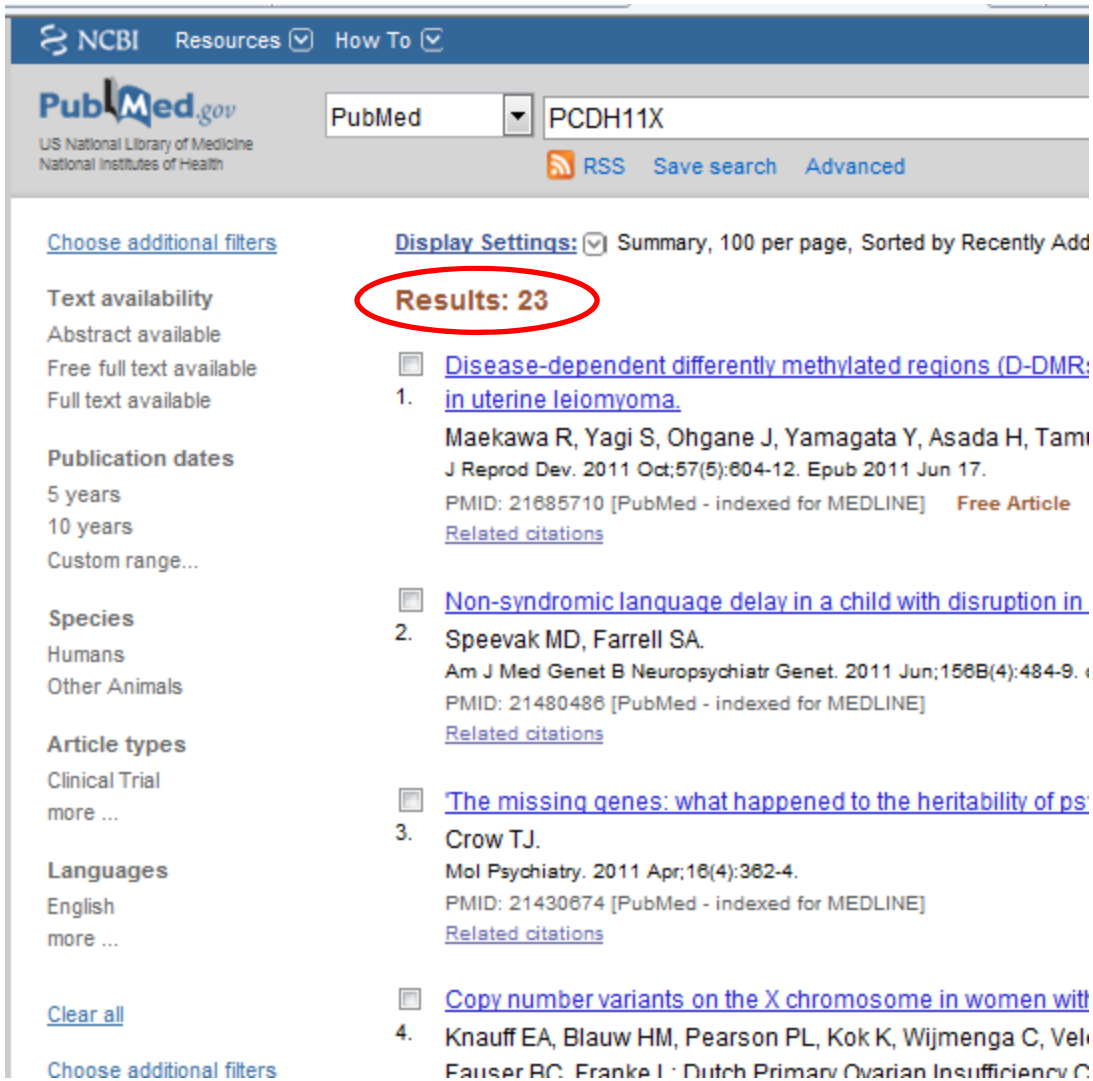
```
AA  ->   0
AB  ->   1
BB  ->   2
```

and additionally sex (0=male,1=female) is also automatically included as a covariate. It is therefore important not to include sex as a separate covariate in a covariate file ever, but rather to use the special `--sex` command that tells PLINK to add sex as coded in the PED/FAM file as the covariate (in this way, it is not double entered for X chromosome markers). If the sample is all female or all male, PLINK will know not to add sex as an additional covariate for X chromosome markers.

The basic association tests that are allelic (`--assoc`, `--mh`, etc) do not need any special changes for X chromosome markers: the above only applies to the linear and logistic models where the individual, not the allele, is the unit of analysis. Similarly, the TDT remains unchanged. For the `--model` test and Hardy-Weinberg calculations, male X chromosome genotypes are excluded.

Not all analyses currently handle X chromosomes markers (for example, LD pruning, epistasis, IBS calculations) but support will be added in future.

A Case of Association Studies of X Chromosome



5. Lack of association between protocadherin 11-X/Y (**PCDH11X** and PCDH11Y) polymorphisms and late onset Alzheimer's disease.
Miar A, Alvarez V, Corao AI, Alonso B, Díaz M, Menéndez M, Martínez E.
Brain Res. 2011 Apr 6;1383:252-6. Epub 2011 Jan 27.
PMID: 21276771 [PubMed - indexed for MEDLINE]
Related citations

6. Lack of association between **PCDH11X** genetic variation and late-onset Chinese population.
Wu ZC, Yu JT, Wang ND, Yu NN, Zhang Q, Chen W, Zhang W, Zhu QX
Brain Res. 2010 Oct 21;1357:152-6. Epub 2010 Aug 11.
PMID: 20707987 [PubMed - indexed for MEDLINE]
Related citations

7. Failure to replicate an association of rs5984894 SNP in the **PCDH11X** g Alzheimer's disease affected patients.
Lescai F, Pirazzini C, D'Agostino G, Santoro A, Ghidoni R, Benussi L, Marchegiani F, Cardelli M, Olivieri F, Nacmias B, Sorbi S, Bagnoli S, Ta Boneschi F, Binetti G, Forloni G, Quadri P, Scarpini E, Franceschi C.
J Alzheimers Dis. 2010;21(2):385-8.
PMID: 20555150 [PubMed - indexed for MEDLINE]
Related citations

8. **PCDH11X** variation is not associated with late-onset Alzheimer disease
Beecham GW, Naj AC, Gilbert JR, Haines JL, Buxbaum JD, Pericak-Va
Psychiatr Genet. 2010 Dec;20(6):321-4.
PMID: 20523261 [PubMed - indexed for MEDLINE] Free PMC Article
Related citations

9. Quantitative analysis of alternative transcripts of human **PCDH11X**/Y ge
Ahn K, Huh JW, Kim DS, Ha HS, Kim YJ, Lee JR, Kim HS.
Am J Med Genet B Neuropsychiatr Genet. 2010 Apr 5;153B(3):736-44.
PMID: 19859901 [PubMed - indexed for MEDLINE]
Related citations

10. Genetic variation in **PCDH11X** is associated with susceptibility to late-o
Carrasquillo MM, Zou F, Pankratz VS, Wilcox SL, Ma L, Walker LP, Yo LH, Bisceglio GD, Ertekin-Taner N, Crook JE, Dickson DW, Petersen R SG.
Nat Genet. 2009 Feb;41(2):192-8. Epub 2009 Jan 11.
PMID: 19136949 [PubMed - indexed for MEDLINE] Free PMC Article
Related citations

Related citations

17. Inactivation status of **PCDH11X**: sexual dimorphisms in gene expression levels in brain.
Lopes AM, Ross N, Close J, Dagnall A, Amorim A, Crow TJ.
Hum Genet. 2006 Apr;119(3):267-75. Epub 2006 Jan 20.
PMID: 16425037 [PubMed - indexed for MEDLINE]
Related citations

18. Expression and genetic variability of PCDH11Y, a gene specific to Homo sapiens and candidate for susceptibility to psychiatric disorders.
Durand CM, Kappeler C, Betancur C, Delorme R, Quach H, Goubran-Botros H, Melke J, Nygren G, Chabane N, Bellivier F, Szoke A, Schurhoff F, Rastam M, Anckarsäter H, Gillberg C, Leboyer M, Bourgeron T.
Am J Med Genet B Neuropsychiatr Genet. 2006 Jan 5;141B(1):67-70.
PMID: 16331680 [PubMed - indexed for MEDLINE]
Related citations

19. Epigenetic silencing of the protocadherin family member PCDH-gamma-A11 in astrocytomas.
Waha A, Güntner S, Huang TH, Yan PS, Arslan B, Pietsch T, Wiestler OD, Waha A.
Neoplasia. 2005 Mar;7(3):193-9.
PMID: 15799819 [PubMed - indexed for MEDLINE] Free PMC Article
Related citations

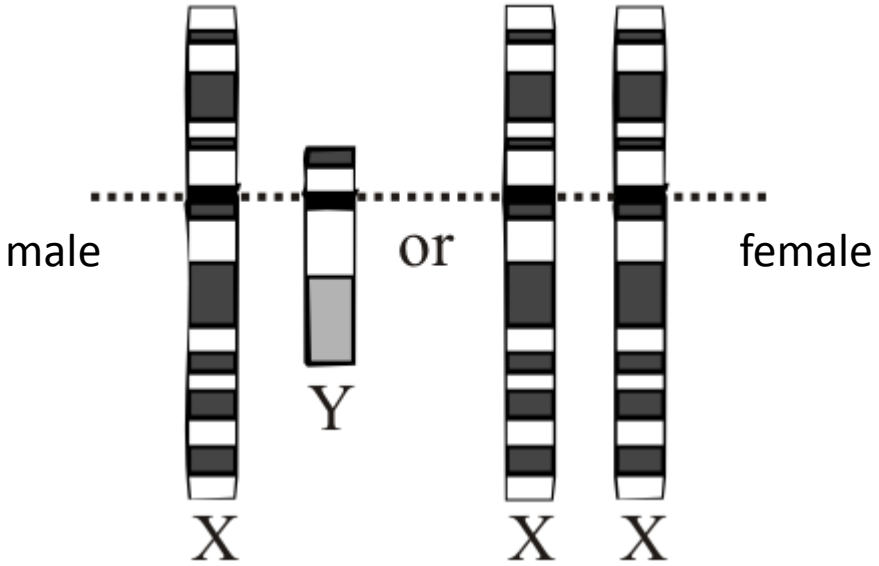
20. A comparative analysis of the pig, mouse, and human PCDHX genes.
Blanco-Arias P, Sargent CA, Affara NA.
Mamm Genome. 2004 Apr;15(4):296-306.
PMID: 15112107 [PubMed - indexed for MEDLINE]
Related citations

21. Protocadherin X (PCDHX) and Y (PCDHY) genes: multiple mRNA isoforms encoding variant signal peptides and cytoplasmic domains.
Blanco-Arias P, Sargent CA, Affara NA.
Mamm Genome. 2004 Jan;15(1):41-52.
PMID: 14727141 [PubMed - indexed for MEDLINE]
Related citations

22. Conservation of PCDHX in mammals; expression of human X/Y genes predominantly in brain.
Blanco P, Sargent CA, Boucher CA, Mitchell M, Affara NA.
Mamm Genome. 2000 Oct;11(10):906-14.
PMID: 11003707 [PubMed - indexed for MEDLINE]
Related citations

X Chromosome

- XY in males and XX in females
- In terms of X chromosome, at each locus, males have one allele but females have two alleles



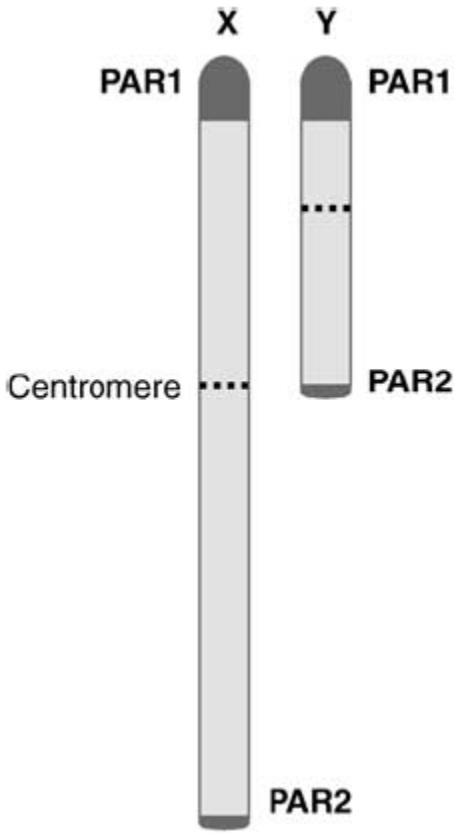
Sex	Genotype
Male	*M* or *m*
Female	*MM*, *Mm* or *mm*

Complexity in X Chromosome

- Pseudoautosomal Regions (PARs) on X Chromosome
  - Homologus DNA sequence on X and Y chromosomes in male, and on both X's in female. But they may behave differently between sexes.
- X Chromosome Inactivation (XCI) in female
  - Dosage compensation by *supressing* on X chromosome in females to equalize the dosage effect in males.
- Skewness of XCI
  - Unequal distribution of the inactivated X chromosome (either maternal or paternal origin) in females may play a key role in disease susceptibility.
- Escaping XCI
  - Active expression of gene(s) on an inactivated X chromosome in females provides different genetic model when analyses of genetic data of mixed sexes.
  - Escaped genes may or may not show equal activity comparing with its counterpart on an activated chromosome. This poses another layer of challenges in statistical modeling

Pseudoautosomal Region (PAR)

- The pseudoautosomal regions (PAR1 and PAR2) are short regions of homology between the mammalian X and Y chromosomes. The PAR behave like an autosome and recombine during meiosis. Thus genes in this region are inherited in an autosomal rather than a strictly sex-linked fashion.
- Their physical lengths are approximately 2.7Mb for PAR1 and 0.33Mb for PAR2.
- The recombination activity in PAR1 is extremely different between sexes.
- To date (2005-08), 24 genes have been reported in PAR1 and 5 genes in PAR2.
- Crossover activity in PAR1 is much higher in men than in women and also higher than for each of the autosomes.
- The rate of recombination in PAR2 is much lower than in PAR1 but still higher than the average rate of the remainder of the X chromosome.
- Genes in PARs express in the *same dosage* in both sexes.



Source:
- A. Helena Mangs and Brian J. Morris "The Human Pseudoautosomal Region (PAR): Origin, Function and Future" Current Genomics, 2007, 8, 129-136
- Antonia Flaquer, et al. "The human pseudoautosomal regions: a review for genetic epidemiologists", European Journal of Human Genetics (2008) 16, 771–779

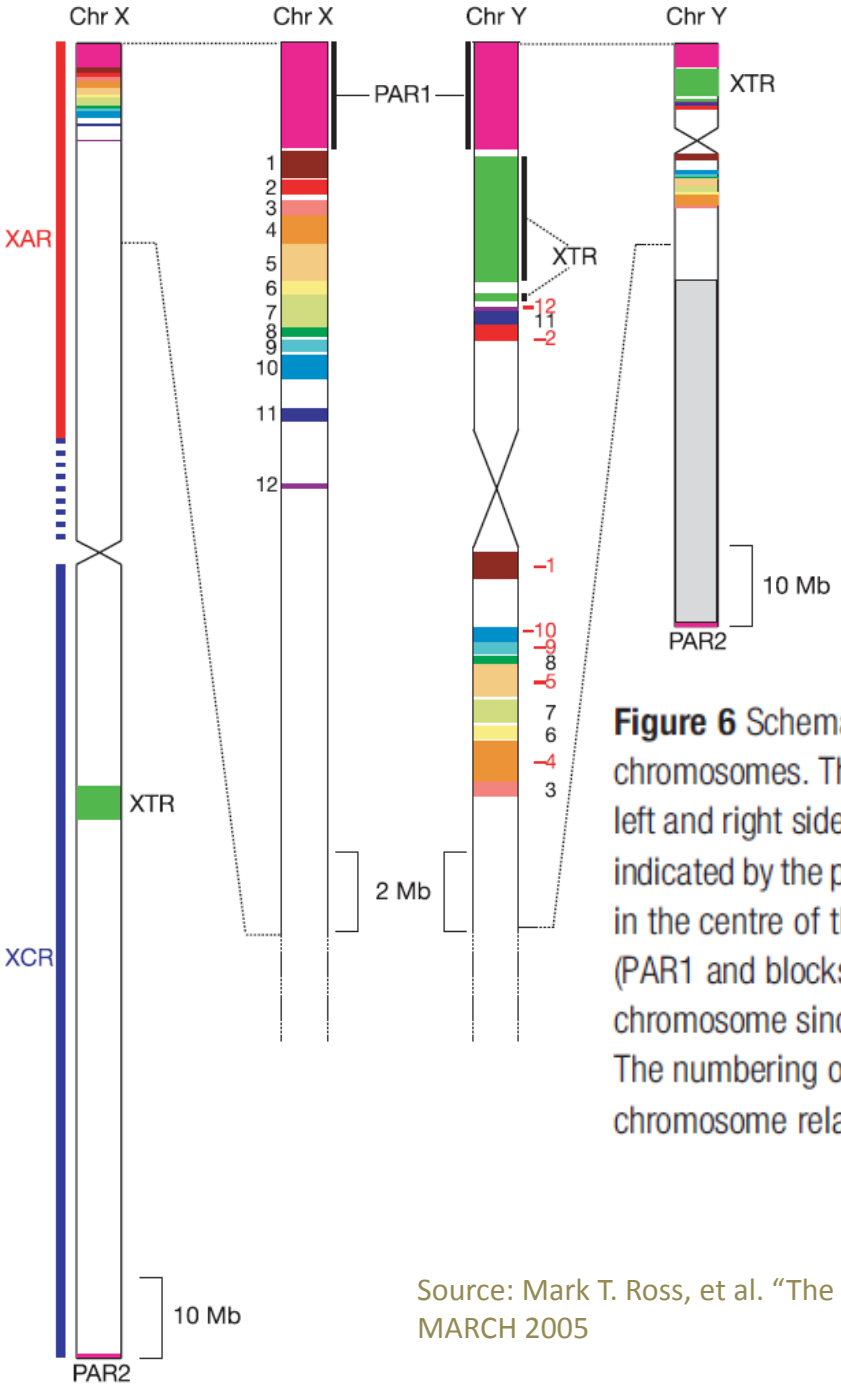


Figure 6 Schematic representation of major homologies between the human sex chromosomes. The entire X and Y chromosomes are shown using the same scale on the left and right sides of the figure, respectively. The major heterochromatic region on Yq is indicated by the pale grey box proximal to PAR2. Expanded sections of X and Y are shown in the centre of the figure. Homologies coloured in the figure are either part of the XAR (PAR1 and blocks 1–12), or were duplicated from the X chromosome to the Y chromosome since the divergence of human and chimpanzee lineages (XTR and PAR2). The numbering of XAR-YAR blocks follows that in Fig. 5b. Blocks inverted on the Y chromosome relative to the X chromosome are assigned red, negative numbers.

Source: Mark T. Ross, et al. "The DNA sequence of the human X chromosome", NATURE |VOL 434 | 17 MARCH 2005

Table 1 **Homologous genes on the human X and Y chromosomes**

Region	Distance from Xpter (Mb)	X gene*	Y gene	Distance from Ypter (Mb)†	XY homology block‡
Pseudoautosomal region PAR1 (XAR)	0.15	cU136G2.1 (*PLCXD1*)	cU136G2.1 (*PLCXD1*)	0.15	PAR1
	0.17	cU136G2.2 (*GTPBP6*)	cU136G2.2 (*GTPBP6*)	0.17	PAR1
	0.25	cM56G10.2§	cM56G10.2§	0.25	PAR1
	0.29	cM56G10.1 (*PPP2R3B*)	cM56G10.1 (*PPP2R3B*)	0.29	PAR1
	0.57	*SHOX*	*SHOX*	0.57	PAR1
	0.92	bA309M23.1§	bA309M23.1§	0.92	PAR1
	1.31	*CRLF2*	*CRLF2*	1.31	PAR1
	1.38	*CSF2RA*	*CSF2RA*	1.38	PAR1
	1.52	*IL3RA*	*IL3RA*	1.52	PAR1
	1.55	*SLC25A6*	*SLC25A6*	1.55	PAR1
	1.56	bA261P4.5§	bA261P4.5§	1.56	PAR1
	1.57	bA261P4.6 (*CXYorf2*)	bA261P4.6 (*CXYorf2*)	1.57	PAR1
	1.59	*ASMTL*	*ASMTL*	1.59	PAR1
	1.66	bA261P4.4 (*P2RY8*)	bA261P4.4 (*P2RY8*)	1.66	PAR1
	1.76	*DXYS155E* (*CXYorf3*)	*DXYS155E* (*CXYorf3*)	1.76	PAR1
	1.79	*ASMT*	*ASMT*	1.79	PAR1
	1.79	bB297E16.3§	bB297E16.3§	1.79	PAR1
	1.91	bB297E16.4§	bB297E16.4§	1.91	PAR1
	1.93	bB297E16.5§	bB297E16.5§	1.93	PAR1
	2.37	*DHRSX*	*DHRSX*	2.37	PAR1
	2.41	*ALTE* (*ZBED1*)	*ALTE* (*ZBED1*)	2.41	PAR1
	2.54	Em:AC097314.2§	Em:AC097314.2§	2.54	PAR1
	2.53	Em:AC097314.3§	Em:AC097314.3§	2.53	PAR1
	2.63	*CD99*	*CD99*	2.63	PAR1
X-added region (XAR)	3.57	*PRKX*	*PRKY*	7.23	2
	5.81	*NLGN4X*	*NLGN4Y*	15.23	5
	6.31	Em:AC108684.1 (*VCX3A*)	*VCY*, *VCY1B*	14.54, 14.6	6
	7.62	*VCX*	*VCY*, *VCY1B*	14.54, 14.6	9
	7.95	Em:AC097626.1 (*VCX2*)	*VCY*, *VCY1B*	14.54, 14.6	10
	8.24	Em:AC006062.2 (*VCX3B*)	*VCY*, *VCY1B*	14.54, 14.6	10
	9.37	*TBL1X*	*TBL1Y*	6.97	11
	11.07	*AMELX*	*AMELY*	6.78	12
	12.75	*TMSB4X*	*TMSB4Y*	14.25	
	16.59	*CXorf15*	*CYorf15A*, *CYorf15B*	20.13, 20.15	
	19.91	*EIF1AX*	*EIF1AY*	21.08	
	23.96	*ZFX*	*ZFY*	2.87	
	40.78	*USP9X*	*USP9Y*	13.33	
	40.96	*DDX3X*	*DDX3Y*	13.46	
	44.61	*UTX*	*UTY*	13.91	
X-conserved region (XCR)	53.00	dJ290F12.2 (*TSPYL2*)	*TSPY* (~35)	9.50	
	53.12	*SMCX*	*SMCY*	20.27	
	71.27	*RPS4X*	*RPS4Y1*, *RPS4Y2*	2.77, 21.27	
X-transposed region (XTR)	88.50	bB348B13.2§	n/a	2.96	XTR
	88.99	*TGIF2LX*	*TGIF2LY*	3.49	XTR
	91.26	*PCDH11X*	*PCDH11Y*	5.28	XTR
X-conserved region (XCR)	135.68	*RNMX* (*RBMX*)	*RBMY* (6)	22.02, 22.04, 22.37, 22.41, 22.66, 22.85	
	139.31	*SOX3*	*SRY*	2.70	
	148.38	Em:AC016940.3 (*HSFX2*)§	*HSFY1*, *HSFY2*	19.3, 19.12	
	148.56	Em:AC016939.4 (*HSFX1*)§	*HSFY1*, *HSFY2*	19.3, 19.12	
Pseudoautosomal region PAR2	154.57	*SPRY3*	*SPRY3*	57.44	PAR2
	154.71	*SYBL1*	*SYBL1*	57.58	PAR2
	154.81	*IL9R*	*IL9R*	57.67	PAR2
	154.81	Em:AJ271736.5§	Em:AJ271736.5§	57.69	PAR2
	154.82	Em:AJ271736.6 (*FAM39A*)§	Em:AJ271736.6 (*FAM39A*)§	57.69	PAR2

Pseudogenes are not included in the table.
*Gene names as shown in Supplementary Fig. 1. HUGO name is in parentheses when the two names differ. Em, EMBL entry.
† Distances refer to Y chromosome sequence assembly NCBI35. Where multiple Y chromosome orthologues exist, the locations of all copies are shown on the Y chromosome. The exception is TSPY, which has ~35 copies in an array centred at approximately 9.5 Mb on the Y chromosome[41].
‡ Major homology blocks as shown in Figs 5 and 6.
§ Novel cases of X genes with Y homologues assigned to these categories.

Source: Mark T. Ross, et al. “The DNA sequence of the human X chromosome”, NATURE |VOL 434 | 17 MARCH 2005

Association Studies of PARs

Table 2 Features of the pseudoautosomal regions in comparison to the sex chromosomes and the autosomal range

Genetic region	*Physical length (Mb)*	*Known protein coding genes (Mb)*	*Male recombination activity (cM/Mb)*	*Female recombination activity (cM/Mb)*	*Male/female quotient of genetic length*
PAR1	2.7	10	4.33–20.48	0.30–1.55	2.8–14.6
PAR2	0.33	15	6.06	—	—
X-specific	165	6	—	1.21	—
Y-specific	60	3	—	—	—
Autosomal range	46–245	3–23	0.80–2.40	1.40–2.80	0.57–0.85

Abbreviation: PAR, pseudoautosomal region.
Physical map lengths and known protein coding genes are taken from the Ensembl database,[41] genetic map lengths for autosomes from the Rutgers map.[30] For genetic lengths of PARs see Table 1.

- Using LD pattern in other autosomal chromosomes, a 540 SNPs in PAR1 and 60 in PAR2 are the least coverage for association test
- The higher recombination rate in PAR1, 540 SNPs for PAR1 would not be enough.
- When a causal variant is strongly associated with male sex, and/or if the sex distribution of the risk allele is different among cases and controls is different, special analytical methods may be used.

Table 3 Available SNP chips and the number of loci contained in PAR1 and PAR2

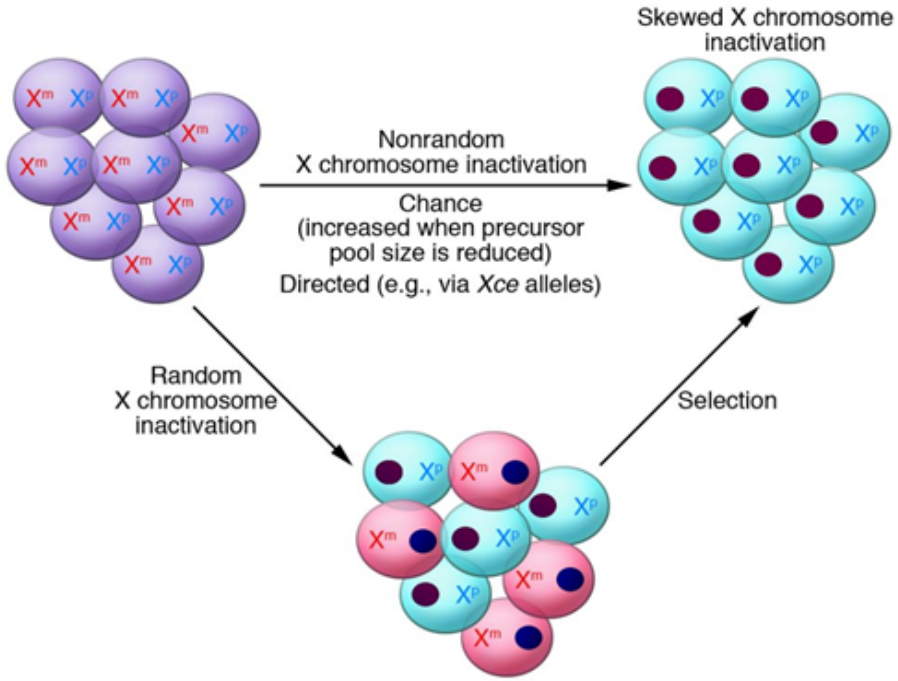
Company	*Chips*	*PAR1*	*PAR2*
Affymetrix	Genome-wide human SNP array 6.0	391	32
	Genome-wide human SNP array 5.0	155	0
	Mapping 500K array set	262	0
	Human mapping 100K set	19	0
	Mapping 10K 2.0 array	5	0
Illumina	Human1M	258	42
	Humanhap550	10	4
	Humanhap300-duo	2	0
	Humanhap240S	9	4
	Linkage V (6k)	18	5

Abbreviations: PAR, pseudoautosomal region; SNP, single nucleotide polymorphism.

Source: Antonia Flaquer, et al. “The human pseudoautosomal regions: a review for genetic epidemiologists”, European Journal of Human Genetics (2008) 16, 771–779

X Chromosome Inactivation (XCI)

- a process by which mammals reduce the number of active X chromosomes to one per diploid set of autosomes, thereby allowing for dosage equality between the sexes.
- XCI in mammals achieves dosage compensation between males and females for X-linked gene products. Inactivation of one X chromosome occurs early in female development and is initiated from the X-inactivation centre (XIC). The XIST transcript is expressed initially on both X chromosomes, but later the transcript from the chromosome that is destined for inactivation becomes more stable than the other. Finally, the transcript is expressed only from the inactive X chromosome (Xi).
- In random XCI, either the Xp or the maternal X chromosome (Xm) is subject to inactivation, and this stochastic choice appears to be made independently in each cell (Lyon, 1961).



Source: (1) Mark T. Ross, et al. "The DNA sequence of the human X chromosome", NATURE |VOL 434 | 17 MARCH 2005; (2) Joshua Starmer and Terry Magnuson "A new model for random X chromosome inactivation" Development 136, 1-10 (2009) doi:10.1242/dev.025908; (3) Jakub Minks, et al. "A skewed view of X chromosome inactivation" *J. Clin. Invest.* 118:20–23 (2008)

X Chromosome Inactivation (XCI)

Table 1
Different haplotype frequencies among HapMap populations for *XIST*

XIST haplotype[A]	CEU	YRI	JPT/CHB
T-A-C-T	0.59	0.17	0.07
C-C-C-T	0.16	0.22	0.01
C-A-C-C	0.2	0.09	0.92
C-A-T-T	0.06	0.22	0
C-A-C-T	0	0.30	0

[A]Haplotype designation is from four of the five SNPs used in Bolduc et al. study (6) (rs1082089 did not have complete population data) ordered by location on chromosome as rs1620574, rs1794213, rs1009948, and rs1894271). The allele information provided by Bolduc et al. did not explicitly state the order of SNPs; however, an estimated 80% of their subject population was homozygous for an *XIST* haplotype. Data retrieved from HapMap (22); release 21a (NCBI build 35). CEU, Utah (USA) residents with ancestry from northern and western Europe; YRI, Yoruba people from Ibadan, Nigeria; JPT/CHB, Japanese people from Tokyo, Japan, and Han Chinese people from Beijing, China.

XCI characteristics could be population dependent.

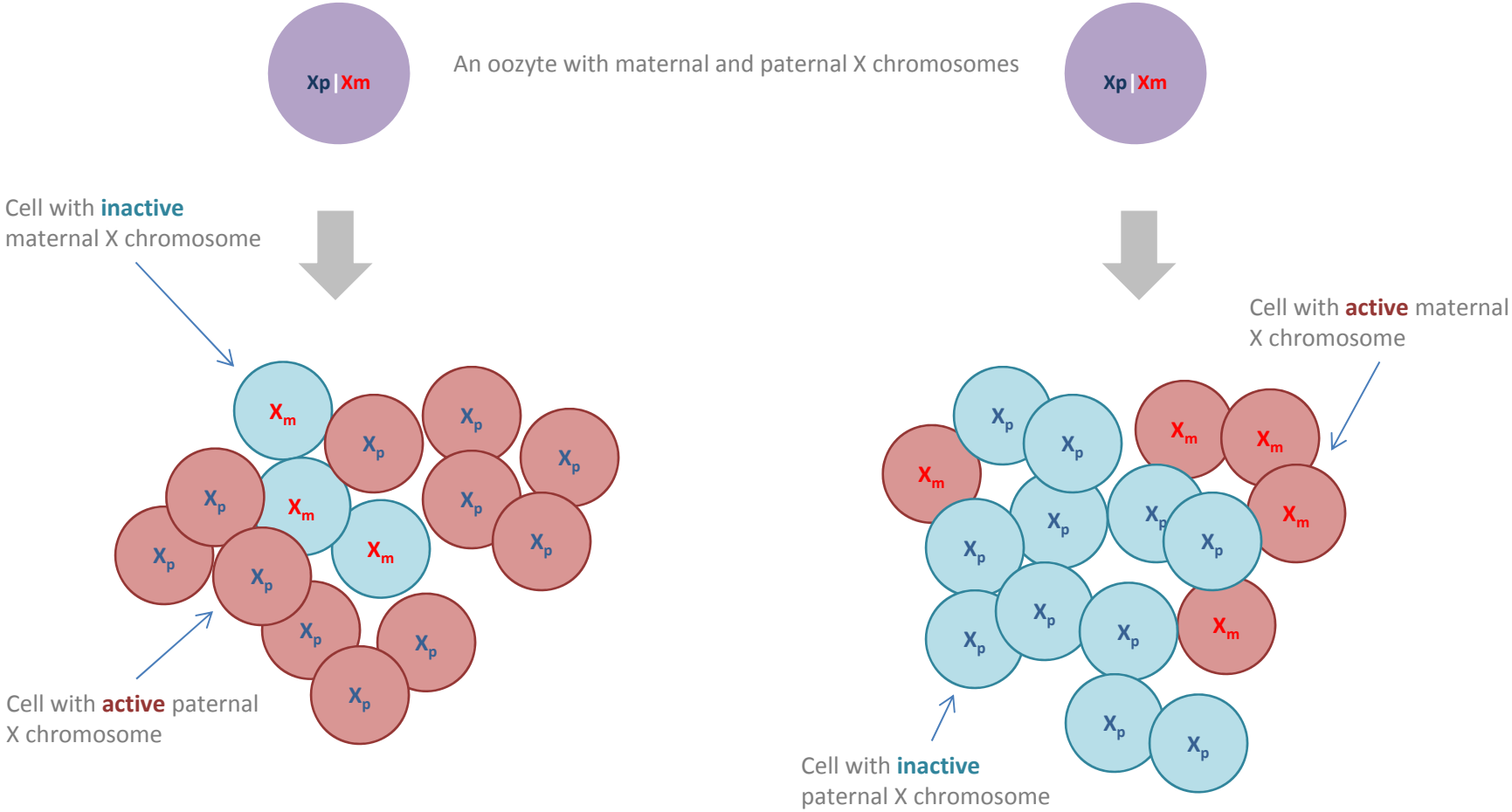
Source: Jakub Minks, et al. "A skewed view of X chromosome inactivation" *J. Clin. Invest.* 118:20–23 (2008)

Skewness of XCI

- XCI, according to the Lyon hypothesis, the fraction of cells with inactivation of each X chromosome should be equivalent. A deviation from the 1:1 XCI ratio is referred to as skewing.
- A popular definition of skewing is an XCI ratio ≥ 3:1.
- Skewness of XCI has been observed in young children but a higher proportion in females at increased ages, e.g. over 60 years of age.
- The HUMARA assay is one of the methods detecting XCI ratio.

Source: (1) Jakub Minks, et al. "A skewed view of X chromosome inactivation" *J. Clin. Invest.* 118:20–23 (2008); (2) Lambert Busque, et al. "Skewing of X-inactivation ratios in blood cells of aging women is confirmed by independent methodologies" BLOOD, 9 APRIL 2009 VOLUME 113, NUMBER 15

Skewness of XCI



Skewness of XCI

A Longitudinal Twin Study of Skewed X Chromosome-Inactivation

Chloe Chung Yi Wong[1], Avshalom Caspi[1,2], Benjamin Williams[1,2], Renate Houts[2], Ian W. Craig[1], Jonathan Mill[1]*

1 MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, London, United Kingdom, **2** Departments of Psychology and Neuroscience, Psychiatry and Behavioral Sciences, Institute for Genome Sciences and Policy, Duke University, Durham, North Carolina, United States of America

Abstract

X-chromosome inactivation (XCI) is a pivotal epigenetic mechanism involved in the dosage compensation of X-linked genes between males and females. In any given cell, the process of XCI in early female development is thought to be random across alleles and clonally maintained once established. Recent studies, however, suggest that XCI might not always be random and that skewed inactivation may become more prevalent with age. The factors influencing such XCI skewing and its changes over time are largely unknown. To elucidate the influence of stochastic, heritable and environmental factors in longitudinal changes in XCI, we examined X inactivation profiles in a sample of monozygotic (MZ) (n = 23) and dizygotic (DZ) (n = 22) female twin-pairs at ages 5 and 10 years. Compared to MZ twins who were highly concordant for allelic XCI ratios, DZ twins showed much lower levels of concordance. Whilst XCI patterns were moderately stable between ages 5 and 10 years, there was some drift over time with an increased prevalence of more extreme XCI skewing at age 10. To our knowledge, this study represents the earliest longitudinal assessment of skewed XCI patterns, and suggests that skewed XCI may already be established in early childhood. Our data also suggest a link between MZ twinning and the establishment of allelic XCI ratios, and demonstrate that acquired skewing in XCI after establishment is primarily mediated by stochastic mechanisms. These data have implications for our understanding about sex differences in complex disease, and the potential causes of phenotypic discordance between MZ female twins.

Citation: Wong CCY, Caspi A, Williams B, Houts R, Craig IW, et al. (2011) A Longitudinal Twin Study of Skewed X Chromosome-Inactivation. PLoS ONE 6(3): e17873. doi:10.1371/journal.pone.0017873

Editor: Catherine Suter, Victor Chang Cardiac Research Institute, Australia

Received December 14, 2010; **Accepted** February 11, 2011; **Published** March 22, 2011

A Longitudinal Twin Study of Skewed X Chromosome-Inactivation

Chloe Chung Yi Wong[1], Avshalom Caspi[1,2], Benjamin Williams[1,2], Renate Houts[2], Ian W. Craig[1], Jonathan Mill[1]*

Subjects

Samples were obtained from Caucasian MZ and DZ twin-pairs enrolled in the Environmental Risk (E-Risk) Longitudinal Twin Study, which has been described in detail elsewhere [30]. In brief, E-Risk investigates how genetic and environmental factors shape children's development. For our XCI analyses, we selected female children, comprising 23 MZ twin-pairs and 22 DZ twin pairs for whom DNA was available at both ages 5 and 10 years. DNA samples (total n = 180) were collected from buccal cells using an established method yielding high molecular weight genomic DNA [31]. All DNA samples were tested for degradation and purity; any degraded or impure samples, as well as those uninformative (i.e. homozygous) for the XCI assay were excluded from analysis (25 and 21 DNA samples were excluded at ages 5 and 10 years, respectively). Ethical approval was granted by the Joint South London and Maudsley and the Institute of Psychiatry Research Ethics Committee and written informed consent were given by all subjects.

Determination of X-inactivation pattern

Allelic XCI ratios were determined by examining DNA methylation in the vicinity of a polymorphic $(CAG)_n$ repeat in the first exon of the human androgen receptor (*AR*) gene as described previously [26,32]. DNA methylation across this region of the *AR* gene is highly correlated with X-inactivation status [32] and the human androgen receptor assay (HUMARA) is a widely used clinical assay for determining XCI status. In brief, 150 ng of genomic DNA was divided into three aliquots and incubated with *Hpa*II, *Msp*I or water. Digested DNA was amplified using primers (one fluorescently labeled) flanking the polymorphic $(CAG)_n$ repeat in the first exon of the *AR* gene (F: 5′-FAM-GCTGTGAAG-GTTGCTGTTCCTCAT-3′ and R:5′-TCCAGAATCTGTTC-CAGAGCGTGC-3′). Fluorescently labeled products were separated on an automated DNA sequencer (Applied Bioscience, Foster City, CA, USA) to accurately quantify the peak heights for each allele. Allelic XCI ratios were calculated as the normalized ratio of peak heights of the shorter to longer alleles of the *Hpa*II digest to those observed in a water (mock) digest.

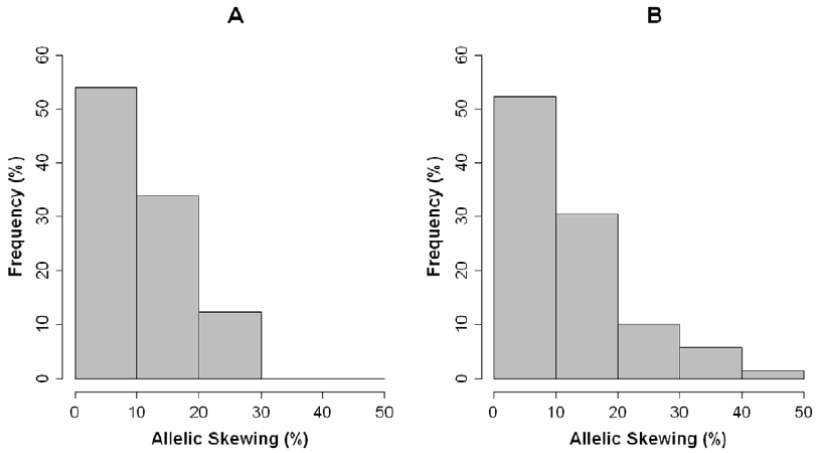
A Longitudinal Twin Study of Skewed X Chromosome-Inactivation

Chloe Chung Yi Wong[1], Avshalom Caspi[1,2], Benjamin Williams[1,2], Renate Houts[2], Ian W. Craig[1], Jonathan Mill[1]*

Table 1. Skewed X-inactivation ratios (deviation from 50:50) in females at ages 5 and 10 years (The percentage reported is cumulative for each group).

			Percentage of Sample with Allelic Skewing of		
Age	**Zygosity**	**Mean (SD)**	**>20%**	**>30%**	**>40%**
5	ALL	48:52 (13.4)	12.3 (n = 8)	0	0
	MZ	47:53 (13.8)	12.9 (n = 4)	0	0
	DZ	49:51 (13.2)	11.8 (n = 4)	0	0
10	ALL	47:53 (14.8)	17.4 (n = 12)	7.3 (n = 5)	1.5 (n = 1)
	MZ	47:53 (16.7)	23.5 (n = 8)	8.8 (n = 3)	2.9 (n = 1)
	DZ	47:53 (12.9)	11.4 (n = 4)	5.7 (n = 2)	0

doi:10.1371/journal.pone.0017873.t001



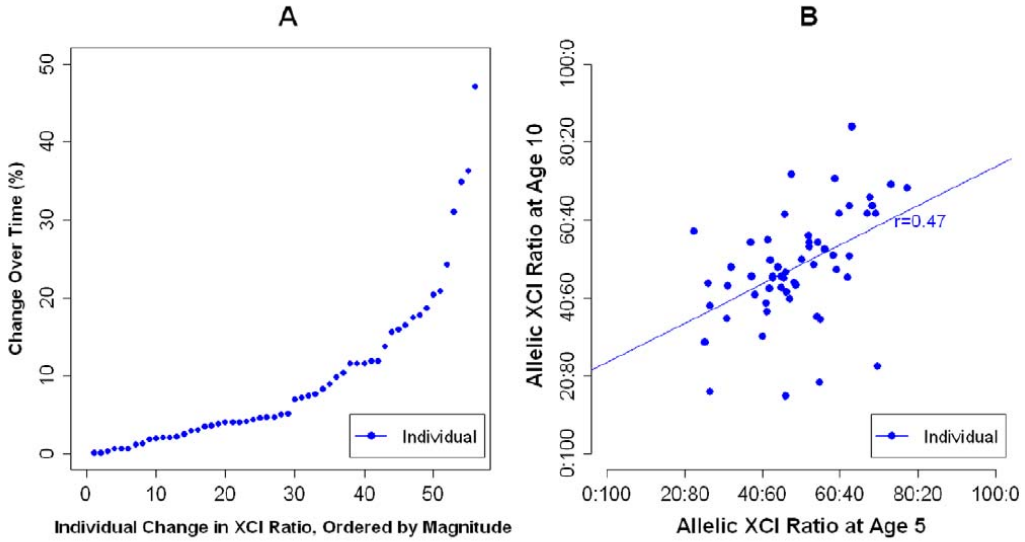


Figure 2. Longitudinal analysis of allelic XCI ratios in female children over early childhood development. (*A*) Individual changes in absolute XCI ratio between ages 5 and 10 years. (*B*) Inter-individual stability correlation for XCI ratio, between ages 5 and 10 years. *XCI = X-Chromosome Inactivation*.

doi:10.1371/journal.pone.0017873.g002

EVIER

Experimental Hematology 33 (2005) 1209–1214

Age-associated skewing of X-inactivation ratios of blood cells in normal females: a candidate-gene analysis approach

Pierre Chagnon[a,b], Sylvie Provost[a], Claude Belisle[a], Véronique Bolduc[a], Marianne Gingras[a], and Lambert Busque[a,c]

[a]*Research Center, Maisonneuve-Rosemont Hospital, University of Montreal, Montreal, Quebec, Canada;*
[b]*Institute for Research in Immunology and Cancer, University of Montreal, Montreal, Quebec, Canada;*
[c]*Department of Hematology, Maisonneuve-Rosemont Hospital, University of Montreal, Montreal, Quebec, Canada*

(Received 10 April 2005; revised 17 May 2005; accepted 17 June 2005)

X-inactivation is a random process that occurs in females early during embryogenesis. Females are mosaics with an equal proportion of cells with the paternal (Xp) or maternal X-chromosome (Xm) in the active state. However, close to 40% of healthy females aged more than 60 y.o. present a significant skewing of X-inactivation ratios (Xp:Xm >3 :1). The exact etiology of this age-associated skewing (AAS) in blood cells is unknown. We hypothesized that AAS is due to hemizygous cell selection caused by allelic variants in hematopoiesis or cell survival genes. To test this hypothesis, we recruited 700 unrelated healthy females of French Canadian ancestry aged more than 60. We determined X-inactivation ratio at the HUMARA locus. We genotyped 81 different SNPs, using TaqMan® technology, in 15 different candidate genes with known role in hematopoiesis, cell cycle, or X-inactivation. Extensive statistical analyses were conducted and demonstrated that none of the 15 candidate genes investigated contribute significantly to AAS.

Am. J. Hum. Genet. 71:168–173, 2002

Report

Skewed X-Chromosome Inactivation Is a Common Feature of X-Linked Mental Retardation Disorders

Robert M. Plenge,[1,*] Roger A. Stevenson,[2] Herbert A. Lubs,[3] Charles E. Schwartz,[2] and Huntington F. Willard[1]

[1]Department of Genetics, Case Western Reserve University School of Medicine, and Center for Human Genetics and Research Institute, University Hospitals of Cleveland, Cleveland; [2]J. C. Self Research Institute, The Greenwood Center, Greenwood, SC; and [3]Department of Pediatrics, Genetics Division, School of Medicine, University of Miami, Miami

Some deleterious X-linked mutations may result in a growth disadvantage for those cells in which the mutation, when on the active X chromosome, affects cell proliferation or viability. To explore the relationship between skewed X-chromosome inactivation and X-linked mental retardation (XLMR) disorders, we used the androgen receptor X-inactivation assay to determine X-inactivation patterns in 155 female subjects from 24 families segregating 20 distinct XLMR disorders. Among XLMR carriers, ~50% demonstrate markedly skewed X inactivation (i.e., patterns ≥80:20), compared with only ~10% of female control subjects ($P < .001$). Thus, skewed X inactivation is a relatively common feature of XLMR disorders. Of the 20 distinct XLMR disorders, 4 demonstrate a strong association with skewed X inactivation, since all carriers of these mutations demonstrate X-inactivation patterns ≥80:20. The XLMR mutations are present on the preferentially inactive X chromosome in all 20 informative female subjects from these families, indicating that skewing is due to selection against those cells in which the XLMR mutation is on the active X chromosome.

Skewed X-Chromosome Inactivation Is a Common Feature of X-Linked Mental Retardation Disorders

Robert M. Plenge,[1,*] Roger A. Stevenson,[2] Herbert A. Lubs,[3] Charles E. Schwartz,[2] and Huntington F. Willard[1]

[1]Department of Genetics, Case Western Reserve University School of Medicine, and Center for Human Genetics and Research Institute, University Hospitals of Cleveland, Cleveland; [2]J. C. Self Research Institute, The Greenwood Center, Greenwood, SC; and [3]Department of Pediatrics, Genetics Division, School of Medicine, University of Miami, Miami

Table 2

Skewed X-Inactivation Patterns in XLMR Carriers

X-Inactivation Pattern	Frequency of Skewed X Inactivation[a] (%)		
	Female Control Subjects	XLMR Noncarriers[b]	XLMR Carriers[c]
≥90:10	3	2	30
≥80:20	9	15	48
≥70:30	30	41	63

[a] To assess statistically the distribution of X-inactivation patterns, a χ^2 test was used to compare the number of female carriers above and below a particular threshold value (≥90:10, ≥80:20, and ≥70:30) to that of the control population, as described by Plenge et al. (1999). To control for multiple hypothesis testing, a Bonferroni correction was applied, and the significance value was set at $P < .01$.

[b] Results were not statistically significant.

[c] All results were significant ($P < .001$).

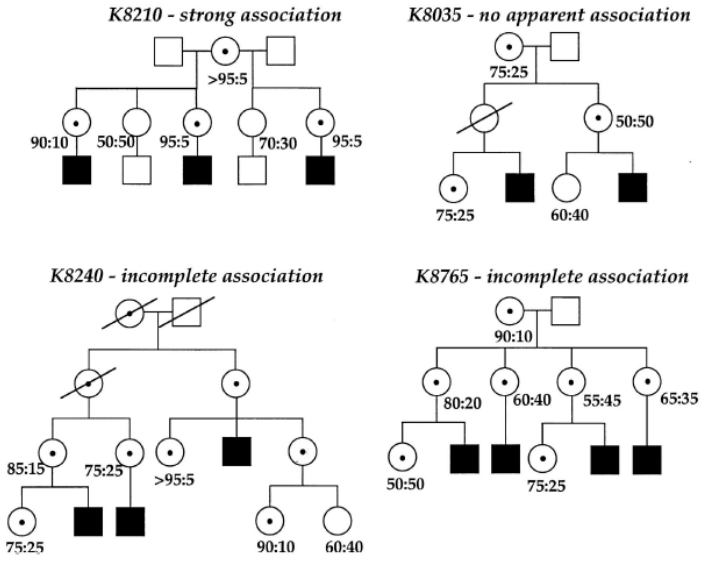


Figure 2 Pedigrees of illustrative families with XLMR. *AR* X-inactivation patterns are shown near each informative female subject. Blackened symbols denote affected individuals, unblackened symbols denote unaffected individuals, and symbols with a black dot denote carriers.

Skewed X-Chromosome Inactivation Is a Common Feature of X-Linked Mental Retardation Disorders

Robert M. Plenge,[1,*] Roger A. Stevenson,[2] Herbert A. Lubs,[3] Charles E. Schwartz,[2] and Huntington F. Willard[1]

[1]Department of Genetics, Case Western Reserve University School of Medicine, and Center for Human Genetics and Research Institute, University Hospitals of Cleveland, Cleveland; [2]J. C. Self Research Institute, The Greenwood Center, Greenwood, SC; and [3]Department of Pediatrics, Genetics Division, School of Medicine, University of Miami, Miami

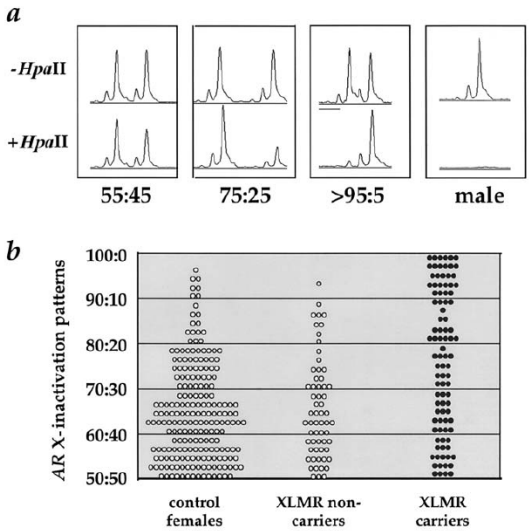


Figure 1 X-inactivation patterns in XLMR disorders. *a*, *AR* X-inactivation–pattern tracings. The top tracings represent the undigested DNA (−*Hpa*II) from three female control subjects and from a male control subject; the bottom tracings represent DNA digested with *Hpa*II prior to PCR (+*Hpa*II). The relative intensity of the two alleles after digestion represents the *AR* X-inactivation pattern for each individual (expressed as a ratio and normalized to the undigested samples). The tracing in males disappears, representing complete digestion of the unmethylated allele on the active X chromosome. Details of the *AR* X-inactivation assay have been described elsewhere, including methods for correcting for unequal peak heights owing to preferential allele amplification (Allen et al. 1992; Naumova et al. 1996; Plenge et al. 1997, 1999). *b*, Distribution of *AR* X-inactivation patterns in families with XLMR and control subjects. The *AR* X-inactivation patterns are shown for two control populations (*unblackened circles*) and for the XLMR carrier population (*blackened circles*).

Skewed X-Inactivation Patterns in XLMR Carriers

X-Inactivation Pattern	Frequency of Skewed X Inactivation[a] (%)		
	Female Control Subjects	XLMR Noncarriers[b]	XLMR Carriers[c]
⩾90:10	3	2	30
⩾80:20	9	15	48
⩾70:30	30	41	63

[a] To assess statistically the distribution of X-inactivation patterns, a χ^2 test was used to compare the number of female carriers above and below a particular threshold value (⩾90:10, ⩾80:20, and ⩾70:30) to that of the control population, as described by Plenge et al. (1999). To control for multiple hypothesis testing, a Bonferroni correction was applied, and the significance value was set at $P < .01$.

[b] Results were not statistically significant.

[c] All results were significant ($P < .001$).

Regulation of X-chromosome inactivation by the X-inactivation centre

Sandrine Augui, Elphège P. Nora* and Edith Heard*

Abstract | X-chromosome inactivation (XCI) ensures dosage compensation in mammals and is a paradigm for allele-specific gene expression on a chromosome-wide scale. Important insights have been made into the developmental dynamics of this process. Recent studies have identified several *cis*- and *trans*-acting factors that regulate the initiation of XCI via the X-inactivation centre. Such studies have shed light on the relationship between XCI and pluripotency. They have also revealed the existence of dosage-dependent activators that trigger XCI when more than one X chromosome is present, as well as possible mechanisms underlying the monoallelic regulation of this process. The recent discovery of the plasticity of the inactive state during early development, or during cloning, and induced pluripotency have also contributed to the X chromosome becoming a gold standard in reprogramming studies.

Escaping XCI

- Not all genes on the inactivated X chromosome are silenced, i.e. they are participating in gene expression.

Source:
(1) Shapiro LJ, Mohandas T, Weiss R, Romeo G (1979) Non-inactivation of an X-chromosome locus in man. Science 204: 1224–1226.
(2) Brown CJ, Greally JM (2003) A stain upon the silence: Genes escaping X inactivation. Trends Genet 19: 432–438.
(3) Carrel L, Willard HF (2005) X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature 434: 400–404.
(4) Carrel L, Park C, Tyekucheva S, Dunn J, Chiaromonte F, et al. (2006) Genomic environment predicts expression patterns on the human inactive X chromosome. PLoS Genet 2(9): e151. DOI: 10.1371/journal.pgen.0020151

NATURE | VOL 434 | 17 MARCH 2005 | www.nature.com/nature

X-inactivation profile reveals extensive variability in X-linked gene expression in females

Laura Carrel[1] **& Huntington F. Willard**[2]

[1]*Department of Biochemistry and Molecular Biology, Pennsylvania State University College of Medicine, Hershey, Pennsylvania 17033, USA*
[2]*Institute for Genome Sciences & Policy, Duke University, Durham, North Carolina 27708, USA*

In female mammals, most genes on one X chromosome are silenced as a result of X-chromosome inactivation[1,2]. However, some genes escape X-inactivation and are expressed from both the active and inactive X chromosome. Such genes are potential contributors to sexually dimorphic traits, to phenotypic variability among females heterozygous for X-linked conditions, and to clinical abnormalities in patients with abnormal X chromosomes[3]. Here, we present a comprehensive X-inactivation profile of the human X chromosome, representing an estimated 95% of assayable genes in fibroblast-based test systems[4,5]. In total, about 15% of X-linked genes escape inactivation to some degree, and the proportion of genes escaping inactivation differs dramatically between different regions of the X chromosome, reflecting the evolutionary history of the sex chromosomes. An additional 10% of X-linked genes show variable patterns of inactivation and are expressed to different extents from some inactive X chromosomes. This suggests a remarkable and previously unsuspected degree of expression heterogeneity among females.

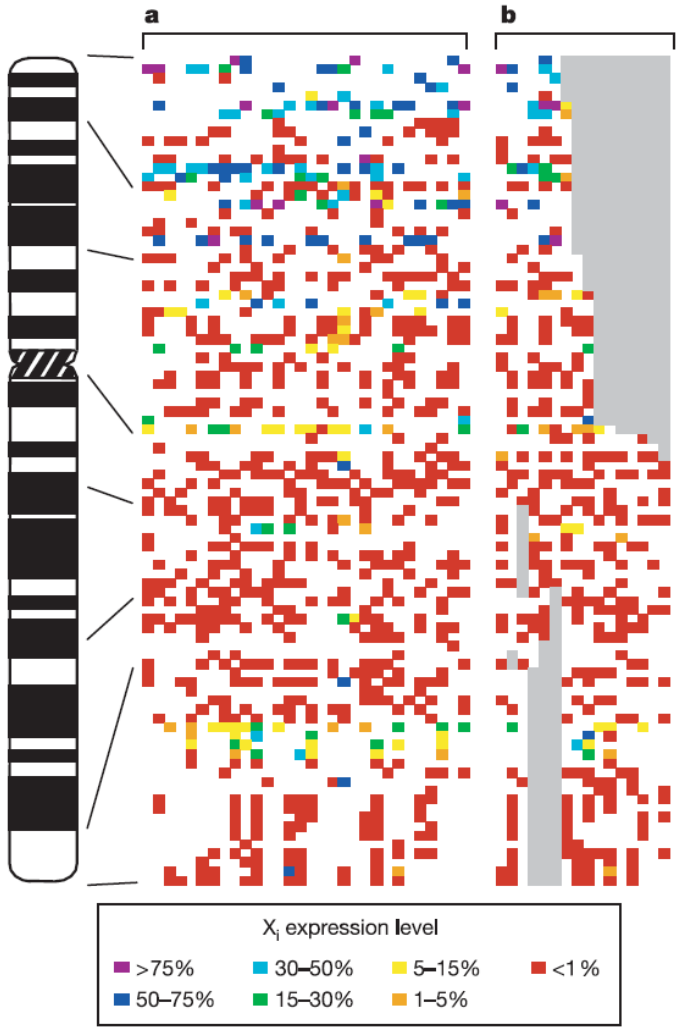


Figure 2 X-inactivation as a measure of allelic expression in non-randomly inactivated primary fibroblasts. **a**, **b**, Gene expression levels from normal X_i chromosomes (**a**) and from structurally abnormal X_i chromosomes (**b**). Each gene is linearly arrayed and approximate correspondence to chromosome location is indicated. The gene order is as in Supplementary Table 1. Each gene was assayed in all heterozygous individuals. Uninformative samples remain uncoloured. Colour-coding reflects relative X_i expression level as indicated. Grey boxes in **b** indicate absent portions of the X chromosome due to deletions or translocations.

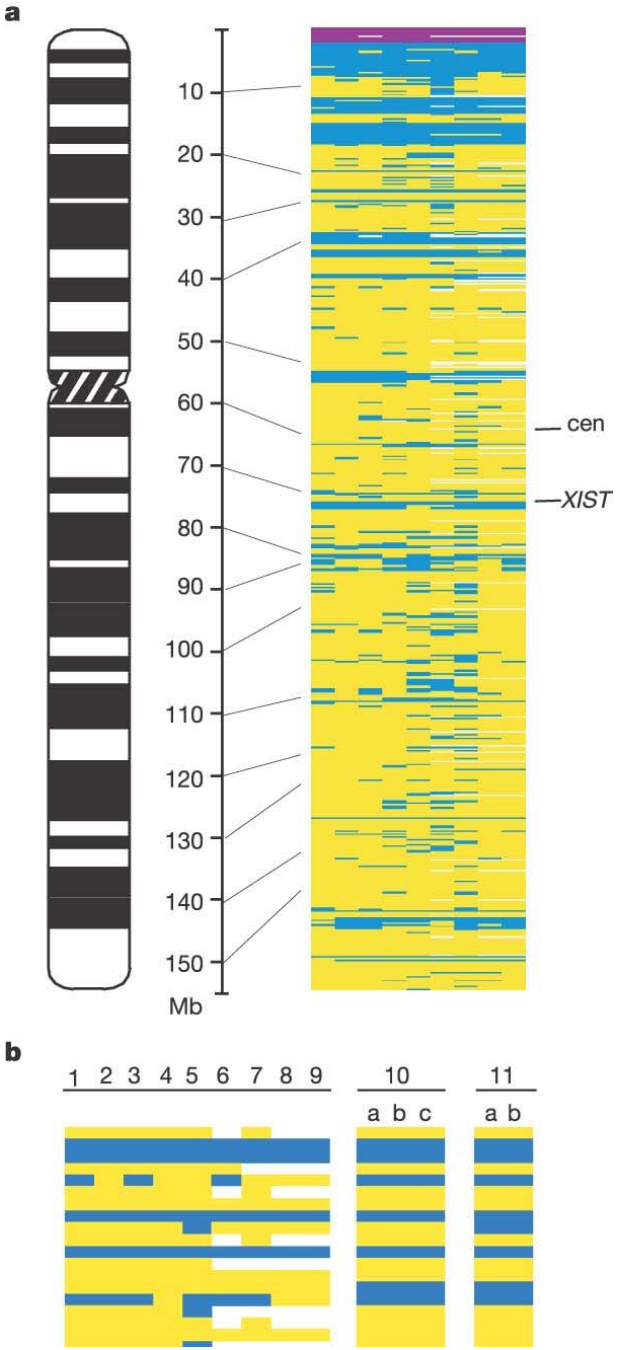
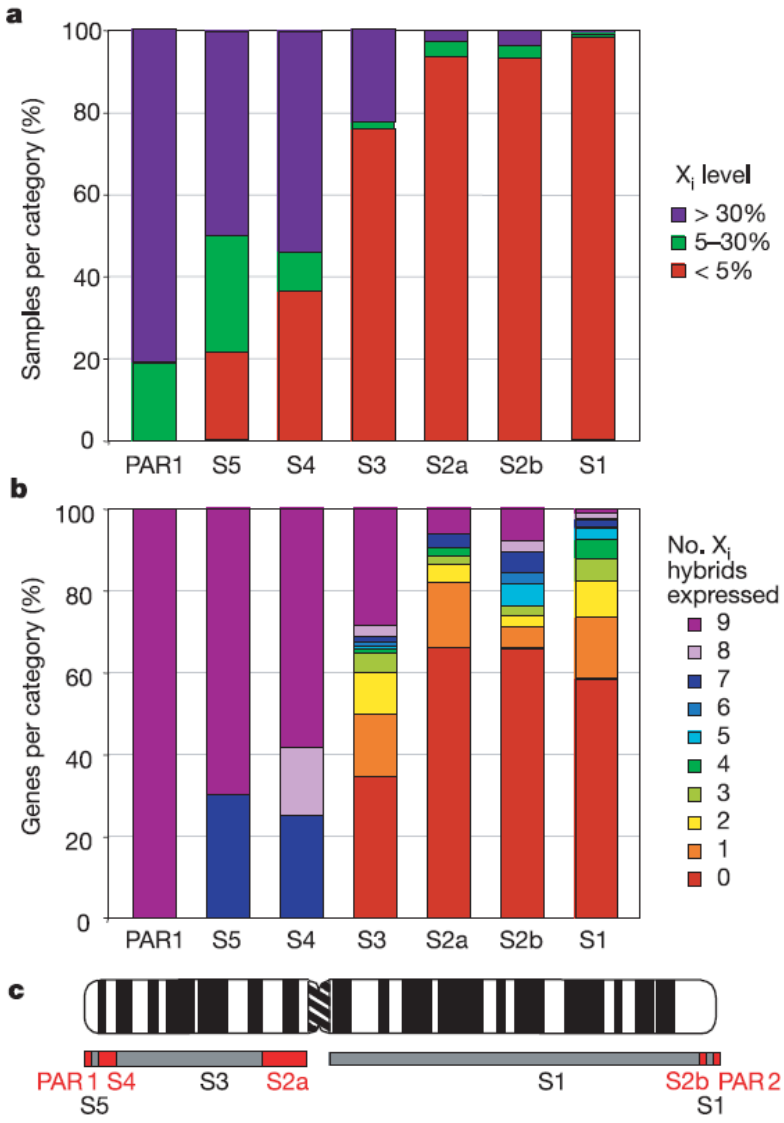


Figure 3 X-inactivation profile of the human X chromosome. **a**, 624 genes were tested in nine X_i hybrids. Each gene is linearly displayed. Blue denotes significant X_i gene expression, yellow shows silenced genes, pseudoautosomal genes are purple, and untested hybrids remain white. Positions of the centromere (cen) and *XIST* are indicated. **b**, To determine whether heterogeneity is largely a property of a specific chromosome, three independently derived hybrids from one X_i chromosome (denoted 10 a, b, c) and two hybrids carrying another X_i chromosome (11 a, b) were isolated, and results for 19 genes are shown adjacent to results for the original nine X_i chromosomes.

Figure 4. The Distribution of Correctly and Incorrectly Classified Genes along the X Chromosome

Dark green indicates correctly classified genes; light green indicates misclassified genes. X inactivation expression patterns [6] for genes included in this study: yellow indicates inactivated genes, and blue indicates escape genes. Not all genes were analyzed at all distances because sequences that included adjacent genes with *different* inactivation patterns were excluded from analysis (see Methods). These gene distances remain uncolored.

DOI: 10.1371/journal.pgen.0020151.g004



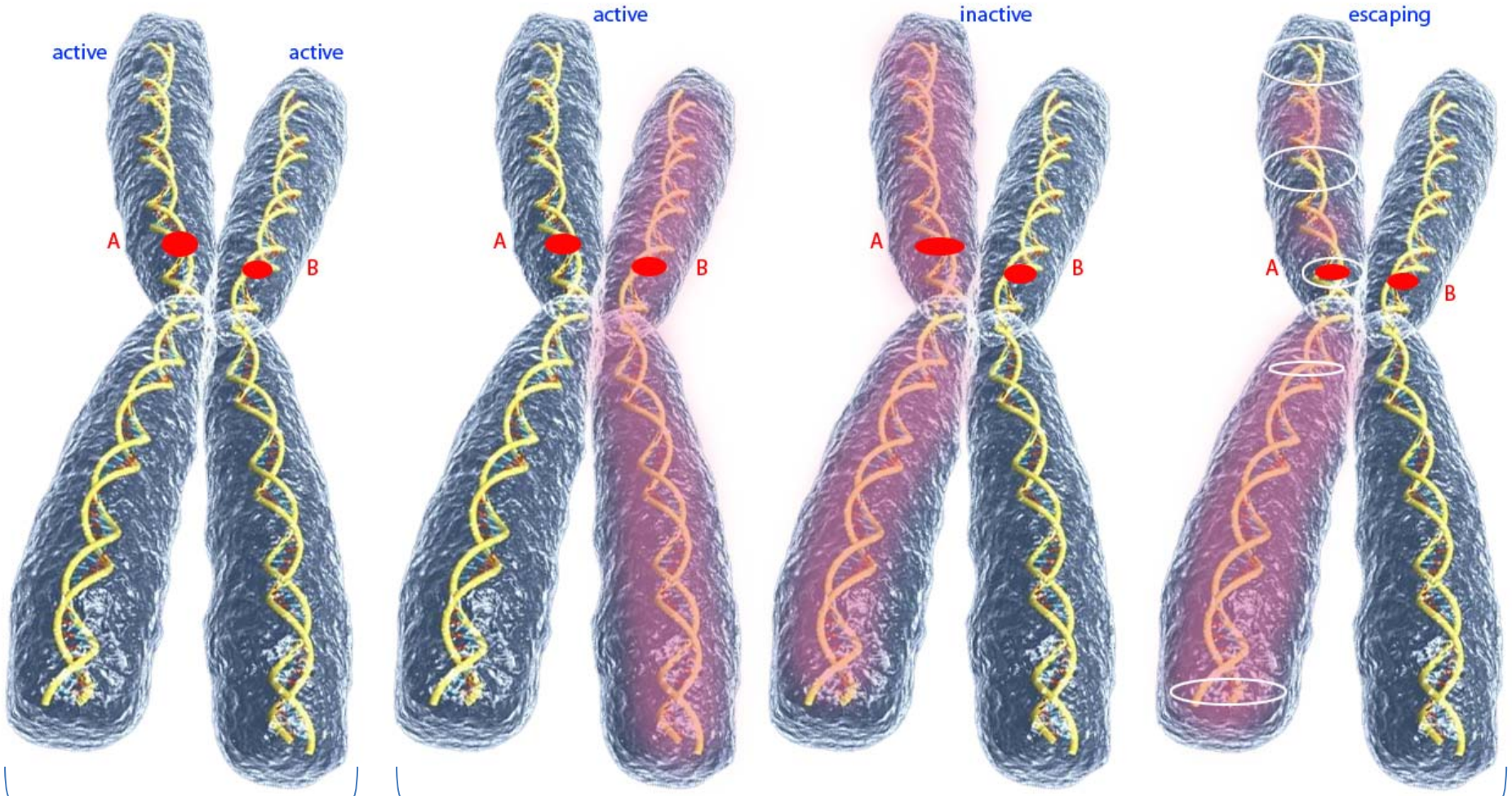
X-inactivation profile reveals extensive variability in X-linked gene expression in females

Laura Carrel[1] **& Huntington F. Willard**[2]

[1]*Department of Biochemistry and Molecular Biology, Pennsylvania State University College of Medicine, Hershey, Pennsylvania 17033, USA*
[2]*Institute for Genome Sciences & Policy, Duke University, Durham, North Carolina 27708, USA*

Figure 4 X_i expression data in primary fibroblasts and X_i hybrids correlate with location on X. Transcripts were subdivided according to X_i results and chromosomal region. **a**, The proportion of samples assayed in primary fibroblasts that show similar X_i results are indicated for the pseudoautosomal region PAR1, and for each X-specific evolutionary strata (S1–S5). **b**, For each chromosomal region, the percentage of genes showing similar X_i hybrid results (out of the nine X_i hybrid lines tested) are also shown. **c**, Location of each evolutionary strata on the X chromosome[13,19,21].

XCI and Escaping



Biostatistics (2008), **9**, 4, *pp.* 593–600
doi:10.1093/biostatistics/kxn007
Advance Access publication on April 25, 2008

Testing for association on the X chromosome

DAVID CLAYTON

Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research, University of Cambridge, Wellcome Trust/MRC Building, Addenbrookes's Hospital, Cambridge CB2 0XY, UK
david.clayton@cimr.cam.ac.uk

Original Paper

Human Heredity

Hum Hered 2011;71:23–36
DOI: 10.1159/000323768

Received: October 12, 2010
Accepted after revision: December 21, 2010
Published online: February 16, 2011

Association Tests for X-Chromosomal Markers – A Comparison of Different Test Statistics

Christina Loley[a, b] Andreas Ziegler[a] Inke R. König[a]

[a]Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, und [b]Medizinische Klinik II, Universitätsklinikum Schleswig-Holstein, Campus Lübeck, Lübeck, Deutschland

Testing for association on the X chromosome

DAVID CLAYTON

Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research, University of Cambridge, Wellcome Trust/MRC Building, Addenbrookes's Hospital, Cambridge CB2 0XY, UK
david.clayton@cimr.cam.ac.uk

SUMMARY

The problem of testing for genotype–phenotype association with loci on the X chromosome in mixed-sex samples has received surprisingly little attention. A simple test can be constructed by counting alleles, with males contributing a single allele and females 2. This approach does assume not only Hardy–Weinberg equilibrium in the population from which the study subjects are sampled but also, perhaps, an unrealistic alternative hypothesis. This paper proposes 1 and 2 degree-of-freedom tests for association which do not assume Hardy–Weinberg equilibrium and which treat males as homozygous females. The proposed method remains valid when phenotype varies between sexes, provided the allele frequency does not, and avoids the loss of power resulting from stratification by sex in such circumstances.

3. THE X CHROMOSOME

Loci on the pseudo-autosomal part of the X chromosome can be treated in exactly the same way as autosomal loci, but others generally require different treatment. For these, males will only carry 1 copy, while, in females, most loci are subject to X inactivation (Chow *and others*, 2005), so that a female will have approximately half her cells with 1 copy active while the remainder of her cells have the other copy activated. Thus, in the absence of interaction with other loci or environmental factors, males should be equivalent to homozygous females in respect to such loci. This suggests that, for X loci in males, A_i should be coded 0 or 2, while D_i should be coded 0. This has several consequences, some of which require modifications to the theory outlined above.