Finding Disease Genes

Dennis Drayna

NIDCD/NIH

Traditional Positional Cloning

- The ability to identify a gene solely on the basis of its location in the genome
- Information about the biochemistry, physiology, or pathology of the gene product is not required

Finding location

- Typically begins with a genetic linkage study

Genetic Linkage

- The co-inheritance of 2 traits as they are passed from parent to offspring
- Violations of Mendel's Second Law
- Linkage occurs because the genes encoding the traits studied reside immediately adjacent to each other on the chromosome
  - Can employ DNA sequence variation itself as the heritable trait

A genome-wide linkage survey

- Test DNA markers at known locations across the genome
- Perform statistical analysis to quantitate the degree of linkage
- Highly developed for mapping Mendelian traits

Linkage to gene

- Observation of linkage reduces the problem by ~ 3 orders of magnitude
  - 25,000 genes to 25 genes, under best case
- Detailed gene-by-gene search is then required
  - Sequencing
  - Mutation detection

How do you know you've found the gene?

- The causative variant is found in affected individuals and not in normals
- Different mutations in the same gene in different families with the same disease

Optimism, then pessimism

Positional cloning for Mendelian disorders led to the great triumphs of human genetics in the 90's

- Positional cloning of Mendelian disease genes remains a powerful and productive approach; niche applications

Attention then turned to non-Mendelian disorders, including many major health problems

Genetic linkage and positional cloning studies were remarkably unsuccessful

Reasons traits can display complex inheritance

- Genetic heterogeneity
  - More than one gene at different locations can give rise to the disorder
- Low penetrance
  - Having the disease genotype confers less than 100% probability of having the disease
- Continuously distributed quantitative traits
  - Not affected vs. normal
- Non-genetic contributions
  - Environmental, developmental, stochastic

New methods of analysis

- Model free
- Based on allele sharing
- Genehunter, Solar, Allegro, Merlin

New study populations

- Genetic isolates
- Inbred populations
- Tribal populations

Renewed emphasis on association studies

- Case-control studies showing association of disease with DNA variants has been historically problematic, but
- Technology for assaying large numbers of DNA variants (SNP's) has improved rapidly
- Association studies have been shown to have greater statistical power than linkage to identify disease gene loci of small effect
  - Common disease – common variant hypothesis

Genome-wide association studies

- Still too many genes/SNP's to evaluate all genes
- Haplotype blocks within the genome may allow use of a reduced set of "tagging SNP's"
- Haplotype structure of the human genome not yet characterized
- Association studies remain vulnerable to several problems

Issues in studies of complex traits

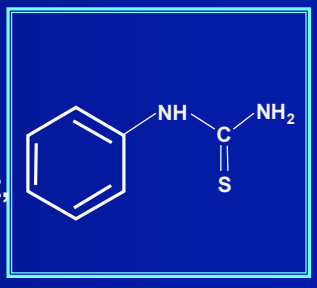
- **Continuously distributed phenotypes**
- **Uncertain mode of inheritance**
- **Low LOD scores**
- **Failures to replicate**
- **Linkage evidence at similar but non-identical loci**

Goal

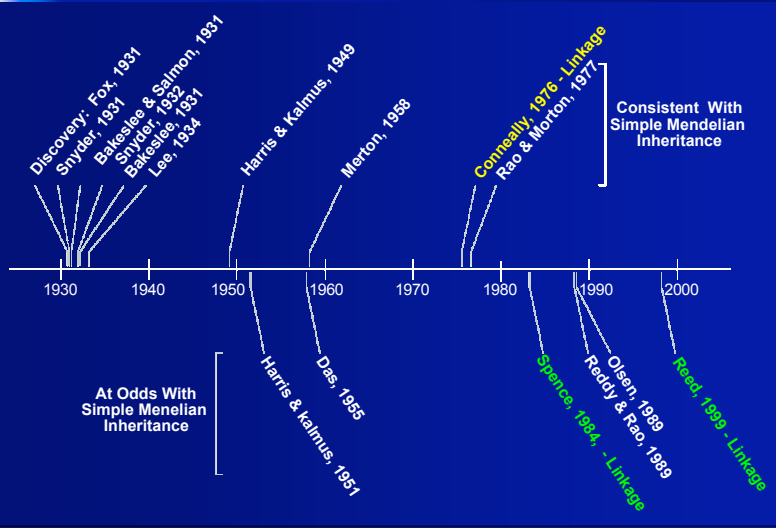
- Identify an "intermediate" trait as a model for complex disease

Phenylthiocarbamide (PTC)

- **Perceived as bitter**
- **PTC taste deficiency discovered by Fox, 1931**
- **Extremely well studied trait, > 500 papers describing results in > 130,000 subjects**



History of Our Understanding of PTC Genetics



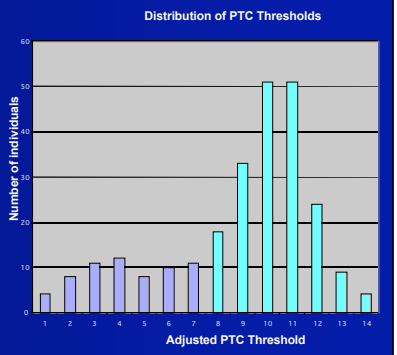
Centre d'Etude Polymorphisme duHumain (CEPH)

(Center for the Study of Human Polymorphism, Paris)

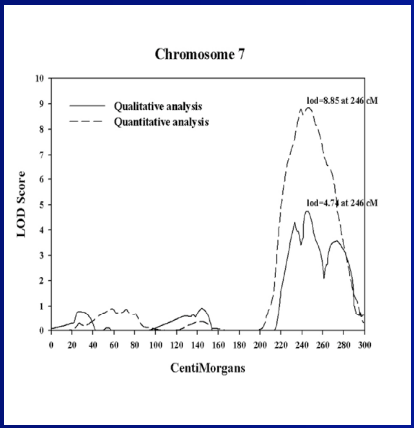
- **Set of standard normal reference families, chosen for maximally informative family structure – 4 grandparents and many sibs**
- **Used to construct the normal human genetic linkage map – worldwide effort**
- **Genotype data at 3,000 – 11,000 loci available on each individual**
  - *Enables genetic linkage studies without additional genotyping*
- **48/60 CEPH families from Utah**

PTC phenotyping

- Classical method of Harris and Kalmus
  - 14 solutions of PTC in water, starting with 1 μM and increasing 2-fold at each step
  - perception of bitter taste, subjects given 3 cups of water and 3 cups of that concentration of PTC, secretly marked
  - Threshold is most dilute correct sort



Linkage Results



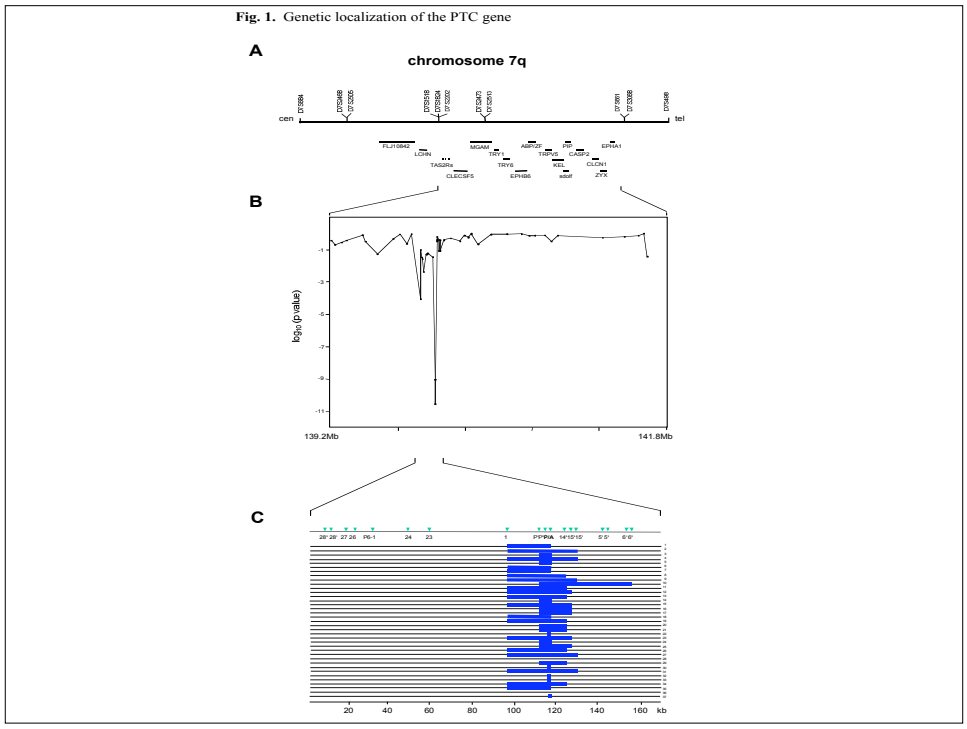
- Qualitative analysis
  - 4.74 (246 cM) on chr. 7
- Quantitative analysis
  - 8.85 (246 cM) on chr. 7
  - 2.01 (4 cM) on chr. 16

Search of the chromosome 7 region

- Bitter taste transduction is believed to be mediated by G protein-coupled receptor (GPCR) signaling pathways
- Numerous GPCR's located in this region
  - 16 genes sequenced : 9 TAS2R genes
    7 OR-like genes
  - 8 individuals evaluated
  - Many sequence differences identified
  - One difference was observed to absolutely correlate with phenotype in chromosome 7-linked families
  - Unrelated non-tasters also appeared to carry this difference
    - Suggested a founder effect and linkage disequilibrium

SNP analysis of critical region

- 60 evenly spaced single nucleotide polymorphisms (SNP's) were genotyped to evaluate linkage disequilibrium
- Unrelated C.E.P.H. individuals - Northern European
  - 27 Families
    - 269 individuals
- Unrelated NIH individuals
  - All races / ethnicities
  - 94 individuals
    - Chose 46 individuals for Linkage Disequilibrium (LD) measurement

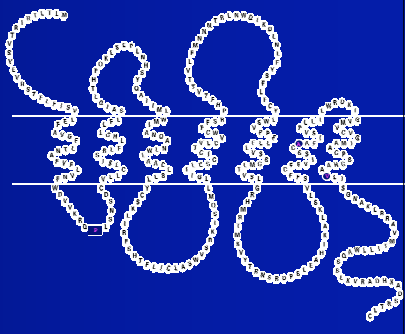


PTC gene and its product

1002 bp, 333 a.a.

member of TAS2R family

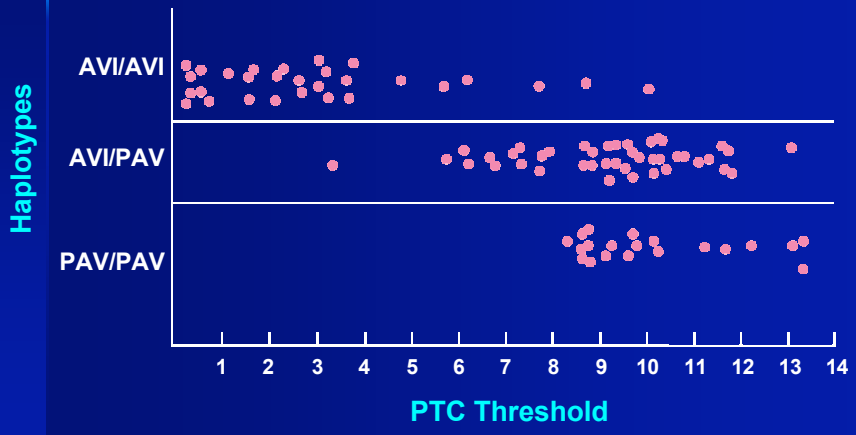
- G protein coupled receptor family
- single coding exon
- 7 transmembrane
- short extracellular N-terminus
- 30% identity with hTAS2R7



Polymorphisms in the PTC Gene

SNP			Amino Acid		Location in predicted protein
Position (b.p.)	Allele	Frequency	Position (a.a.)	a.a. encoded	
145	C	.36	49	Proline	1st intracellular loop
	G	.64		Alanine	
785	C	.38	262	Alanine	6th transmembrane
	T	.62		Valine	
886	G	.38	296	Valine	7th transmembrane
	A	.62		Ile	

Correlation of Haplotypes with Phenotype in Unrelated Individuals



Haplotype Association with Taste Phenotypes

Haplotypes	Sample	No. of subjects	
		Nontasters	Tasters
AVI / AVI	Utah	38	14
	NIH	21	0
AVI / AAV	Utah	10	7
	NIH	1	3
* / PAV	Utah	3	108
	NIH	1	58

* indicates any haplotype found in the sample. No AAV homozygotes were observed in either sample

Conclusions – Phenotype to molecular structure and function

- Using positional candidate gene and SNP association methods, found strong LD in TAS2R gene
- Multiple haplotypes observed in this gene
- Haplotypes explain phenotype
  - Heterozygote effect
  - All the bimodality
  - 55 - 85% of variance

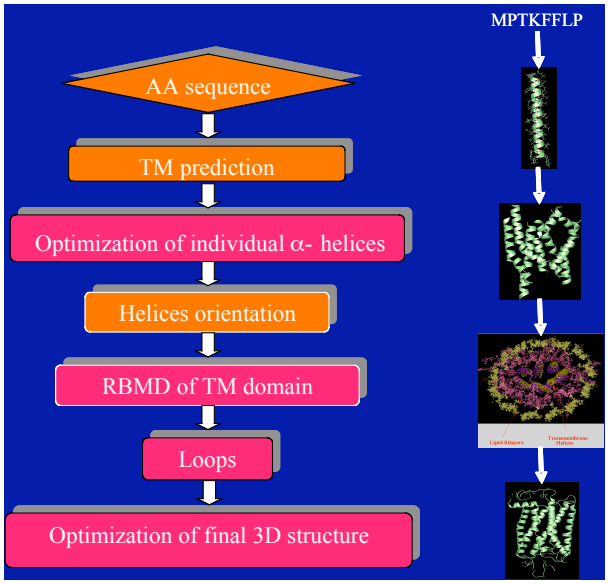
General applicability

- Quantitative vs. qualitative trait specification
- Linkage followed by targeted SNP association studies are now a highly promising strategy for complex traits
- NOD2 in Crohn's Disease
- Others

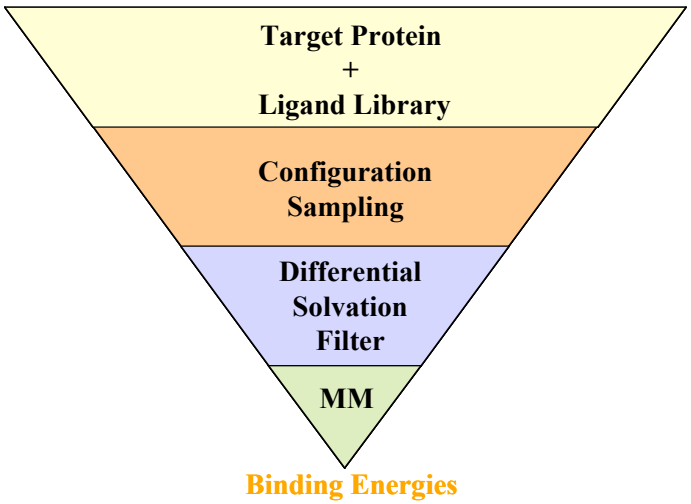
Protein structural difference predictions

- Integral membrane proteins are not conducive to traditional crystalographic approaches
- Employed computational predictive methods

MembStruk



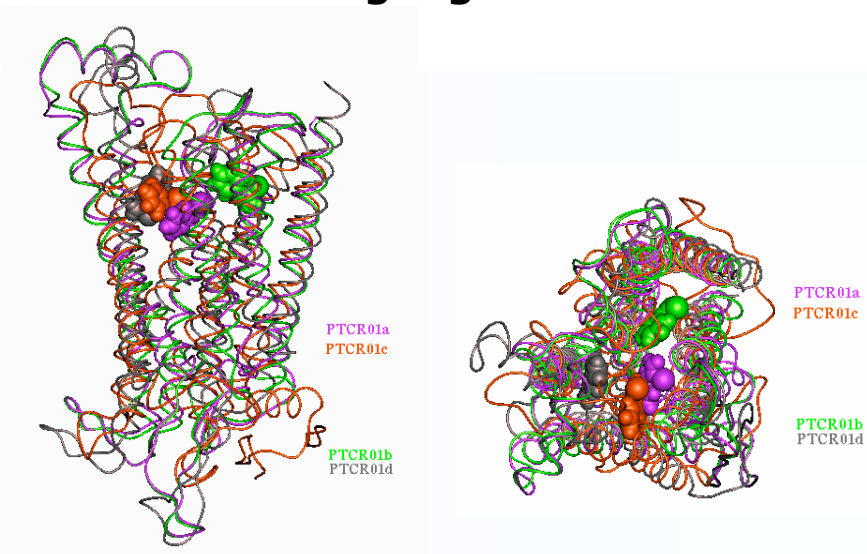
Hierarchical docking



Multiple 3D Structures

PTCR01a and PTCR02a	(a) MembStruk 3.5
PTCR01b and PTCR02b	(b) MembStruk 3.5 with helices rotations adjusted for Cys-Cys bound.
PTCR01c and PTCR02c	(c) Homology to bovine rhodopsin (PDB 1L9H)
PTCR01d and PTCR02d	(d) Helices bending and translations from 1L9H, TM predictions from MembStruk 3.5, rotations based on Cys-Cys bounds and arbitrary loops.

Best binding region for PTC



Initial Conclusions

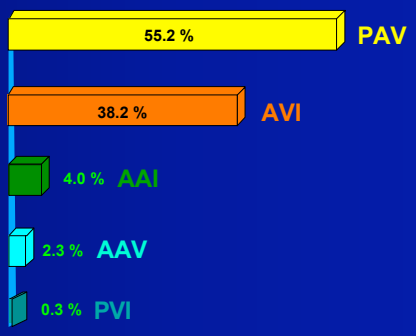
- Binding affinities of the taster and non-taster forms of the protein for PTC are predicted to be the same
- Suggests that the non-taster form of the protein is non-functional due to G-protein coupling or signaling

Population genetics and evolution

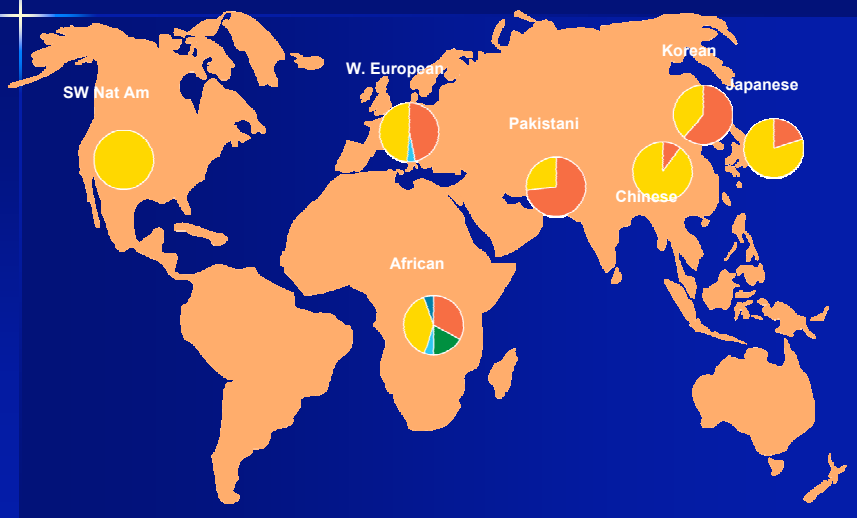
- PTC has been intensively studied by population geneticists for 70 years
- Numerous predictions about PTC gene variation have been made
- Can these studies provide additional understanding about gene function?

PTC gene haplotypes

- Examined haplotypes within and surrounding gene in 11 populations worldwide
  - Within the gene, five haplotypes observed
  - PVI and AAI were exclusively found in Africa



Worldwide Haplotype Distribution



Why are non-tasters so common?

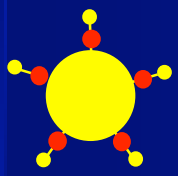
- PTC and structurally related compounds are bitter and toxic to the thyroid
- Their bitter taste causes strong aversion, a presumed selective advantage

3 Possibilities

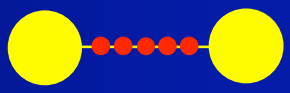
- Drift
  - Cascade of bottlenecks
- Population subdivision
  - Allele frequencies in Africa not significantly different than in other populations – not the likely cause
- Selection

Analysis of DNA sequence variation to evaluate drift vs. selection

- Tajima's D statistic
  - Compares Π and S
  - Affected by "shape" of haplotype network



D = -1.9

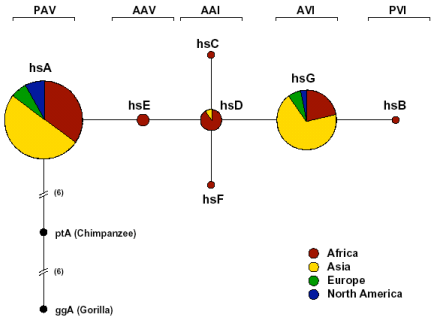


D = 3.4

PTC

D = 1.59

Analysis of DNA sequence variation to evaluate drift vs. selection



Conclusion/paradox

- The non-taster allele (along with the taster allele) has come to high frequency because it's selected for
- How could a non-functional allele of a protective gene be selected for?

Hypothesis

- The non-taster form of the protein does not contain disabling differences
  - Not a deletion, no stop codons
- We now predict the non-taster allele of the PTC gene encodes a perfectly functional receptor for another bitter toxic substance, not yet identified

Collaborators

University of Utah
- Mark Leppert
- Hilary Coon
- Andrew Peiffer
- Stephen Wooding
- Lynn Jorde

Stanford University
- Neil Risch
- Eric Jorgenson

NIDCD/NIH
- Un-kyung Kim

California Institute of Technology
- Wely Floriano
- William Goddard