



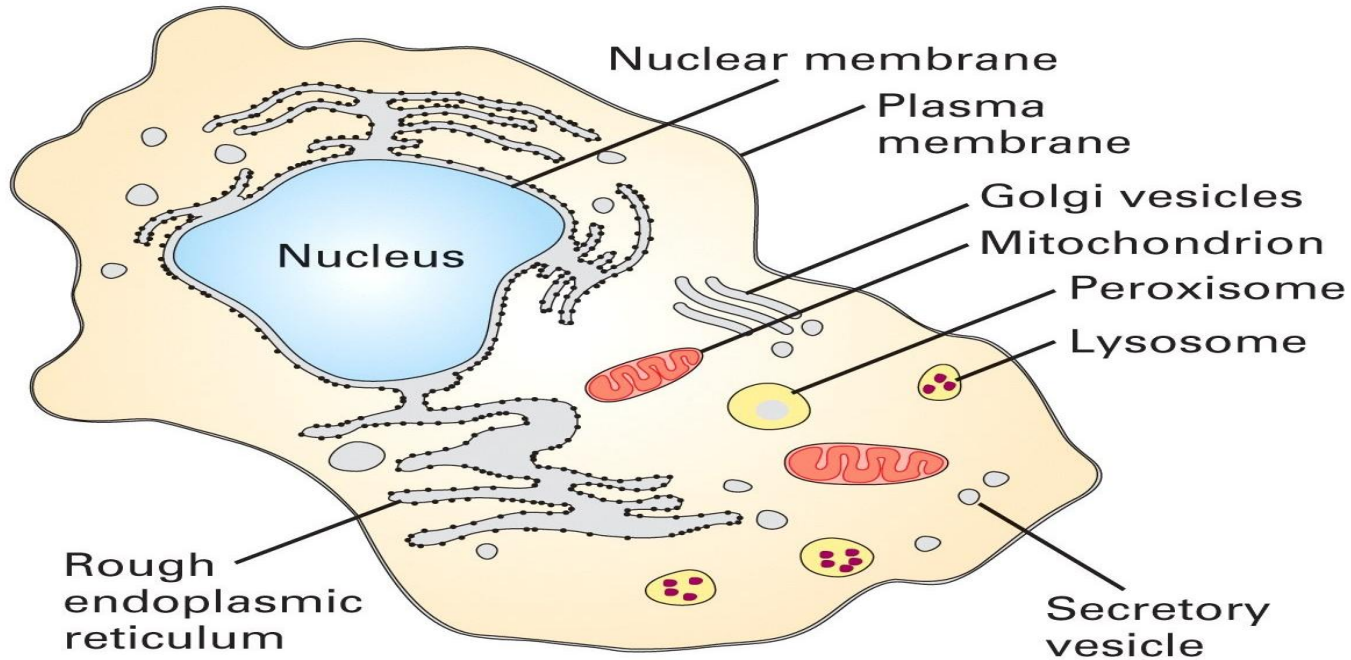
Trimethylaminuria (TMAU)

Yiran Guo, Ph.D.
Center for Applied Genomics
Children's Hospital of Philadelphia

TMAU Genetics

- Background in Human Genetics
- Human genome variants and methods to detect them
- Rare Mendelian disorders and inheritance
- Genetic analysis of TMAU

Life begins with Cell



- A cell is a smallest structural unit of an organism that is capable of independent functioning
- All cells have some common features

Cells Information and Machinery

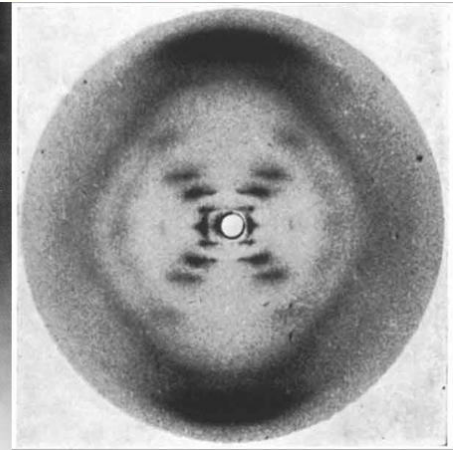
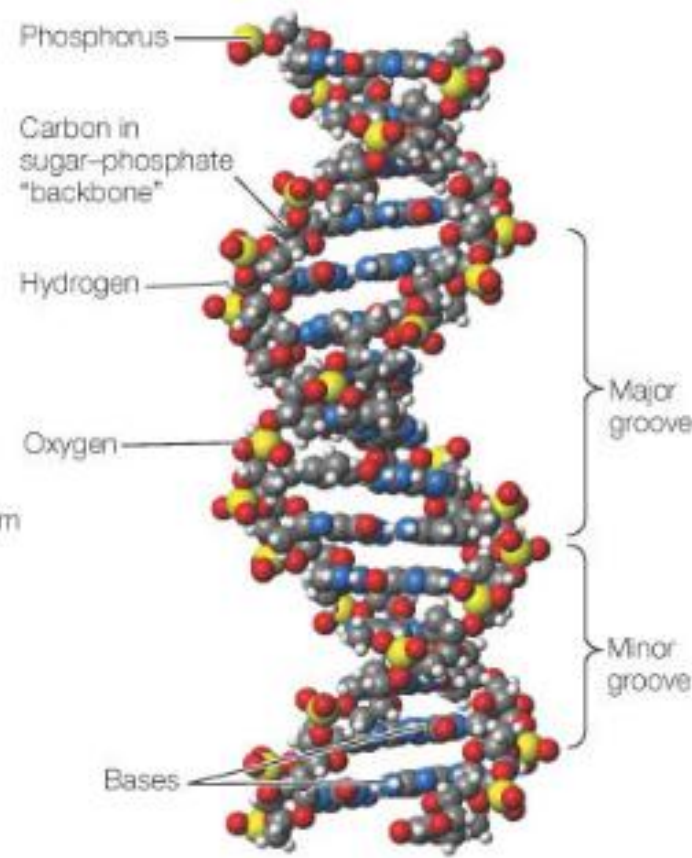
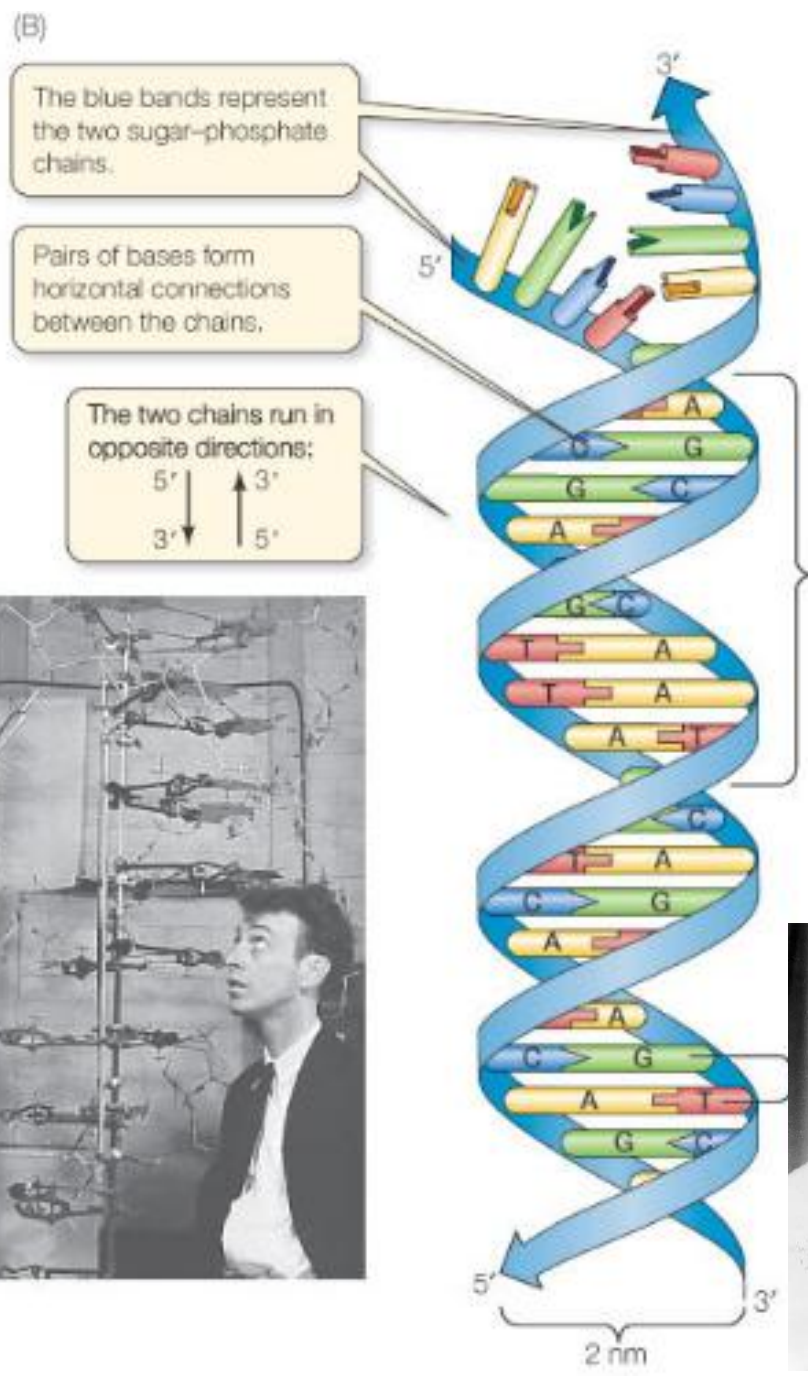
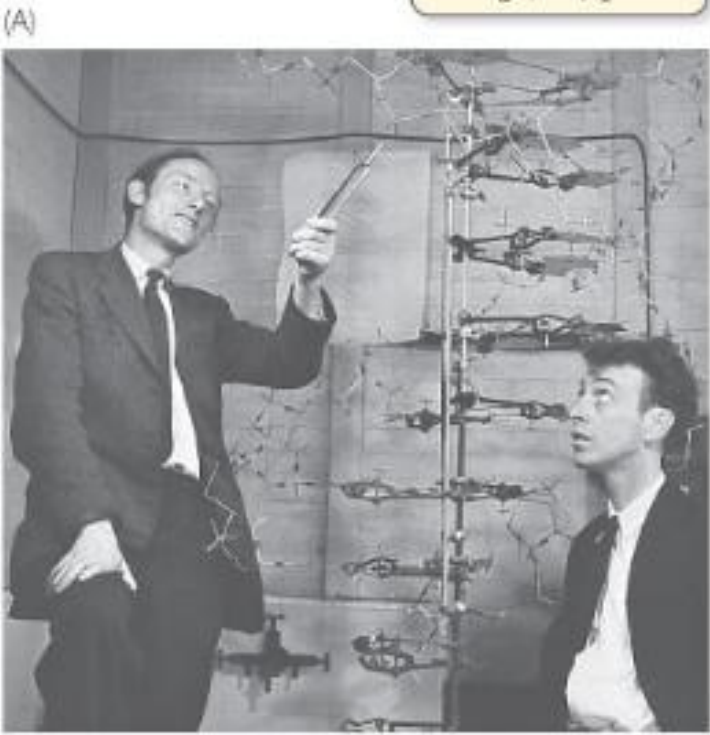
- Cells store all information to replicate itself
 - Human genome is around 3 billions base pair long
 - **Almost** every cell in human body contains same set of genes
 - But not all genes are used or expressed by those cells
- Machinery:
 - Collect and manufacture components
 - Carry out replication
 - Kick-start its new offspring(A cell is like a car factory)

Key Terminology

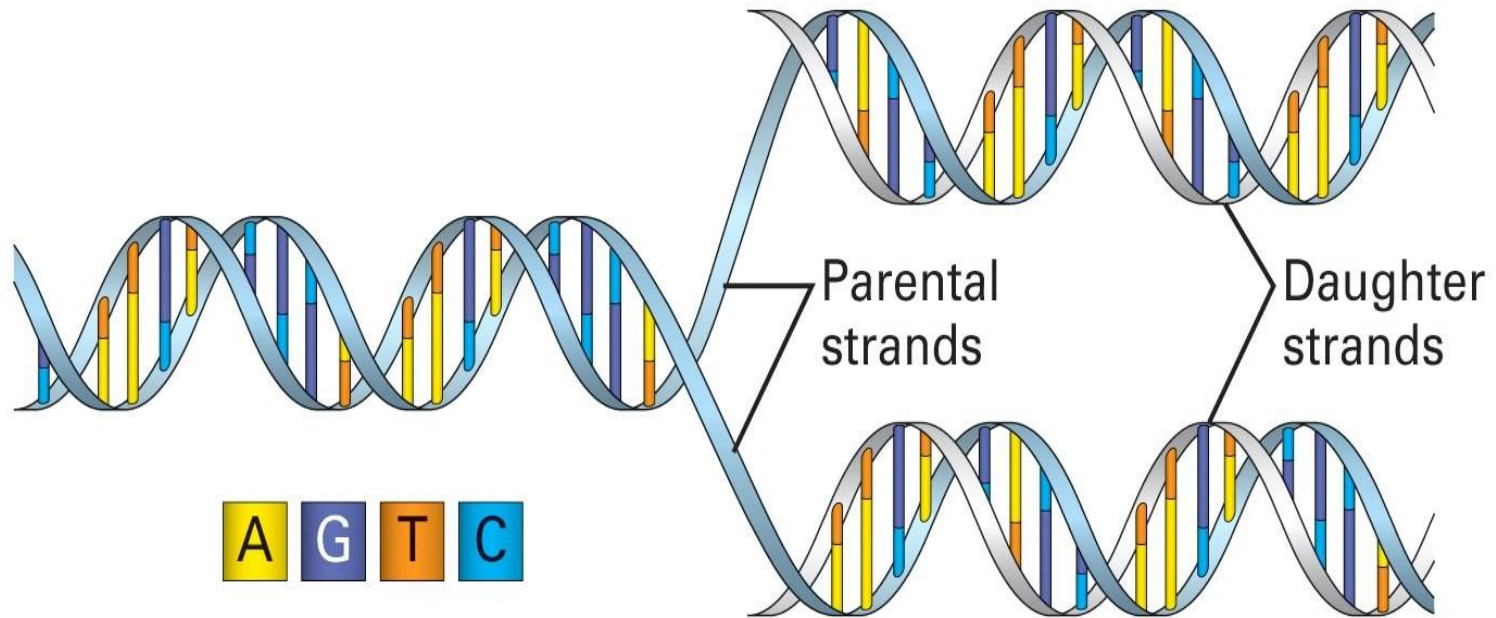
- The **genome** is an organism's complete set of DNA (genetic material).
 - a bacteria contains about 600,000 DNA base pairs
 - human and mouse genomes have some 3 billion.
 - human genome has 24 distinct chromosomes, each chromosome contains many **genes**.
- **Gene**
 - basic physical and functional units of heredity.
 - specific sequences of DNA bases that encode instructions on how to make **proteins**.
- **Genotype**
 - the genetic makeup of an organism
- **Phenotype**
 - the physical expressed traits of an organism
- **Nucleic acid**
 - biological molecules(RNA and DNA) that allow organisms to reproduce;
- **Proteins**
 - make up the cellular structure
 - large, complex molecules made up of smaller subunits called **amino acids**.

All Life depends on 3 critical molecules

- DNAs
 - Hold information on how cell works
- RNAs
 - Act to transfer short pieces of information to different parts of cell
 - Provide templates to synthesize into protein
- Proteins
 - Form enzymes that send signals to other cells and regulate gene activity
 - Form body's major components (e.g. hair, skin, etc.)

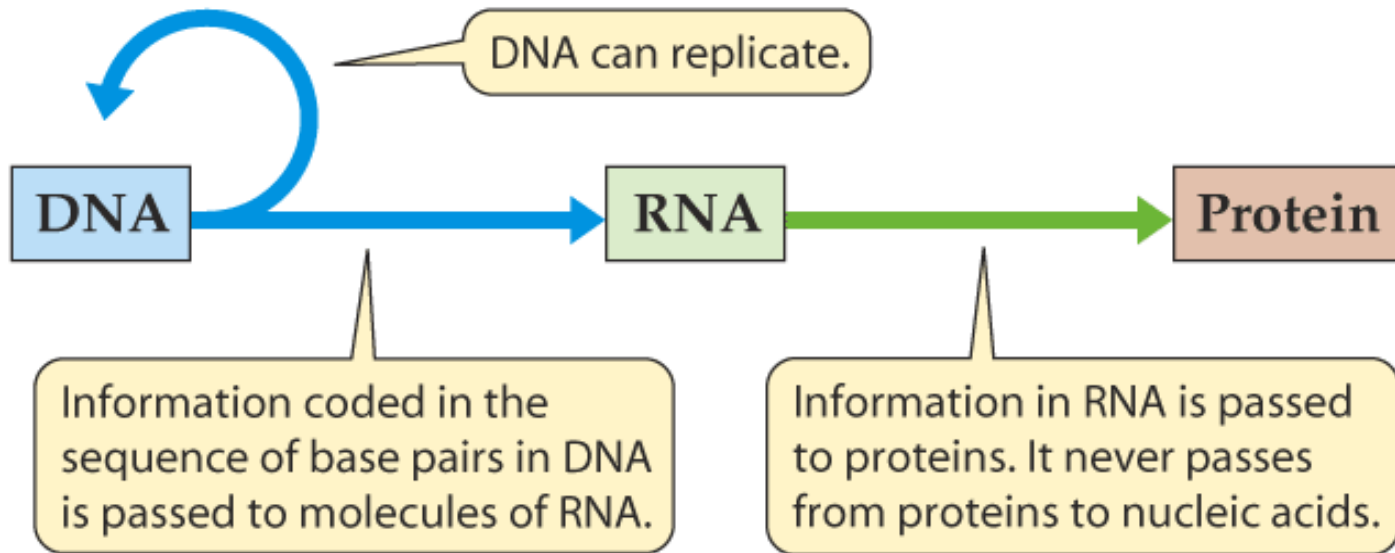


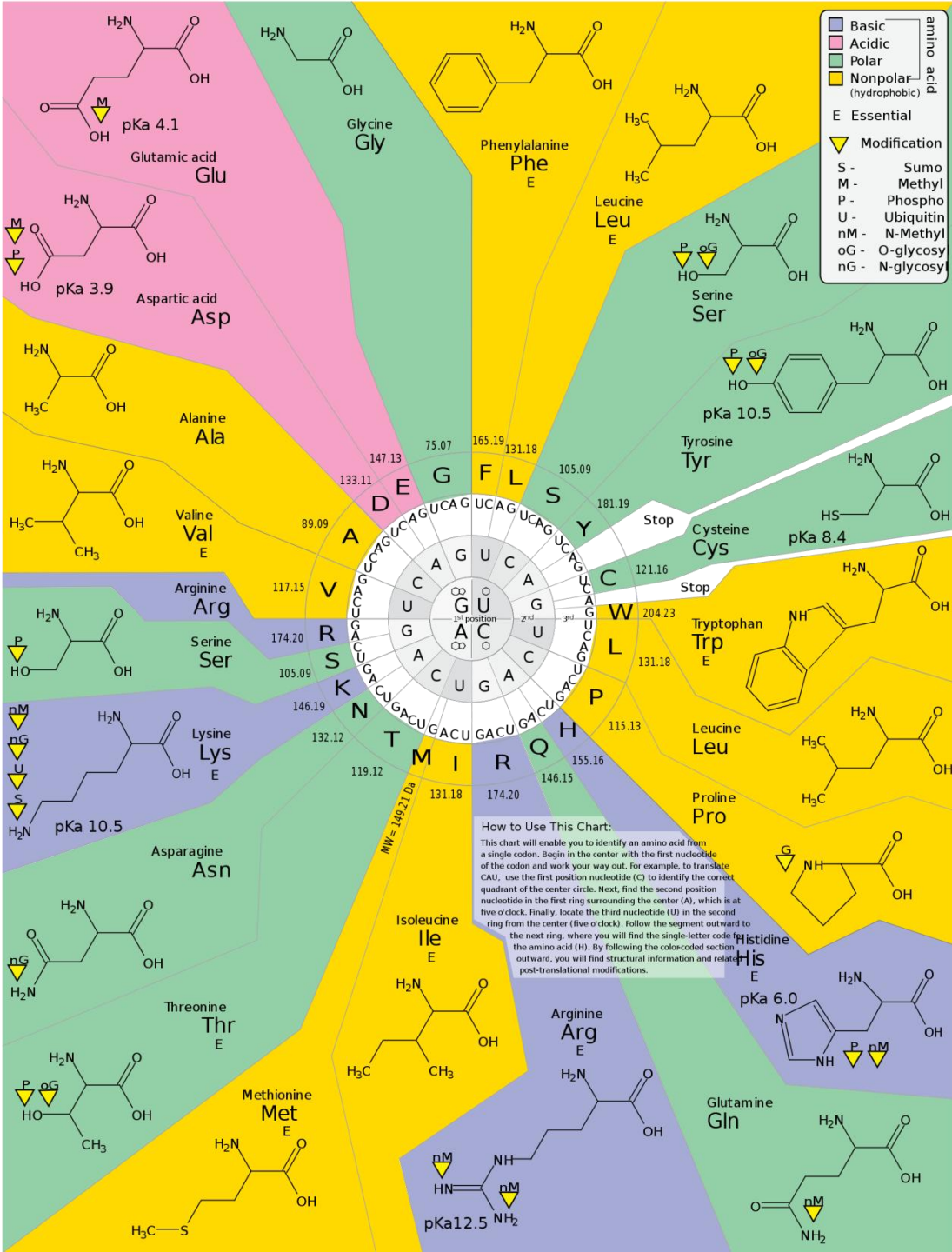
DNA: The Code of Life



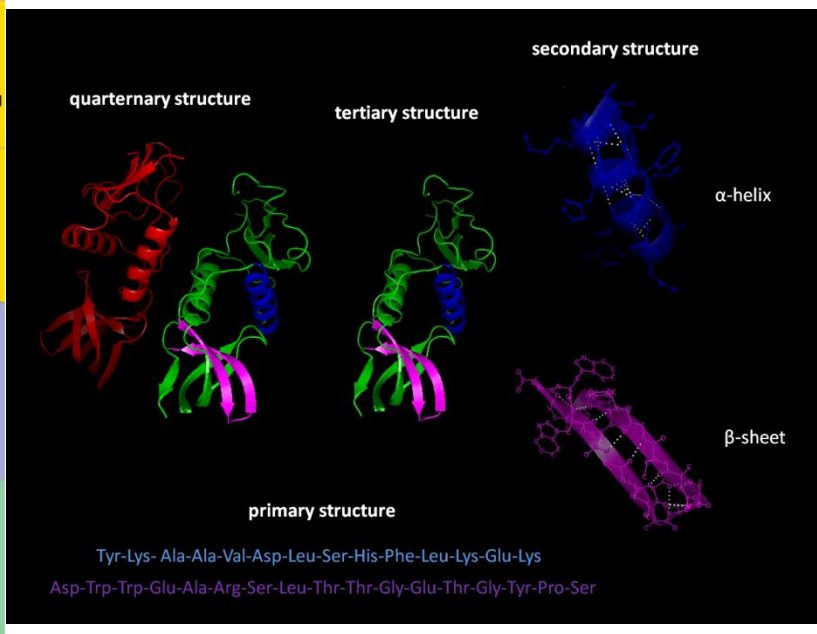
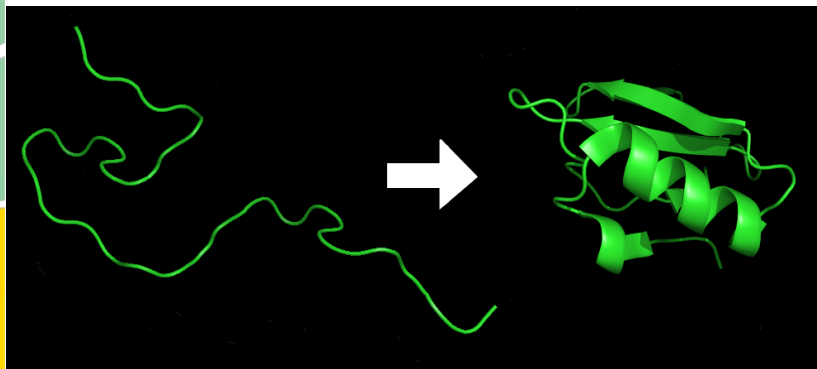
- The structure and the four genomic letters code for all living organisms
- Adenine, Guanine, Thymine, and Cytosine which pair A-T and C-G on complimentary strands.

DNA, RNA, and the Flow of Information (Central Dogma of molecular biology)



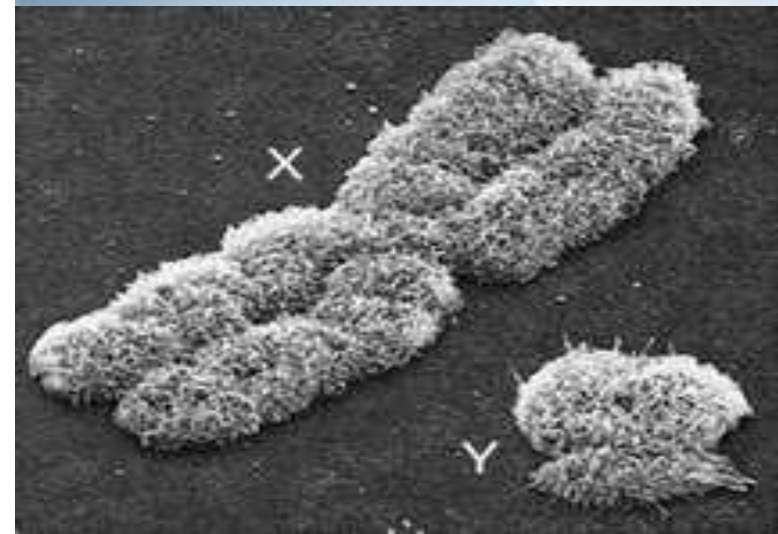
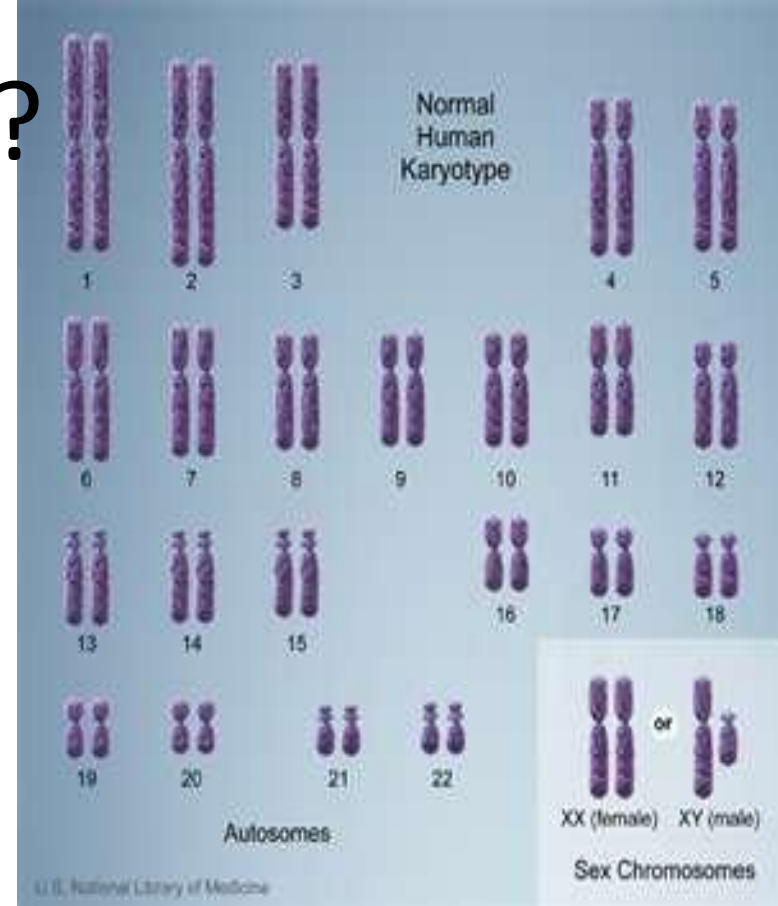


Post-Translational Modifications and Protein Folding



What makes us human?

- Analyze human chromosome...
- Karotype
 - Picture of all the chromosomes in an organism
 - Autosomes
 - Chromosomes 1~44 (pairs 1~22)
 - A.k.a. autosomal chromosomes
 - Sex chromosomes
 - Determine a person's sex (male XY or female XX)
 - Chromosome 45 and 46 (set 23)



15 February 2001

nature

\$16.00

www.nature.com

the human genome

Nuclear fission

Five-dimensional
energy landscapes

Seafloor spreading

The view from under
the Arctic ice

Career prospects

Sequence creates new
opportunities

naturejobs

genomics special

Science

16 February 2001

Vol. 291 No. 5507
Pages 1145-1434 \$9

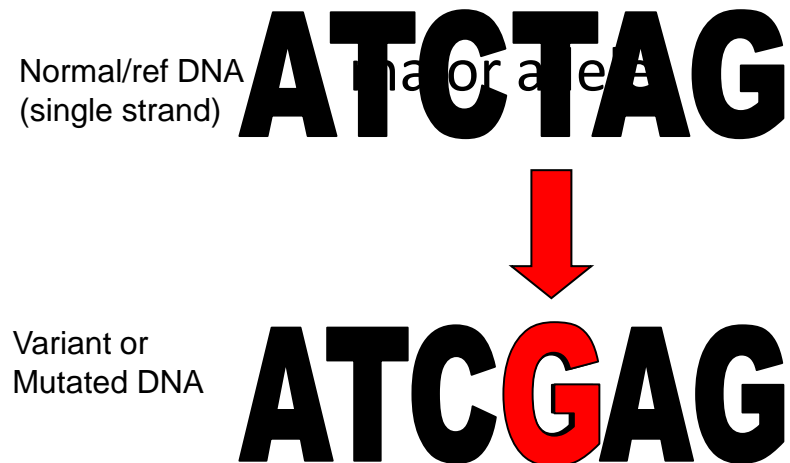
THE HUMAN GENOME



AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

Compare to the reference ...

- Terms for variation in genomes
 - Single Nucleotide Polymorphism/Variant (SNP/SNV)
 - Insertion-deletion (indel; mostly short, e.g. <10bp)
 - Copy Number Variation/Variant (CNV): duplication, deletion
 - Structural Variation/Variant (SV): translocation, inversion, ... (Feuk *et al.* 2006 *Nat. Rev. Genet.*)
 - “Normal/wild type/mutation negative/mutation free/variant free”: with regard to the diploid (2X) reference genome
- The most frequent variations are bi-allelic SNPs:
and minor allele



Single variants: the Good, the Bad, and the Silent

- Single variants can serve the organism in three ways:
- **The Good:** A mutation can cause a trait that enhances the organism's function: Mutation in the sickle cell gene provides resistance to malaria.
- **The Bad:** A mutation can cause a trait that is harmful, sometimes fatal to the organism: Huntington's disease, a symptom of a gene mutation, is a degenerative disease of the nervous system.
- **The Silent:** A mutation can simply cause no difference in the function of the organism. (Vast majority of human genetic variation, population specific)

Technologies to detect variation

- Genotyping
 - Microarray, hybridization based, highly automated (Illumina Omni/Infinium, Affymetrix Axiom, ExomeChip etc.)
 - Predefined SNPs, a few hundred thousand ~ several million markers
 - Can also detect CNVs (Copy Number Variations)
 - GWAS arrays: for genome-wide association studies (GWAS), mostly contain common SNPs (minor allele frequency [MAF] > 5% in the general population)
 - Gene-centric arrays: focusing on specific regions (candidate-wide), contain uncommon ($1\% < \text{MAF} < 5\%$) and/or rare SNPs ($\text{MAF} < 1\%$). E.g. CardioChip (IBC array), MetaboChip, ImmunoChip, ForensicChip, TransplantChip etc.

Technologies to detect variation

- Traditional Sanger sequencing
 - Primer design -> PCR -> microfluidic capillary sequencing -> chromatograms
 - Semi-automated, max length ~1000bp, can detect novel variants
- Next generation sequencing
 - Began in early 2000 (e.g. Solexa which is originated from Cambridge U)
 - Illumina GA/HiSeq/MiSeq, Life Tech SOLiD/Ion Torrent, Complete Genomics, Roche 454 (discontinued) and others (PacBio)
 - Whole Genome Sequencing (WGS) and Whole Exome Sequencing (WES): minimal prior information, thus could lead to new discoveries
 - Exome capture: microarray assisted, coding regions only (Agilent/Nimblegen/Illumina ...)
 - High throughput, high accuracy, low cost, large amount of data (typically in the order of terabytes, ie ~1,000,000,000,000 Bytes), and computationally intensive -> from linux/unix cluster to cloud computing

Analysis: Mendelian disease sequencing

- Mendelian inheritance

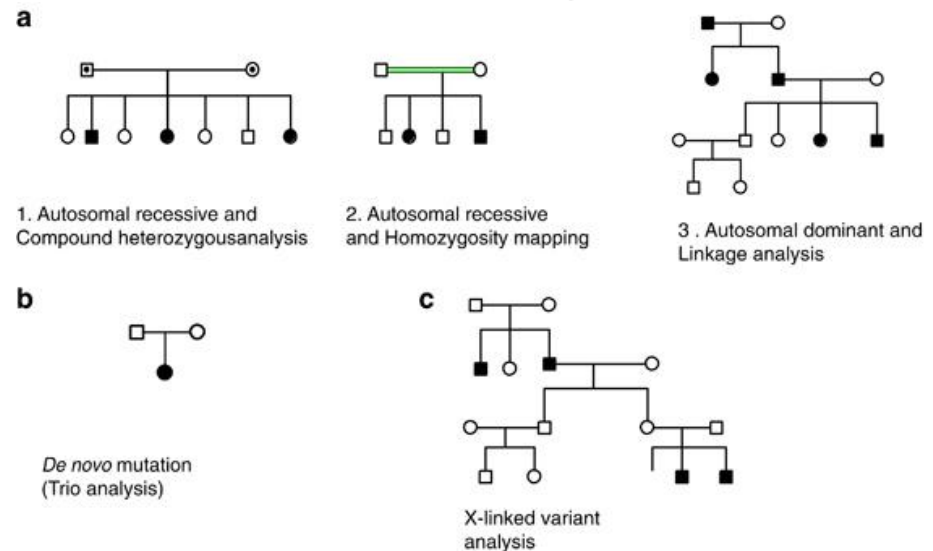
- Autosomal recessive, comp het
- Autosomal dominant, *de novo*
- X-linked recessive
- X-linked dominant
- Y-linked (paternally inherited)
- Mitochondrial (maternally inherited)

- Variable expressivity and incomplete (reduced) penetrance

- Caused by single mutations (a.k.a. monogenic disorders), mostly in the coding region of exons, thus alters protein structure/function

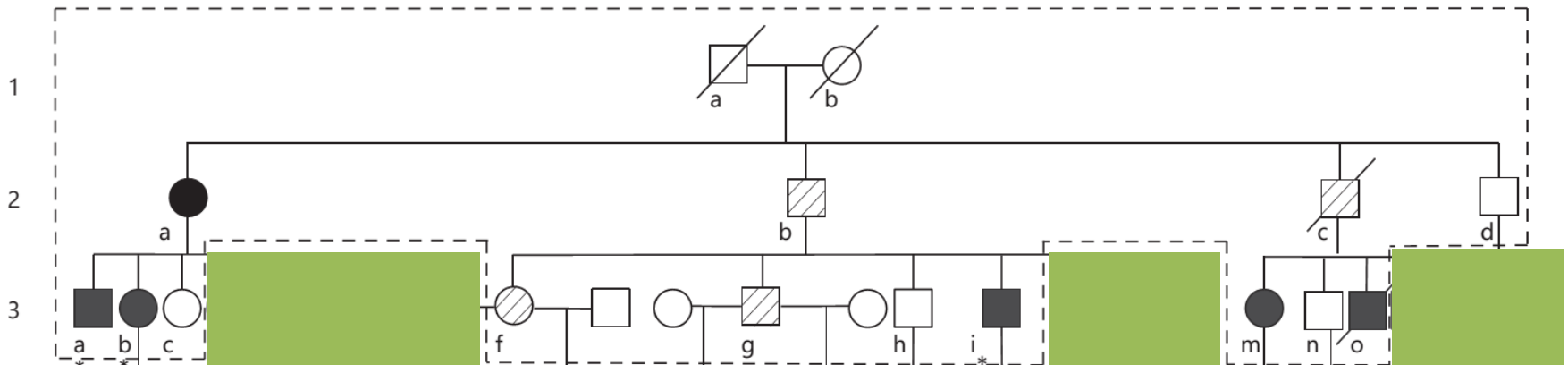
- Traditional method of linkage analysis using genetic markers, then trace the mutation across generations

- low resolution due to marker size (over several hundred/thousand , even million bps) and marker location
- requires large multi-generation pedigrees



Analysis: Mendelian disease sequencing

- The disease is partially resolved ~60 years after the original report on the pedigree within dashed lines



Analysis: Mendelian disease sequencing

- NGS based method, esp. WES
 - exome-wide screening, base-pair resolution
 - small families are fine (trios, even multiple sporadic cases as for TMAU)
- General procedure
 - Raw fastq files (short read of ~100bp)
 - Alignment to reference genome using BWA, generating BAM files
 - Variant calling across BAM files using GATK (checking BAM/variants using IGV see right)



Analysis: Mendelian disease sequencing

- General procedure

- Sequencing
- Alignment/short read mapping back to the reference genome assembly
- Variant calling
- Variant annotation according to gene sets
- Variant filtering based on frequency, inheritance, previous report, biological relevance, gene-gene interaction and other evidences

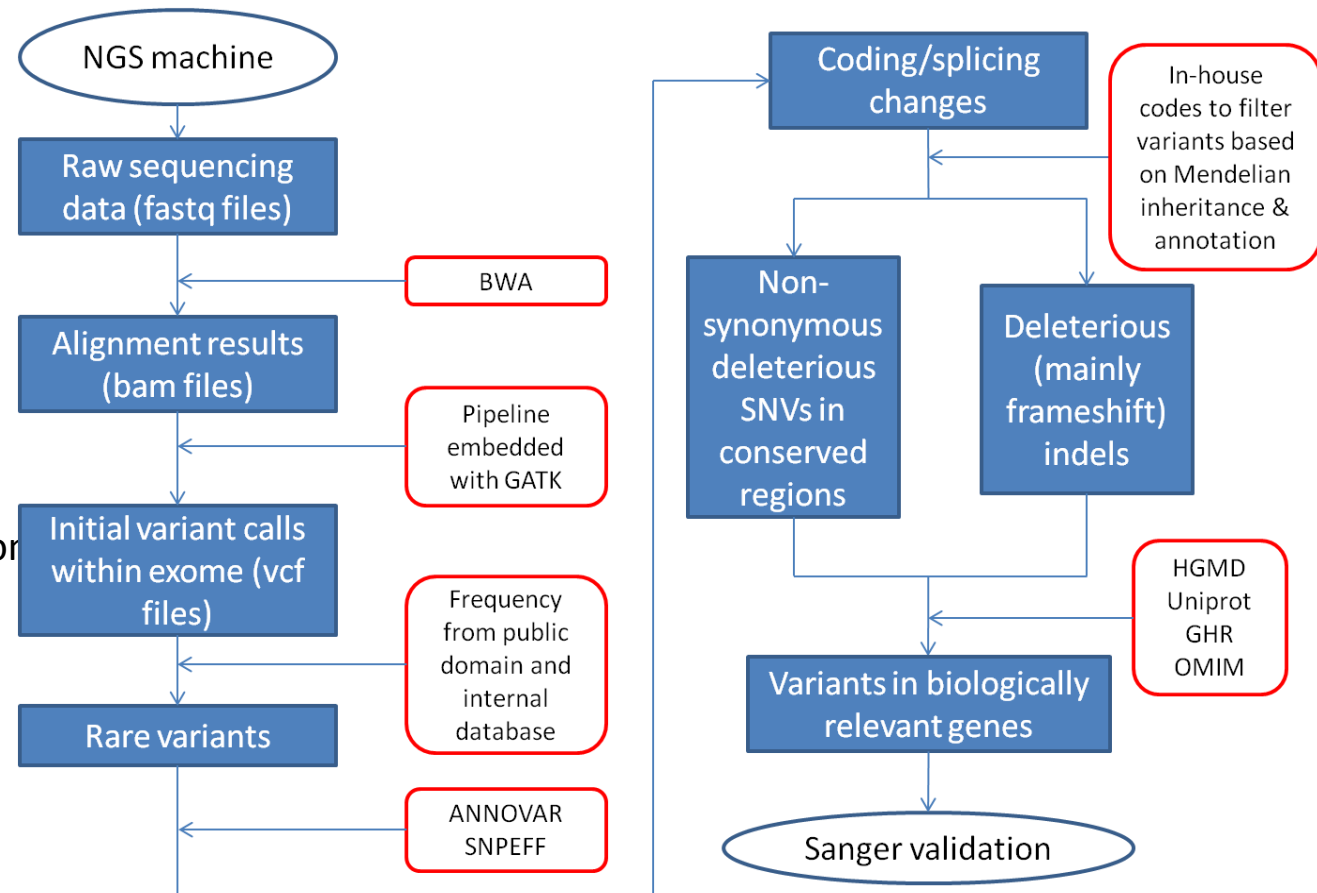


Figure 2. Variant filtering pipeline in exome sequencing data analysis.

Analysis: Mendelian disease sequencing

- RD1000
 - A large rare disease sequencing effort in progress, collaboration between
 - BGI (providing free sequencing)
 - CHOP (coordination and analysis), and
 - multiple sample sources from US as well as around the world (Australia, Italy, Netherlands, Saudi Arabia, Spain ...)
 - Phenotypic information: pedigrees, diagnosis, medical records, disease specific knowledge
 - Genotypic data: raw fastq files, alignment, variants called (with annotation)
 - Authorship in manuscripts
 - Opportunity to grant applications
- TMAU genetics study is part of RD1000

TMAU genetics

- Trimethylamine (TMA) has unpleasant odor of fish
- choline $\xrightarrow{\text{gut bacteria}}$ TMA $\xrightarrow{\text{FMO3 (human liver enzyme flavin containing monooxygenase 3)}}$ TMA-N-oxide (TMAO; odorless)
- Specifically variants in the FMO3 gene are related to impaired TMA metabolism, and genetic variants of the *FMO3* gene can inactivate or lower the oxygenation efficacy, resulting the odorous TMA in urine (TMAU), sweat and other body fluids
- FMO3 variants don't explain all TMAU cases: other genes could contribute too, especially those within the same/similar pathway of FMO3

TMAU genetics

- Thus we conducted the current study to investigate the genetics of TMAU
 - odor evaluation by a trained sensory panel
 - analysis of their urine concentration of TMA relative to TMAO before and after choline ingestion
 - whole exome sequencing (WES) as well as subsequent variant analysis of all 10 samples
- Published in **BMC Medical Genetics** (2017 Feb 15)
Genetic analysis of impaired trimethylamine metabolism using whole exome sequencing
<https://bmcmmedgenet.biomedcentral.com/articles/10.1186/s12881-017-0369-8>

Yiran Guo[†] ✉, Liang-Dar Hwang[†], Jiankang Li[†], Jason Eades, Chung Wen Yu, Corrine Mansfield, Alexis Burdick-Will, Xiao Chang, Yulan Chen, Fujiko F. Duke, Jianguo Zhang, Steven Fakharzadeh, Paul Fennessey, Brendan J. Keating, Hui Jiang, Hakon Hakonarson, Danielle R. Reed ✉ and George Preti

[†] Contributed equally

TMAU patients

We informed adult subjects with body odor complaints who contacted the study investigators about participating in our ongoing sensory and genetic studies. Approximately 130 subjects made contact during an 8-year period from 1999 to 2007 and we evaluated them for the sensory and metabolic arms of the study. From this pool of subjects, we chose ten at random for WES.

Table 1 Characteristics of subjects with body odor complaints

ID	Age (yr)	Sex	Race	TMAO: TMA ^a	Fishy body odor	Fishy oral odor	Other, body odor	Common oral odors
52	78	F	C	0.13	None	None	None	Pungent, sulfurous ^b
114	18	M	AA	0.37	None	None	None	Sulfurous/fecal ^b
122	64	F	AA	0.47	None	None	None	Fecal, pungent ^b
35	45	F	C	0.54	None	None	None	Mild sulfurous ^b
99	59	F	AA	0.58	None	None	None	Mild metallic, smoky
64	51	F	C	0.61	None	None	None	Mild
62	51	F	AA	0.79	None	None	None	Sulfurous ^b
113	47	M	C	0.79	None	None	None	Strong sulfurous ^b
98	44	F	C	0.86	None	None	Musty ^c	Unremarkable
56	70	F	C	0.87	None	None	None	Unremarkable

ID subject identifier number, *age* age at assessment in years (yr), *M* male, *F* female, *AA* American of African descent, *C* Caucasian American of European descent. Race/ethnicity was determined by self-report. ^aTable is sorted by TMAO:TMA ratio; Less than 0.90 is criterion for TMAU (i.e., reference > 0.90). ^bOral odor secondary to plaque located on the posterior dorsal surface of the tongue. Plaque often contains odor-causing bacteria ([7] and references therein). ^cAlso described as 'damp'

Genetic Variants in known genes

```

#contig<ID=GL000225.1,length=211173,assembly=b37>
#contig<ID=GL000192.1,length=547496,assembly=b37>
#contig<ID=NC_007605,length=171822,assembly=b37>
#contig<ID=h3745,length=394779
##reference=file:///mnt/zeilon/ce
  
```

Very rare

.../gatk/bundle/2.8/b37/human_g1k_v37_decoy.fasta

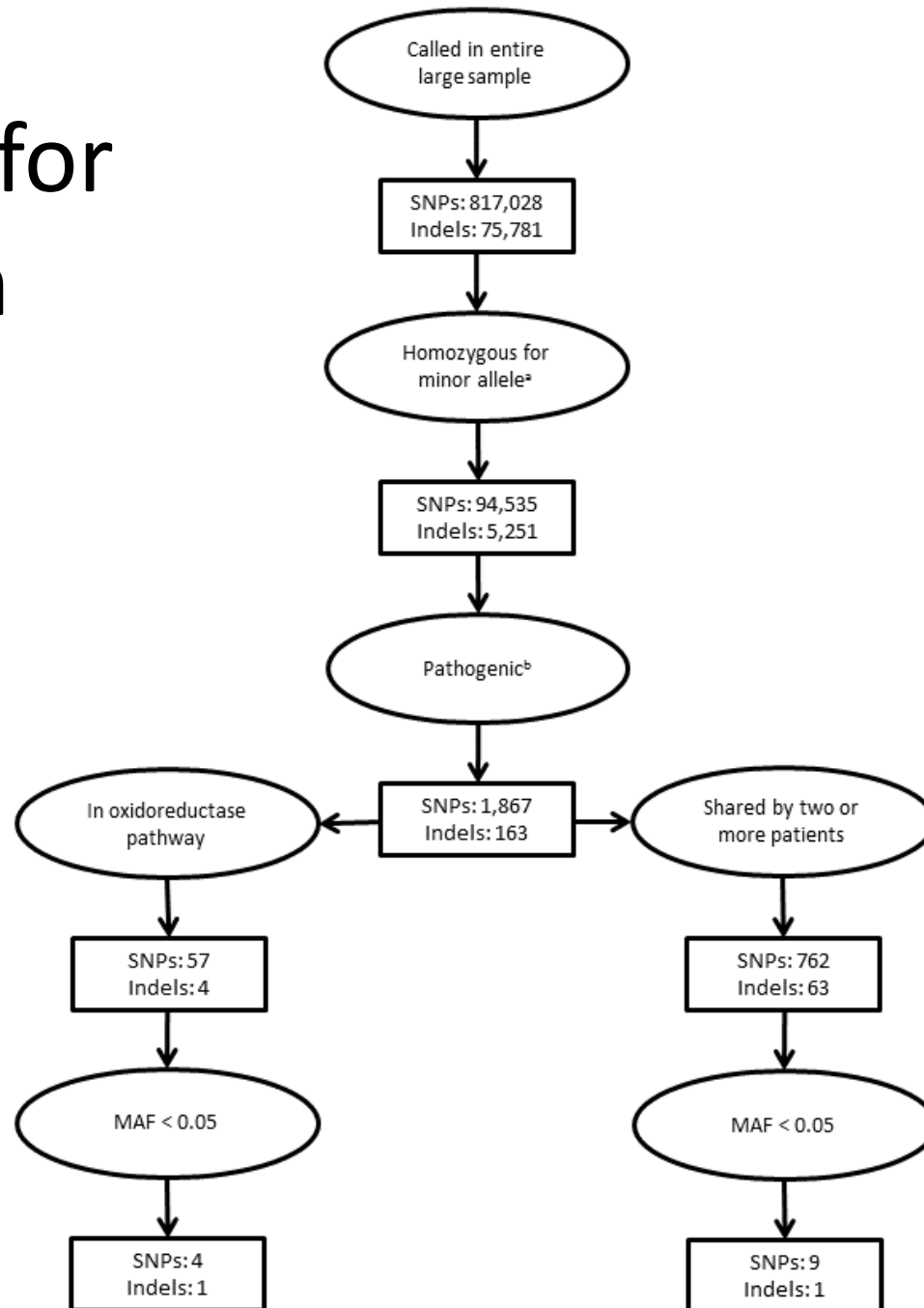
Table 3 Known rare and common pathogenic variants for TMAU observed in this study

Gene	Variants (hg19)	MAF	Amino acid change in HGVS format	Subject ID ^a	Ref
<i>FMO3</i>	rs72549325 chr1:g.171076936G > A/T NM_001002294.2:c.442G > A/T	4.118e-05/ 8.236e-06 ^b	NP_001002294.1:p.Gly148Arg/ NP_001002294.1:p.Gly148Ter	99 [†]	[40]
<i>FMO3</i>	rs2266782 chr1:g.171076966G > A NM_001002294.2:c.472G > A	0.383 ^b	NP_001002294.1:p.Glu158Lys	64 [†] , 98 [†] , 113 [†] , 122 [†]	[37, 50]
<i>FMO3</i>	rs1736557 chr1:g.171080080G > A NM_001002294.2:c.769G > A	0.080 ^b	NP_001002294.1:p.Val257Met	122 [†]	[37, 51]
<i>FMO3</i>	rs2266780 chr1:g.171083242A > G NM_001002294.2:c.923A > G	0.153 ^b	NP_001002294.1:p.Glu308Gly	98 [†]	[50]
<i>PYROXD2</i>	rs7072216 ^c chr10:g.100156853 T > C NM_032709.2:c.625 + 249A > G	0.4012 ^d	Intronic	52 [†] , 56 [†] , 35 [§] , 64 [§] , 113 [§]	[21]

Common in population, and associated with impairments in metabolism

^aZygosity of samples: [†]heterozygous; [§]homozygous
^bFrequency from ExAC database (<http://exac.broadinstitute.org/>)
^cThis intronic variant is not seen in the exome data. We selected it based on literature search and genotyped it in the ten study samples
^dFrequency from the 1000 Genomes Project (<http://www.1000genomes.org/>)

Analysis workflow for variants in unknown genes



Genetic Variants in Unknown Genes

- at least one of the ten subjects was homozygous for the alternative alleles, predicted to be harmful or deleterious by SIFT/PolyPhen-2, and with $MAF < 0.05$; three patients are with the lowest TMAO:TMA ratio
- nine rare putatively pathogenic SNVs and one rare putatively pathogenic indel were shared by two patients
- Gene-gene interaction study found three variants

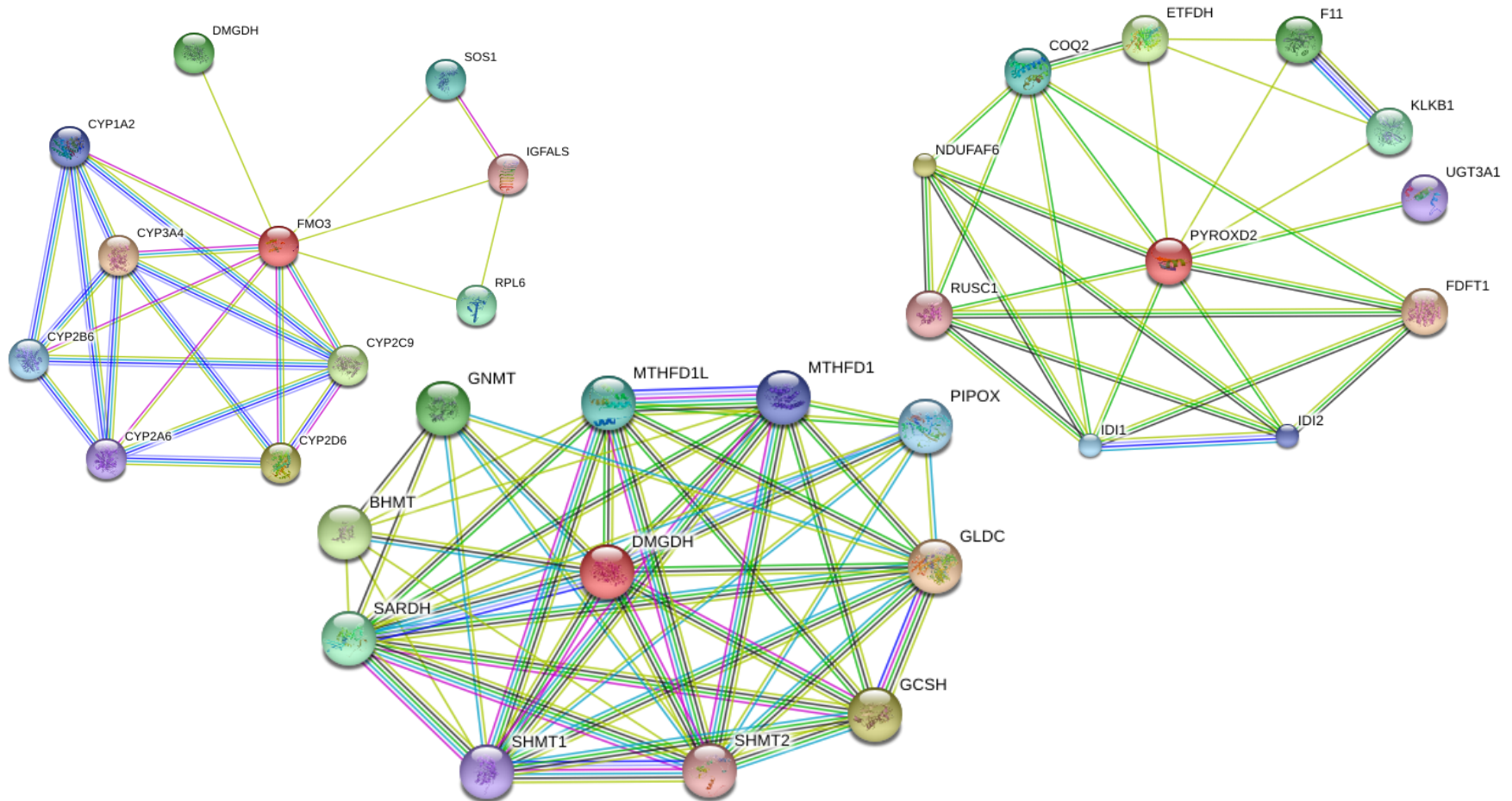
Table 4 Novel genetic variants with potential involvement in TMAU

Variants	Chr	Pos	Gene	MAF	Subject ID ^a
SNPs in oxidoreductase pathways					
rs61733458	3	148,916,215	<i>CP</i>	0.0110	52 ←
rs34625494	17	41,002,169	<i>AOC2</i>	0.0032	62
rs72947591	18	9,887,167	<i>TXNDC2</i>	0.0179	52 ←
rs116368403	19	41,600,254	<i>CYP2A13</i>	0.0046	122 ←
Indels in oxidoreductase pathways					
1 bp insertion	10	102,295,637	<i>HIF1AN</i>	0.0200	114 ←

SNPs shared by at least two subjects						
rs73891273	3	196,235,191	<i>SMCO1</i>	0.0445	62 & 99	
rs77469804	6	110,679,450	<i>METTL24</i>	0.0262	114 & 122	
rs77749341	7	149,462,317	<i>ZNF467</i>	0.0142	99 & 114	
rs7091756	10	1,094,906	<i>IDI1</i>	0.0207	35 & 56	
rs7956250	12	93,966,693	<i>SOCS2</i>	0.0257	99 & 122	
rs55739813	15	41,803,754	<i>LTK</i>	0.0367	35 & 52	
rs138735905	19	17,638,121	<i>FAM129C</i>	0.0193	64 & 98	
rs114989947	22	17,265,194	<i>XKR3</i>	0.0344	99 & 113	
rs41305431	X	103,495,552	<i>ESX1</i>	0.0266	113 & 114	
Indels shared by at least two subjects						
1 bp insertion	—	1	158,533,298	<i>OR6P1</i>	0.0100	62 & 122

rs#: reference SNP identifier (does not apply to indels). *Chr* Chromosome, *Pos* base pair position in map GRCh37/hg19, *MAF* minor allele frequency ^aAll subjects are homozygous for the minor allele

Gene-gene interaction



Confirmation and follow up

- Confirmation by Sanger and Taqman sequencing
 - concordance rate of 100%
- More study is needed, and we are seeking funding opportunities to test DNA of another 100 TMAU patients in order to identify new genes involved in this disorder

Acknowledgement

- Colleagues in the Center for Applied Genomics at the Children's Hospital of Philadelphia, esp. director Dr. Hakon Hakonarson
- Collaborators at Monell Chemical Senses Center (George, Danielle, Corrine, Daniel, and others)
- BGI – for continuous collaboration in rare disease research

Thank you and any questions?