

# Common Conventions

BIG  
BIO

*Sam Jensen*

**BIG  
BIO** THANKS

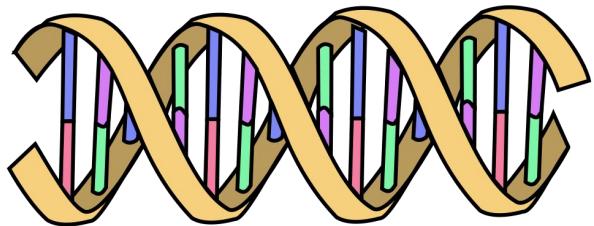
**UCLA** Computational Medicine

# REVIEW





# REVIEW

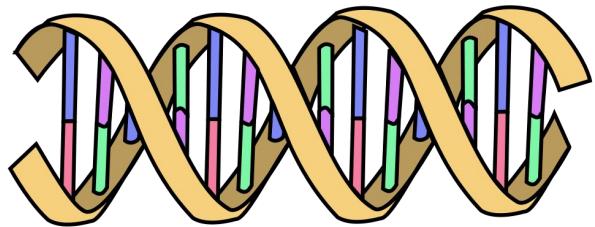


...CTCGTCACTTCACGTATG...

||| ||| ||| ||| ||| |||

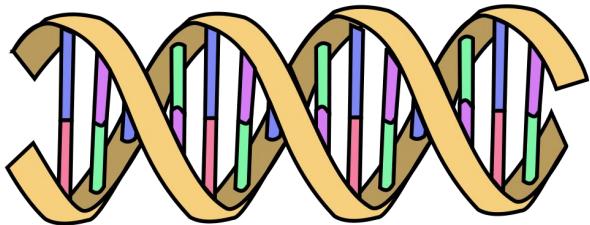
...GAGCAGTGAAGTGCATAC...

# REVIEW

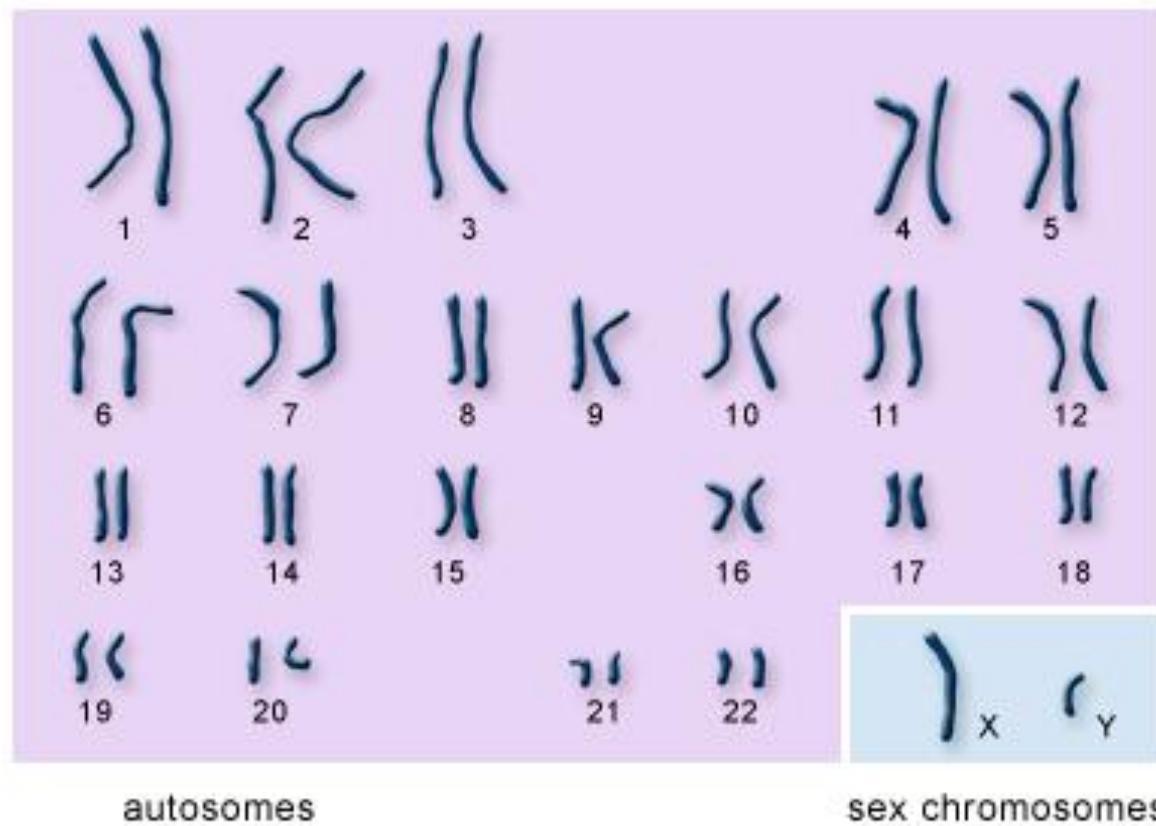


...CTCGTCACTTCACGTATG...

# REVIEW



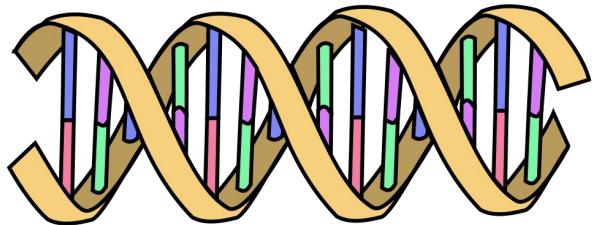
...CTCGTCACTTCACGTATG...



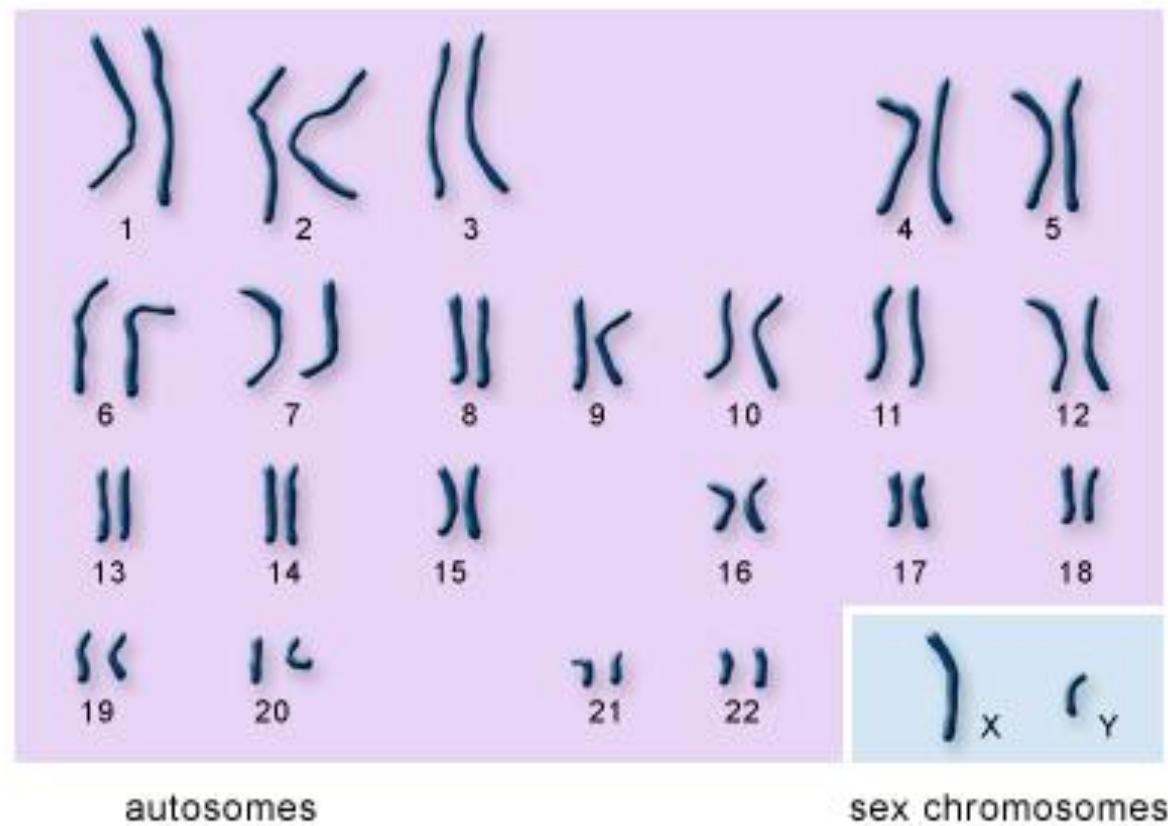
autosomes

sex chromosomes

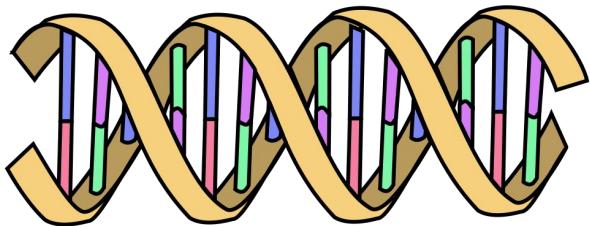
# REVIEW



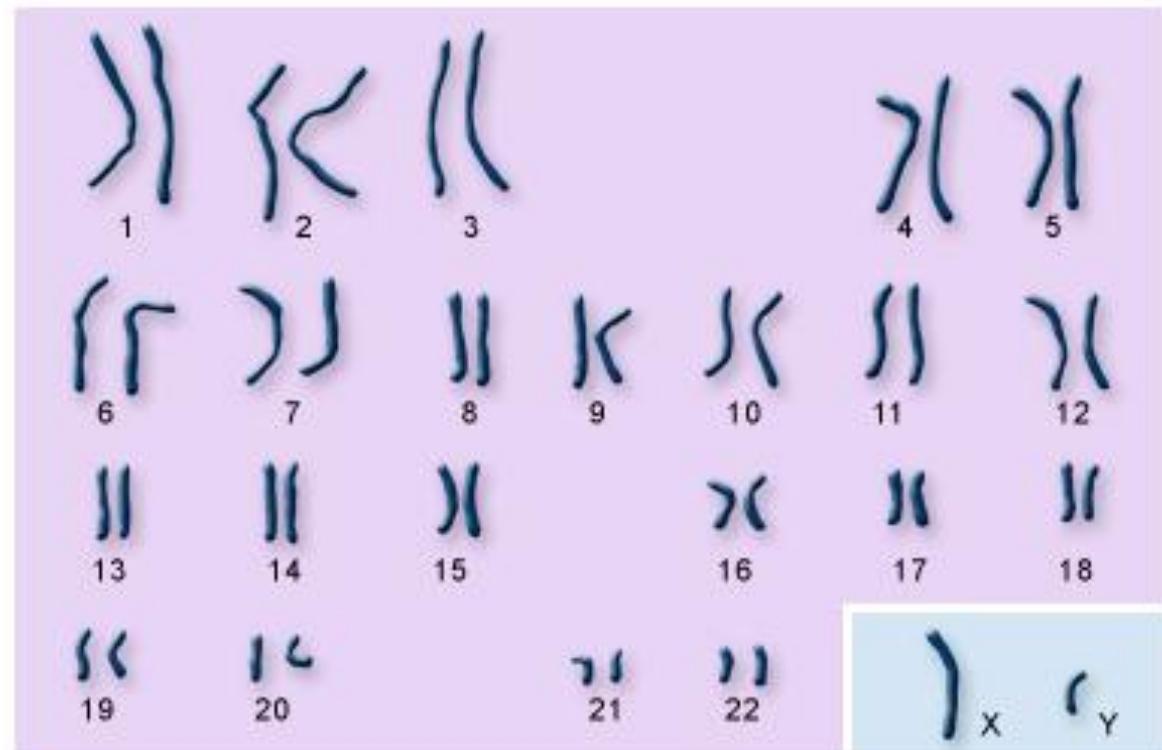
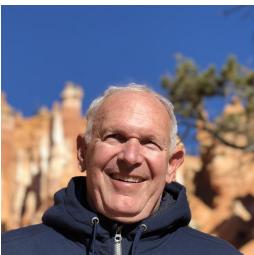
...CTCGTCACTTCACGTATG...



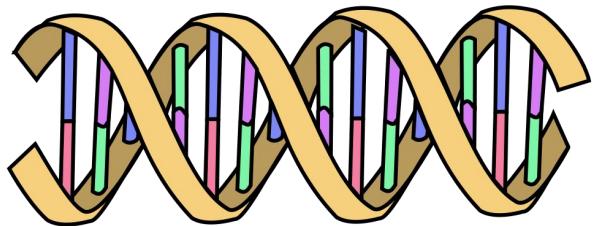
# REVIEW



...CTCGTCACTTCACGTATG...



# REVIEW



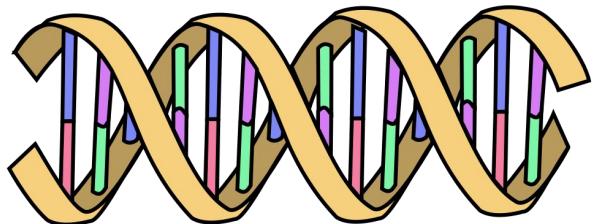
...CTCGTCACTTCACGTATG...



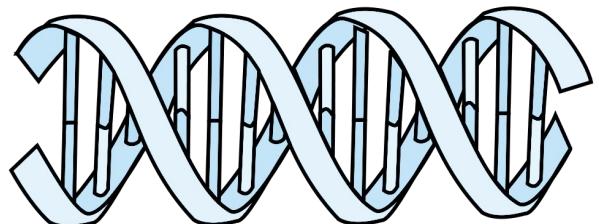
...CACGTCACTTCACGTATG...

...CTCCTCTCATCAC---TG...

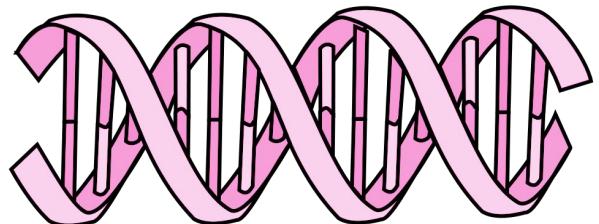
# REVIEW



...CTCGTCACTTCACGTATG...

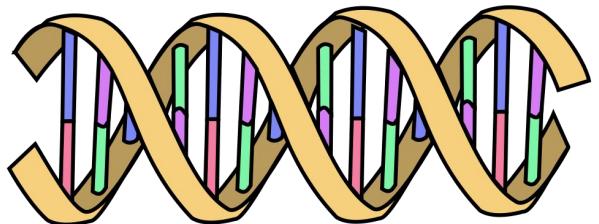


...CACGTCACTTCACGTATG...

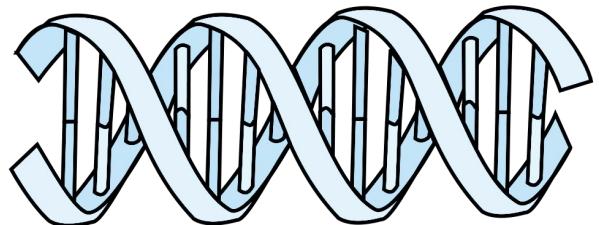


...CTCCTCTCATCAC---TG...

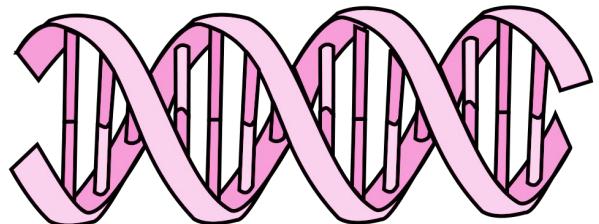
# REVIEW



...CTCGTCACTTCACGTATG...

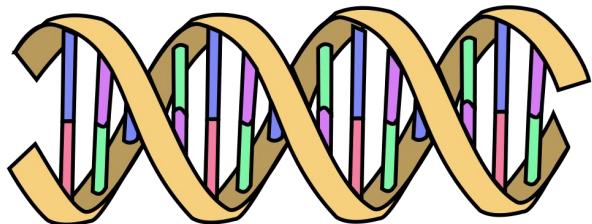


...CACGTCACTTCACGTATG...

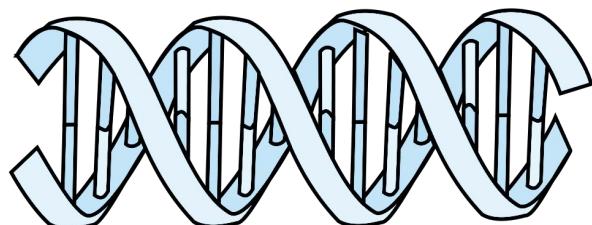


...CTCCTCTCATCAC---TG...

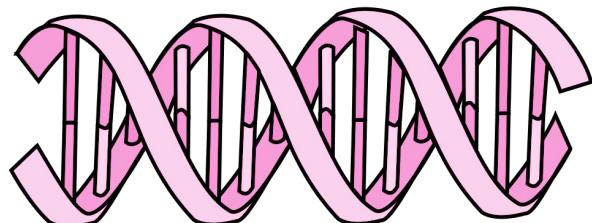
# REVIEW



...CTCGTCACTTCACGTATG...



...**C**ACGTCACTTCACGTATG...



...CTC**C**TCT**C**A**T**CAC - - - TG...



Pos 2



Pos 4

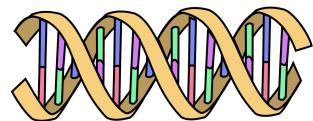


Pos 7



Pos 14

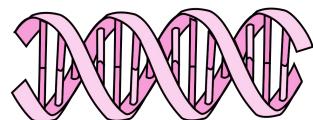
# REVIEW



...CTCGTCACTTCACGTATG...



...C<sub>A</sub>CGTCACTTCACGTATG...



...CTC<sub>C</sub>TCT<sub>C</sub>ATCAC - - - TG...

Pos 2

Pos 4

Pos 7

Pos 14

*Unphased*

T/A      G/C      ACT/TCA      - - - /GTA

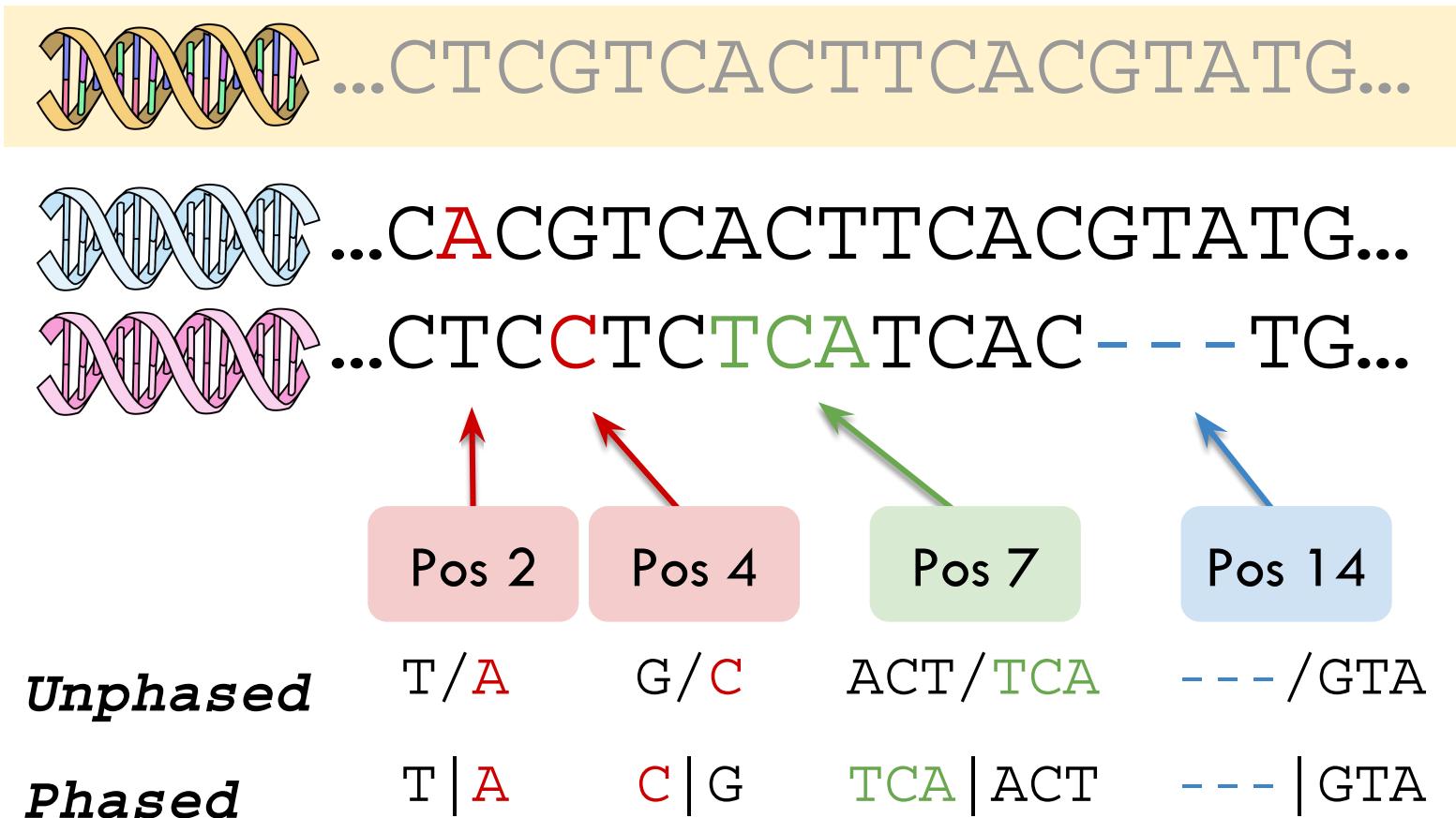
*Phased*

T | A      C | G      TCA | ACT      - - - | GTA

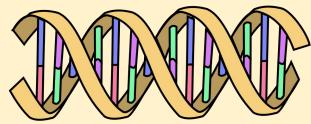
# REFERENCE GENOME



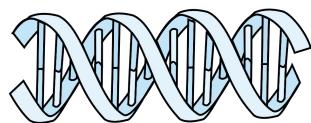
# REFERENCE GENOME



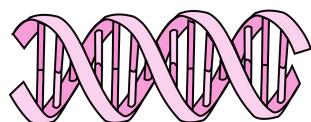
# REFERENCE GENOME



...CTCGTCACTTCACGTATG...



...CAGTCACTTCACGTATG...



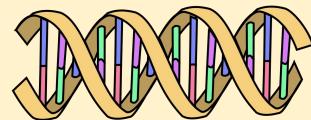
...CTCCCTCTCAATCAC - - TG...

<b>GRC</b>	<b>UCSC</b>	<b>Year</b>
1. NCBI34	hg16	2003
2. NCBI35	hg17	2004
3. NCBI36	hg18	2006
4. GRCh37	hg19	2009
5. GRCh38	hg38	2014

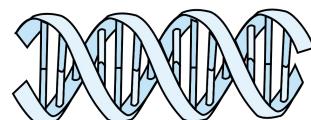


printout of human reference genome  
Wellcome Collection, London

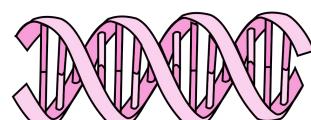
# REFERENCE GENOME



...CTCGTCACTTCACGTATG...



...CACGTCACTTCACGTATG...



...CTCCCTCTCAATCAC - - - TG...

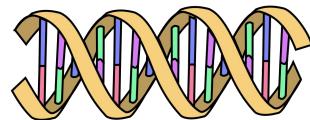
GRC	UCSC	Year
1. NCBI34	hg16	2003
2. NCBI35	hg17	2004
3. NCBI36	hg18	2006
4. GRCh37	hg19	2009
5. GRCh38	hg38	2014

**Reference  
genomes do  
not represent  
the genome of  
ONE person.**

# ALLELES



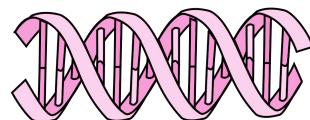
# ALLELES



...CTCGTCACTTCACGTATG...



...C~~A~~CGTCACTTCACGTATG...



...CTC~~C~~TCT~~C~~ATCAC --- TG...

Pos 2

Pos 4

Pos 7

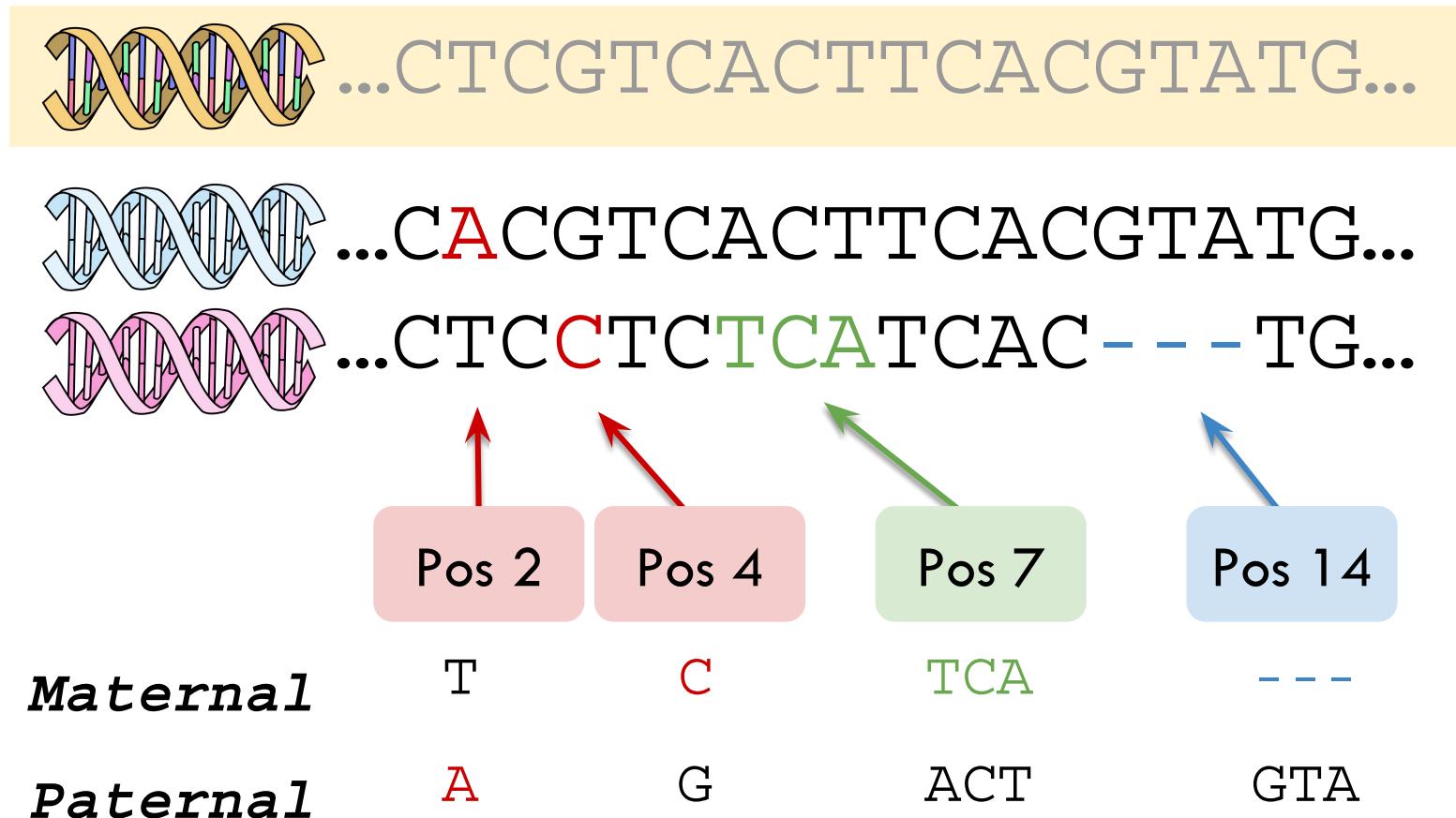
Pos 14

*Allele 1*      T                    G                    ACT                    ---

*Allele 2*      A                    C                    TCA                    GTA

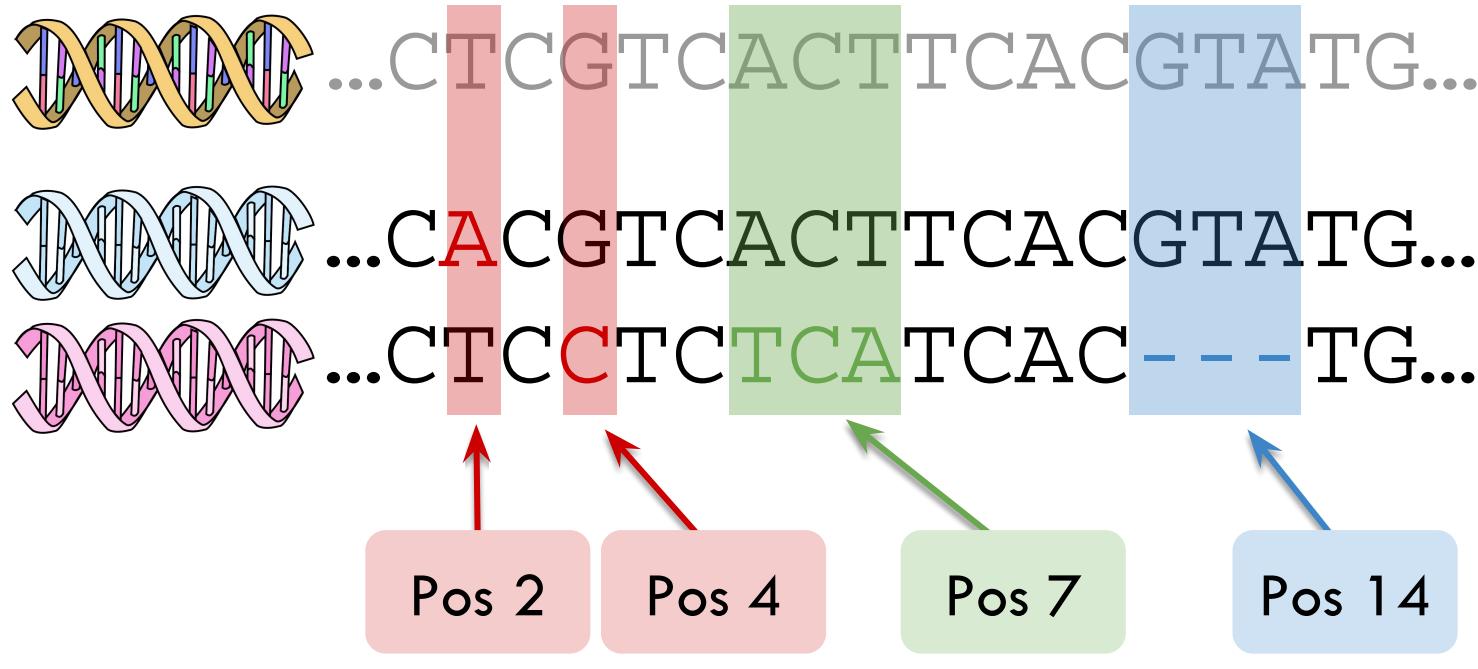
How can I refer to these alleles?

# ALLELES



How can I refer to these alleles?

# ALLELES



Reference	T	G	ACT	GTA
Alternate	A	C	TCA	---

How can I refer to these alleles?

# ALLELES



...CTCGTCACTTCTC---TG...

...CACGTCACTTCACCGTATG...

...CTCCCTCTCACTCAC---TG...

Pos 2

Pos 4

Pos 7

Pos 14

*Ancestral*

T

G

ACT

---

*Derived*

A

C

TCA

GTA

How can I refer to these alleles?

# ALLELLE FREQUENCY

---

# ALLELLE FREQUENCY



...CACGTCACTTCACGTATG...

...CTC**C**TCTCATCAC --- TG...



...CTC**C**TCACCTTCACGTATG...

...CTC**C**TCACCTTCAC --- TG...



...CACGT**C**TCACTCACGTATG...

...CACGT**C**TCACTCACGTATG...



...CTC**C**TCACCTTCAC --- TG...

...CTC**C**TCACCTTCAC --- TG...



...CTC**C**TCACCTTCAC --- TG...

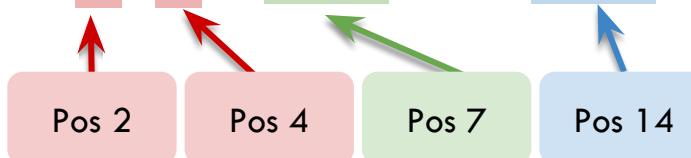
...CAC**C**TCACCTCACGTATG...

Pos 2

Pos 4

Pos 7

Pos 14



# ALLELLE FREQUENCY



...CACGTCACTTCACCGTATG...

Pos 2

...CTCCTCTCATCAC---TG...

Pos 4

...CTCCTCACTTCACCGTATG...

*Allele 1*

T

G

ACT

---

...CTCCTCACTTCAC---TG...

*Allele 2*

A

C

TCA

GTA



...CACGTCTCATCACCGTATG...



...CACGTCTCATCACCGTATG...



...CTCCTCACTTCAC---TG...

...CTCCTCACTTCAC---TG...



...CTCCTCACTTCAC---TG...

...CACCTCACTTCACCGTATG...

Pos 2

Pos 4

Pos 7

Pos 14



# ALLELLE FREQUENCY



...CACGTCACTTCACCGTATG...

Pos 2 Pos 4 Pos 7 Pos 14



...CTCCTCTCATCAC---TG...

Allele 1 T G ACT ---  
Allele 2 A C TCA GTA



...CTCCTCTCATCAC---TG...

Allele 1 6 3 7 5  
Allele 2 4 7 3 5



...CTCCTCTCATCAC---TG...



...CTCCTCTCATCAC---TG...

...CACCTCACTTCACCGTATG...

Pos 2

Pos 4

Pos 7

Pos 14

Pos 2

Pos 4

Pos 7

Pos 14

# ALLELLE FREQUENCY



...CACGTCACTTCACGTATG...  
...CTCCTCTCATCAC---TG...

Pos 2 Pos 4 Pos 7 Pos 14

**Allele 1** T G ACT ---

**Allele 2** A C TCA GTA



...CTCCTCTCATTCACGTATG...  
...CTCCTCATTCAC---TG...

**Allele 1** 6 3 7 5

**Allele 2** 4 7 3 5



...CACGTCTCATCACGTATG...  
...CACGTCTCATCACGTATG...

**Allele 1** 60% 30% 70% 50%

**Allele 2** 40% 70% 30% 50%



...CTCCTCATTCAC---TG...  
...CTCCTCATTCAC---TG...



...CTCCTCATTCAC---TG...  
...CACCTCATTCACGTATG...

Pos 2 Pos 4 Pos 7 Pos 14

# ALLELLE FREQUENCY



...CACGTCACTTCACGTATG...  
...CTCCTCTCATCAC---TG...

Pos 2	Pos 4	Pos 7	Pos 14
-------	-------	-------	--------



...CTCCTCTCATTCACGTATG...  
...CTCCTCATTCAC---TG...

<b>Allele 1</b>	T	G	ACT	---
-----------------	---	---	-----	-----



...CACGTCTCATCACGTATG...  
...CACGTCTCATCACGTATG...

<b>Allele 2</b>	A	C	TCA	GTA
-----------------	---	---	-----	-----



...CTCCTCATTCAC---TG...  
...CTCCTCATTCAC---TG...

<b>Allele 1</b>	6	3	7	5
-----------------	---	---	---	---



...CTCCTCATTCAC---TG...  
...CACCTCATTCACGTATG...

<b>Allele 2</b>	4	7	3	5
-----------------	---	---	---	---

Pos 2      Pos 4      Pos 7      Pos 14

<b>Major</b>	T	C	ACT	---
--------------	---	---	-----	-----

<b>Minor</b>	A	G	TCA	GTA
--------------	---	---	-----	-----

# VARIANT REPRESENTATION



# letters bad, numbers good



C|T  
T|T  
ATC|G--



T|T  
T|A  
ATC|ATC

Chr	Pos	Ref	Alt	Ind1-H1	Ind1-H2	Ind2-H1	Ind2-H2
12	2,147,839	C	T	C	T	T	T
12	2,147,913	T	A	T	T	T	A
12	2,152,882	G--	ATC	ATC	G--	ATC	ATC

Chr	Pos	Ref	Alt	Ind1-H1	Ind1-H2	Ind2-H1	Ind2-H2
12	2,147,839	C	T	0	1	1	1
12	2,147,913	T	A	0	0	0	1
12	2,152,882	G--	ATC	1	0	1	1

# letters bad, numbers good

Haplotype Matrix (Phased necessary)

Chr	Pos	Ref	Alt	Ind1-H1	Ind1-H2	Ind2-H1	Ind2-H2
12	2,147,839	C	T	0	1	1	1
12	2,147,913	T	A	0	0	0	1
12	2,152,882	G--	ATC	1	0	1	1

Genotype Matrix (Unphased or Phased)

Chr	Pos	Ref	Alt	Ind1	Ind2
12	2,147,839	C	T	1	2
12	2,147,913	T	A	0	1
12	2,152,882	G--	ATC	1	2

Other column options: Ancestral Allele, Derived Allele, rsID, genome feature, error

# VCF files

```
##fileformat=VCFv4.0
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=1000GenomesPilot-NCBI36
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=.,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT NA00001 NA00002 NA00003
20 14370 rs6054257 G A 29 PASS NS=3;DP=14;AF=0.5;DB;H2 GT:GQ:DP:HQ 0|0:48:1:51,51 1|0:48:8:51,51 1/1:43:5:...
20 17330 . T A 3 q10 NS=3;DP=11;AF=0.017 GT:GQ:DP:HQ 0|0:49:3:58,50 0|1:3:5:65,3 0/0:41:3
20 1110696 rs6040355 A G,T 67 PASS NS=2;DP=10;AF=0.333,0.667;AA=T;DB GT:GQ:DP:HQ 1|2:21:6:23,27 2|1:2:0:18,2 2/2:35:4
20 1230237 . T . 47 PASS NS=3;DP=13;AA=T GT:GQ:DP:HQ 0|0:54:7:56,60 0|0:48:4:51,51 0/0:61:2
20 1234567 microsat1 GTCT G,GTACT 50 PASS NS=3;DP=9;AA=G GT:GQ:DP 0/1:35:4 0/2:17:2 1/1:40:3
```

# VCF files

```
##fileformat=VCFv4.0
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=1000GenomesPilot-NCBI36
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=.,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">
```

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	NA00001	NA00002	NA00003
20	14370	rs6054257	G	A	29	PASS	NS=3;DP=14;AF=0.5;DB;H2	GT:GQ:DP:HQ	0 0:48:1:51,51	1 0:48:8:51,51	1/1:43:5:...
20	17330	.	T	A	3	q10	NS=3;DP=11;AF=0.017	GT:GQ:DP:HQ	0 0:49:3:58,50	0 1:3:5:65,3	0/0:41:3
20	1110696	rs6040355	A	G,T	67	PASS	NS=2;DP=10;AF=0.333,0.667;AA=T	DB GT:GQ:DP:HQ	1 2:21:6:23,27	2 1:2:0:18,2	2/2:35:4
20	1230237	.	T	.	47	PASS	NS=3;DP=13;AA=T	GT:GQ:DP:HQ	0 0:54:7:56,60	0 0:48:4:51,51	0/0:61:2
20	1234567	microsat1	GTCT	G,GTACT	50	PASS	NS=3;DP=9;AA=G	GT:GQ:DP	0/1:35:4	0/2:17:2	1/1:40:3

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT
20	14370	rs6054257	G	A	29	PASS	NS=3;DP=14;AF=0.5;DB;H2	GT
20	17330	.	T	A	3	q10	NS=3;DP=11;AF=0.017	GT
20	1110696	rs6040355	A	G,T	67	PASS	NS=2;DP=10;AF=0.333,0.667;AA=T;DB	GT
20	1230237	.	T	.	47	PASS	NS=3;DP=13;AA=T	GT
20	1234567	microsat1	GTCT	G,GTACT	50	PASS	NS=3;DP=9;AA=G	GT

# VCF files

```
##fileformat=VCFv4.0
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=1000GenomesPilot-NCBI36
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=.,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT NA00001 NA00002 NA00003
20 14370 rs6054257 G A 29 PASS NS=3;DP=14;AF=0.5;DB;H2 GT:GQ:DP:HQ 0|0:48:1:51,51 1|0:48:8:51,51 1/1:43:5;...
20 17330 . T A 3 q10 NS=3;DP=11;AF=0.017 GT:GQ:DP:HQ 0|0:49:3:58,50 0|1:3:5:65,3 0/0:41:3
20
20
20
20
20
##reference=1000GenomesPilot-NCBI36
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=.,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">
```

# VCF files

```
##fileformat=VCFv4.0
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=1000GenomesPilot-NCBI36
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=.,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">
```

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	NA00001	NA00002	NA00003
20	14370	rs6054257	G	A	29	PASS	NS=3;DP=14;AF=0.5;DB;H2	GT:GQ:DP:HQ	0 0:48:1:51,51	1 0:48:8:51,51	1/1:43:5:...
20	17330	.	T	A	3	q10	NS=3;DP=11;AF=0.017	GT:GQ:DP:HQ	0 0:49:3:58,50	0 1:3:5:65,3	0/0:41:3
20	1110696	rs6040355	A	G,T	67	PASS	NS=2;DP=10;AF=0.333,0.667;AA=T	DB GT:GQ:DP:HQ	1 2:21:6:23,27	2 1:2:0:18,2	2/2:35:4
20	1230237	.	T	.	47	PASS	NS=3;DP=13;AA=T	GT:GQ:DP:HQ	0 0:54:7:56,60	0 0:48:4:51,51	0/0:61:2
20	1234567	microsat1	GTCT	G,GTACT	50	PASS	NS=3;DP=9;AA=G	GT:GQ:DP	0/1:35:4	0/2:17:2	1/1:40:3

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT
20	14370	rs6054257	G	A	29	PASS	NS=3;DP=14;AF=0.5;DB;H2	GT
20	17330	.	T	A	3	q10	NS=3;DP=11;AF=0.017	GT
20	1110696	rs6040355	A	G,T	67	PASS	NS=2;DP=10;AF=0.333,0.667;AA=T;DB	GT
20	1230237	.	T	.	47	PASS	NS=3;DP=13;AA=T	GT
20	1234567	microsat1	GTCT	G,GTACT	50	PASS	NS=3;DP=9;AA=G	GT

# VCF files

```

##fileformat=VCFv4.0
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=1000GenomesPilot-NCBI36
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=.,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">
#CHROM POS ID REF ALT QUA FILTER INFO
20 14370 rs6054257 G A 29 PASS NS=3;DP=14;AF=0.5;DB;H2
20 17330 . T A 3 q10 NS=3;DP=11;AF=0.017
20 1110696 rs6040355 A G,T 67 PASS NS=2;DP=10;AF=0.333,0.667;AA=T;DB
20 1230237 . T . 47 PASS NS=3;DP=13;AA=T
20 1234567 microsat1 GTCT G,GTACT 50 PASS NS=3;DP=9;AA=G

```

FORMAT	NA00001	NA00002	NA00003
GT:GQ:DP:HQ	0 0:48:1:51,51 1 0:48:8:51,51 1/1:43:5:...		
GT:GQ:DP:HQ	0 0:49:3:58,50 0 1:3:5:65,3 0/0:41:3		
T;DB	GT:GQ:DP:HQ 1 2:21:6:23,27 2 1:2:0:18,2 2/2:35:4		
GT:GQ:DP:HQ	0 0:54:7:56,60 0 0:48:4:51,51 0/0:61:2		
GT:GQ:DP	0/1:35:4 0/2:17:2 1/1:40:3		

#CHROM	POS	ID	REF	ALT	QUA	FILTER	INFO
20	14370	rs6054257	G	A			
20	17330	.	T	A			
20	1110696	rs6040355	A	G,T			
20	1230237	.	T	.			
20	1234567	microsat1	GTCT	G,GTACT			

FORMAT	NA00001	NA00002	NA00003
GT:GQ:DP:HQ	0 0:48:1:51,51 1 0:48:8:51,51 1/1:43:5:...		
GT:GQ:DP:HQ	0 0:49:3:58,50 0 1:3:5:65,3 0/0:41:3		
T;DB	GT:GQ:DP:HQ 1 2:21:6:23,27 2 1:2:0:18,2 2/2:35:4		
GT:GQ:DP:HQ	0 0:54:7:56,60 0 0:48:4:51,51 0/0:61:2		
GT:GQ:DP	0/1:35:4 0/2:17:2 1/1:40:3		

# VCF files

```
##fileformat=VCFv4.0
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=1000GenomesPilot-NCBI36
##phasing=partial
##INFO=<ID=NS,Number=1>Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1>Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=.,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1>Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0>Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0>Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1>Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1>Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1>Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2>Type=Integer,Description="Haplotype Quality">
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT NA00001 NA00002 NA00003
20 14370 rs6054257 G A 29 PASS NS=3;DP=14;AF=0.5;DB;H2 GT:GQ:DP:HQ 0|0:48:1:51,51 1|0:48:8:51,51 1/1:43:5:...
20 17330 . T A 3 q10 NS=3;DP=11;AF=0.017 GT:GQ:DP:HQ 0|0:49:3:58,50 0|1:3:5:65,3 0/0:41:3
20 1110696 rs6040355 A G,T 67 PASS NS=2;DP=10;AF=0.333,0.667;AA=T;DB GT:GQ:DP:HQ 1|2:21:6:23,27 2|1:2:0:18,2 2/2:35:4
20 1230237 . T . 47 PASS NS=3;DP=13;AA=T GT:GQ:DP:HQ 0|0:54:7:56,60 0|0:48:4:51,51 0/0:61:2
20 1234567 microsat1 GTCT G,GTACT 50 PASS NS=3;DP=9;AA=G GT:GQ:DP 0/1:35:4 0/2:17:2 1/1:40:3
```

- vcftools and bcftools can manipulate and filter and extract information really easily from these files
- Many software tools take vcf files as input to run genome analyses
- vctools and bcftools and also convert these files to easy to use formats!

**BIG  
BIO** THANKS

**UCLA** Computational Medicine