

**2014 Esri International User Conference**

July 14–18, 2014 | San Diego, California

# Using GIS for Delivering Molecular Diagnostics

Alice Rathjen, CEO, Founder DNA Guide

# The MAP of the Human Genome



# State of Art Bioinformatics: Keyword Search Returns Points on the Map

**Genome Reference Consortium**

[GRC Home](#) | [Data](#) | [Help](#) | [Report an Issue](#) | [Contact Us](#) | [Credits](#) | [Curators Only](#)

[Human Overview](#) | [Human Issues under Review](#) | [Human Assembly Data](#) | [Report a Problem](#)

## Human Genome Overview

Information concerning the continuing improvement of the human genome.

The GRC is working hard to provide the best possible reference assembly for human. We do this by both generating multiple representations (alternate loci) for regions that are too complex to be represented by a single path. Additionally, we are releasing regional fixes known as patches. This allows users who are interested in a specific locus to get an improved representation without affecting users who need chromosome coordinate stability.

### Getting Data

GRCh38 (latest major release): [FTP](#)  
Information on regions under review: [FTP](#)  
[Current Tiling Path Files \(TPFs\)](#)

**Transitioning to GRCh38? Try the NCBI Remapping Service, which uses the same assembly-assembly alignments used by the GRC.**

### GRC Blog

**Chromosome 9 peri-centromeric assembly improvement** Apr 04, 2014

**GRCh38: Incorporating Modeled Centromere Sequence** Jan 14, 2014

Centromeres are specialized chromatin structure...

[see all](#)

### Recently Resolved Human Issues

**Human (HG-2040)** Jun 11, 2014

Clone name corrected.

**Human (HG-1956)** Jun 5, 2014

CR381562.10 now has a placement on Chr X between components BX640545 and AL954722 close to the p arm

[see all](#)

◀ Region containing alternate loci

An ideogram representation of the latest human assembly, GRCh38 (not showing unplaced or unlocalized sequences).

# State of Art Bioinformatics:

**Thematic Maps** – European Bioinformatics Institute



# State of Art Bioinformatics:

## Pan and Zoom, Annotation on Tooltip





# State of Art Bioinformatics: Maps Finally Getting Crowded

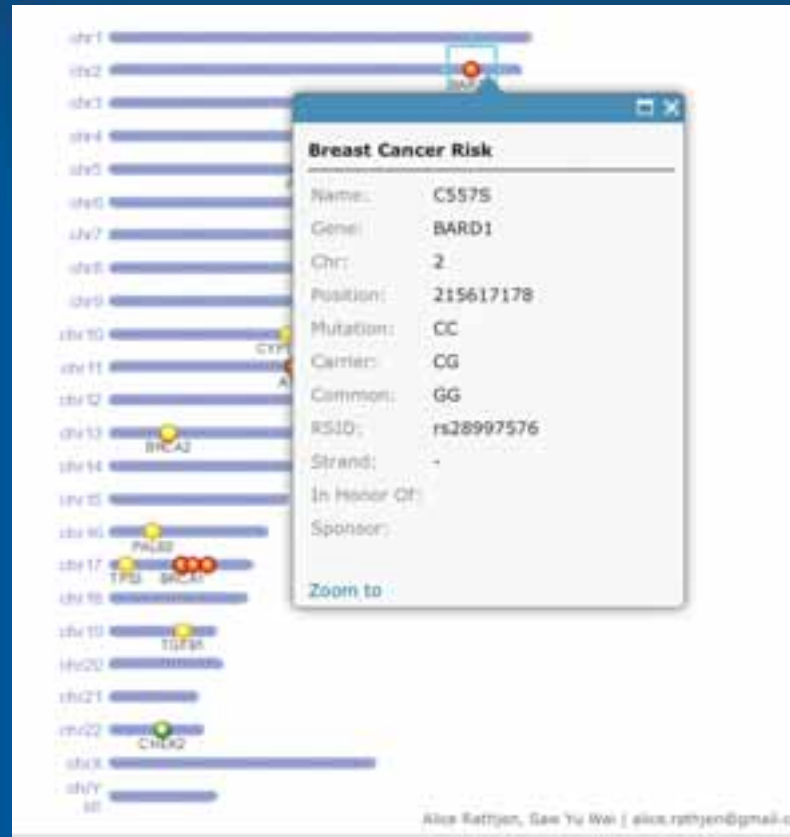


**Rapid Scientific Advances  
Yet, Current Lab Format...**

**Molecular Diagnostic  
Delivered as a Static Text via  
PDF Email Attachment**

# Bringing Genomic Data to the Clinic using GIS

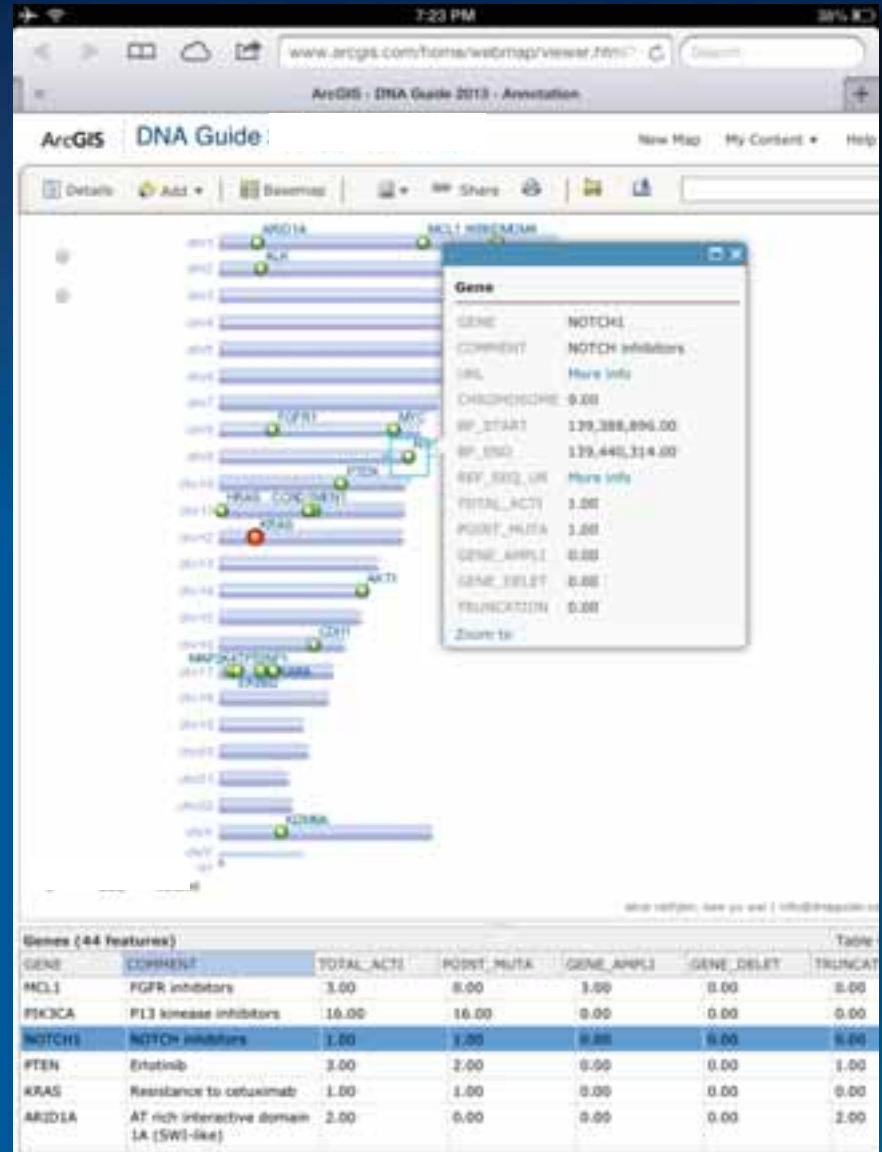
- GIS Already Up and Running for most Large Entities  
Facilities Mapping  
Infrastructure
- Proven Technology for distributing location based information to diverse users
- Rapid Deployment
- Easy Adoption
- Scales Indefinitely





# Types of Molecular Diagnostics

- Cancer Tumor Analysis (200 genes)
- Carrier Testing (1-50 points of data)
- Drug Response (1-15 points of data)
- Disease Risk (1 to 60 points of data per disease)



# Genomic Coordinate Sources

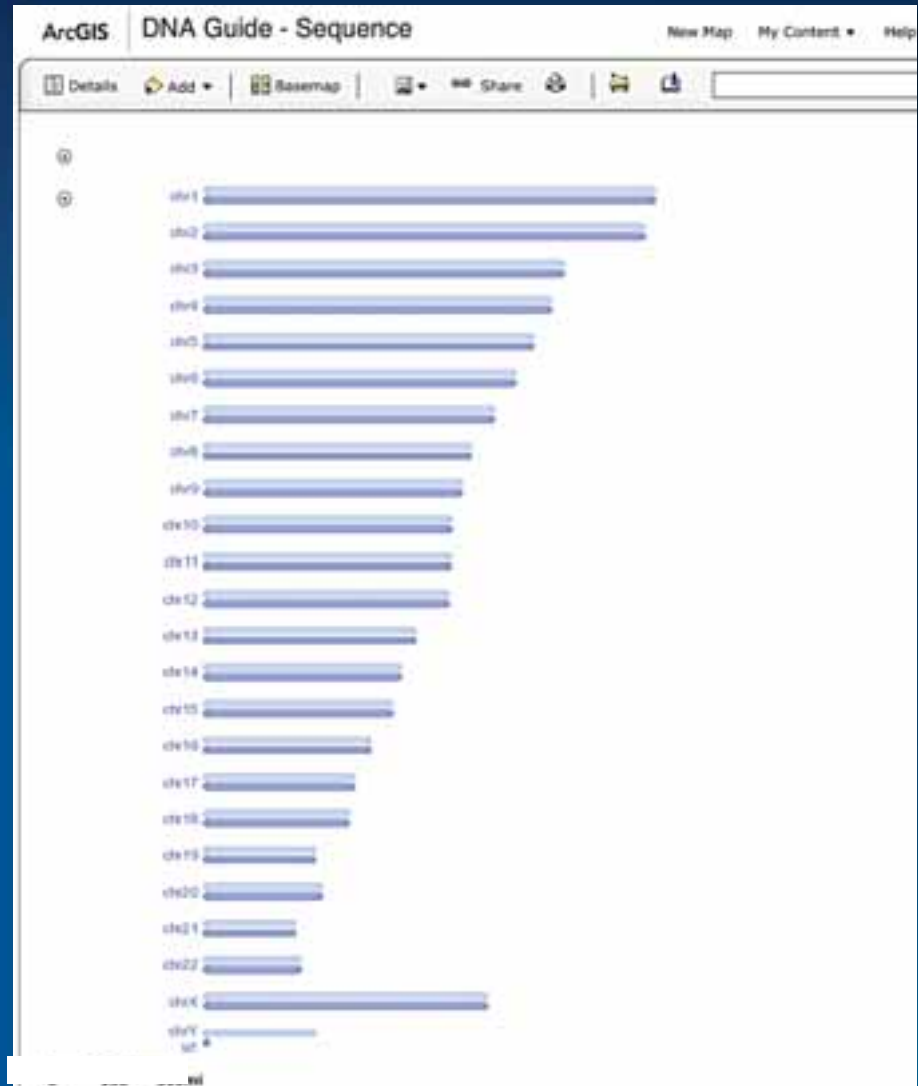
- Human Reference Genome Builds  
<http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human>
- Human Genes  
**HUGO** [www.genenames.org](http://www.genenames.org)
- SNPs **DBSNP**  
[www.ncbi.nlm.nih.gov/SNP](http://www.ncbi.nlm.nih.gov/SNP)



# Bringing Genomic Data into GIS

## Calculating x Coordinate

- Longest Chromosome , Chr 1, has 250,000,000 base pairs
- Using WebMercator, 1 base pair = .01 map units
- X Coordinate = genetic marker position divided by 100
- Example coordinate/position (1234567 = 12345.67)



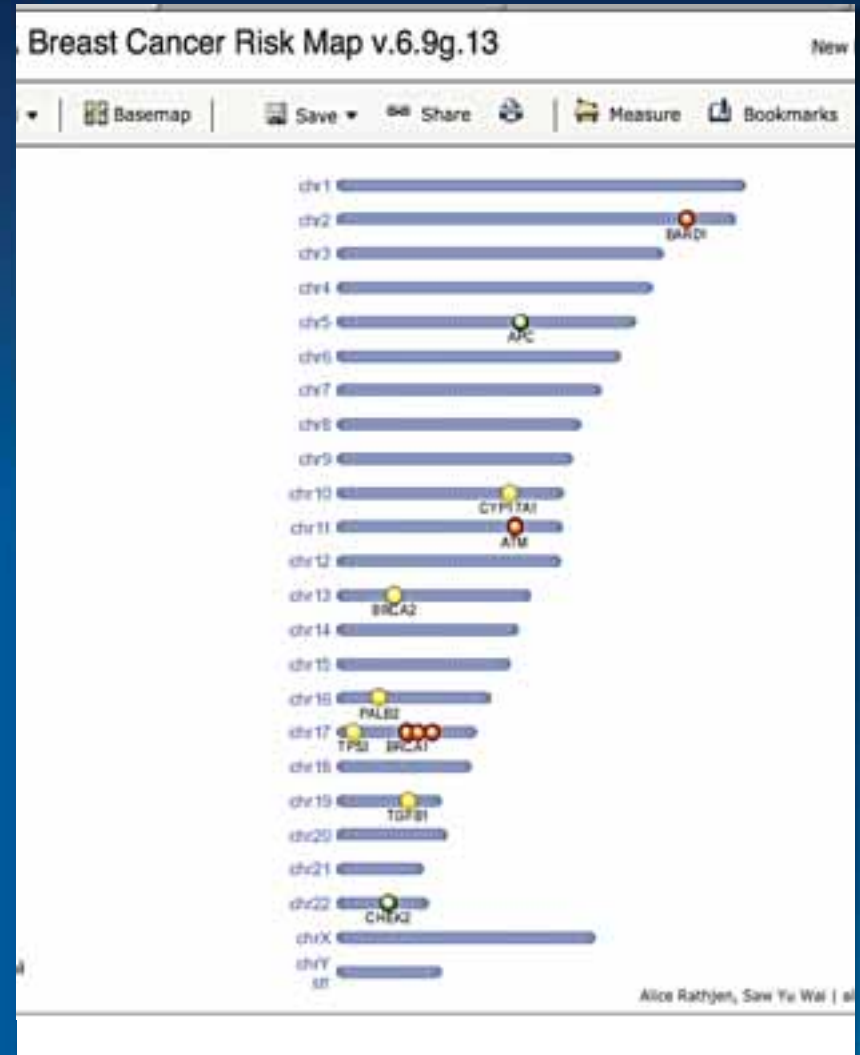
# Bringing Genomic Data into GIS

## Calculating $y$ Coordinate\*

### 23 Chromosome Pairs

- Chromosome number times -200.000.00 =  $y$  coordinate (i.e. chr 22 = -4400000.00)
- Chromosome X coordinate = -4600000.00
- Chromosome Y coordinate = -4800000.00

\*midline, for diploid browser add .01 and subtract .01 for each chromosome pair



# Custom Genomic Symbols for GIS

Genomes consist of large .txt files full of A, T, C and G's

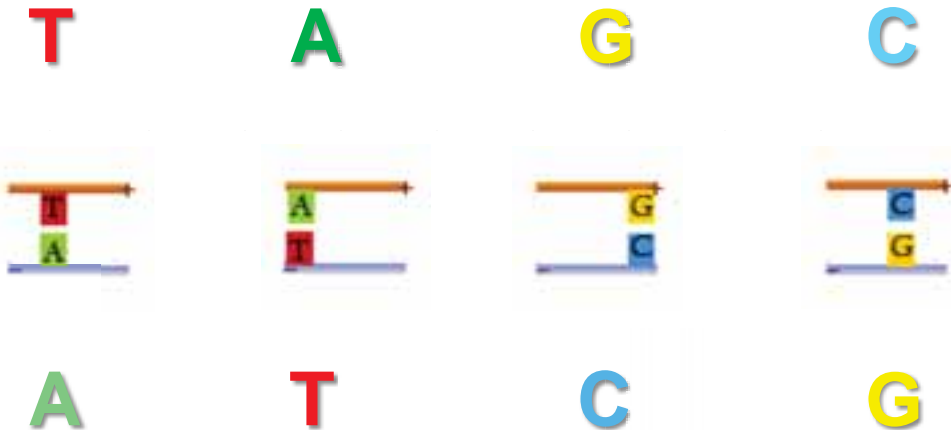
DNA Bases: A, T, C, G's.



# Bringing Genomic Data into GIS

A's and T's bond together, C's and G's bond together.

Double Helix (A's & T's, C's & G's)

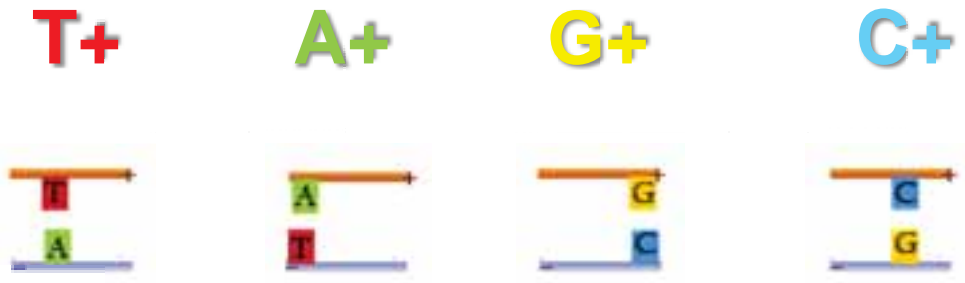




# Bringing Genomic Data into GIS

The data in .txt files is for one side of the DNA Strand.

Top Strand is the **Plus (+) Strand**



**A-**      **T-**      **C-**      **G-**

Bottom Strand is the **Minus(-) Strand**

# DNA Mapping

## Sample Dataset

- RSID number is the research ID for the marker.
- The Genotype “AG” means the person has an “A” at this location on one of their Chromosome 1s and a G at this location on their second Chromosome 1

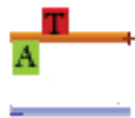
# Human Reference Build 36 , + orientation			
#rsid	chromosome	position	genotype
rs3094315	1	742429	AG
rs3131972	1	742584	AG
rs11240777	1	788822	AG
rs6681049	1	789870	CC
rs4970383	1	828418	CC
rs4475691	1	836671	CC
rs7537756	1	844113	AA
rs13302982	1	851671	GG
rs1110052	1	863421	GT
rs2272756	1	871896	AG
rs3748597	1	878522	CC
rs13303106	1	881808	AG
rs28415373	1	883844	CC
rs13303010	1	884436	AA

# Custom Symbols for Spatial Analysis

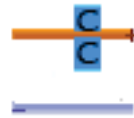
Points in genome where humans vary are called SNPs

Patient data often shown in pairs of letters

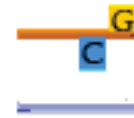
**AT+**



**CC+**



**CG+**



(One letter for each chromosome pair)

# Benefits of GIS

## Enabling Information Flow

- Topological Analysis across layers
- Medical Utility (4 stars rating - Pathologists, AMA)
- Viewing Risk (E=Everyone, PG = Physician Guidance, R = Restricted –via genetic counselors)

## Patient Engagement

- Real Time Consent (fine grain control for sharing info)



# Thank You ESRI

Alice Rathjen

CEO, Founder, DNA Guide  
alice@dnaguide.com

Key Contributors:

Saw Yu Wai

Mike Hargreaves

Richard Couch

