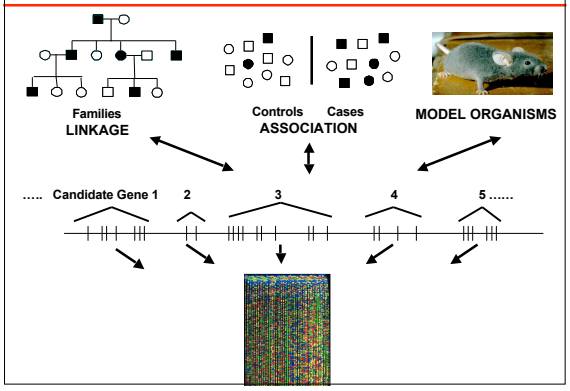
Medical Resequencing

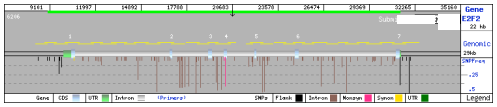
Debbie Nickerson

Department of Genome Sciences
University of Washington

Genetic Studies



Overview of a Candidate Gene

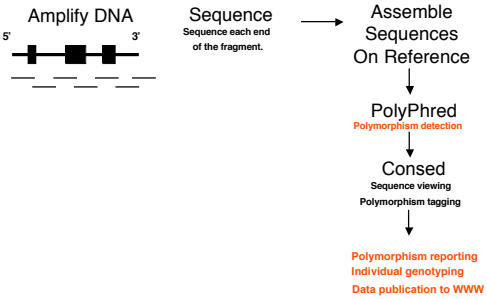


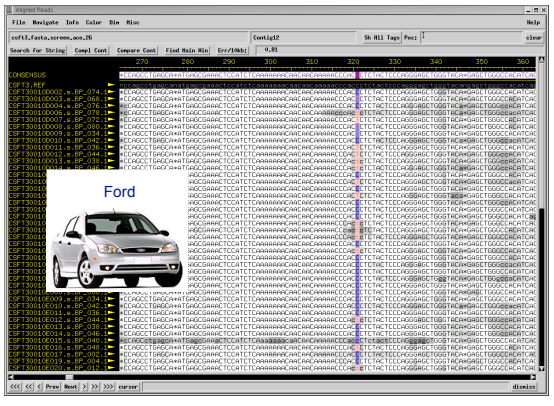
Average Gene Size - 26.5 kb ~ Compare 2 haploid - 1 in 1,200 bp

~130 SNPs (200 bp) - 15,000,000 SNPs

~ 44 SNPs ≥ 0.05 MAF (600 bp) - 6,000,000 SNPs

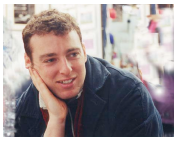
Sequencing production and data analysis pipeline

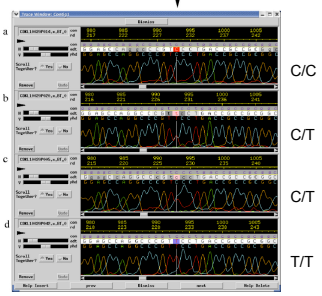




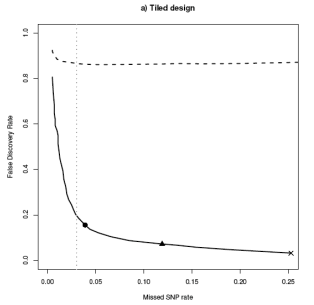
Aston-Martin of SNP Detection - PolyPhred 5.0

* Matthew Stephens
Peggy Dyer-Robertson
Jim Sloan

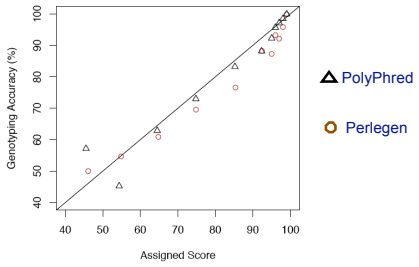




Comparison PolyPhred v4.29 versus v5.0

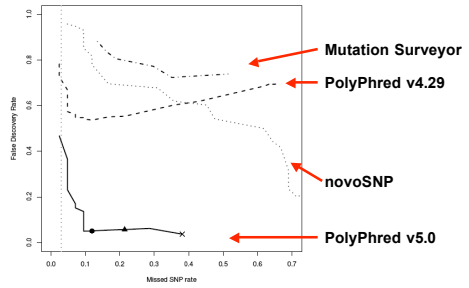


PolyPhred 5 Scores - Provide Quantitative Assessment of SNP Genotype



Double-Coverage - Automation = 93% of all SNPs, 100% of high-frequency SNPs, with no false positive SNPs identified, and 99.9% genotyping accuracy.

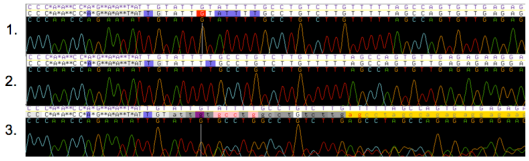
Comparison PolyPhred v5.0 to others



PolyPhred Update - Indels

Short Indels < 300 bp

95% less than 15 bp



Bhangale et al (2005) Hum Mol Genet. 14: 59-69

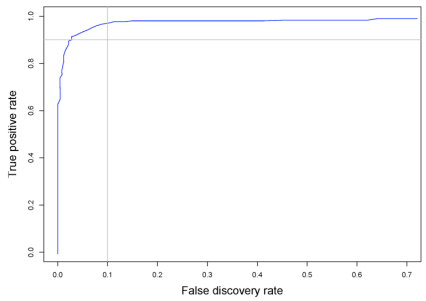
Importance of short indels

- Indels are common and in LD with substitutions and can be used to improve the marker densities
- Indels are overrepresented as disease-causing mutations
  - ~24% of mutations in the HGMD are indels

type	no. of entries	percent
Micro lesions		
missense/nonsense	21013	57.21%
splicing	4011	9.51%
regulatory	433	1.03%
small deletions	7042	16.69%
small insertions	2780	6.59%
small indels	398	0.94%
Gross lesions		
repeat variations	86	0.20%
gross insertions/duplications	383	0.91%
complex rearr./inversion	491	1.16%
gross deletions	2356	5.58%
total	42178	100.00%

24.22%

Indel-Detection Accuracy

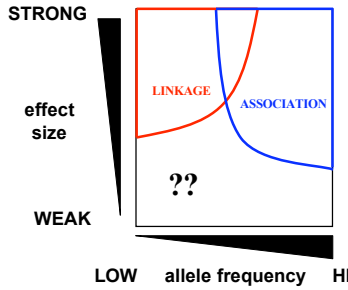


For Every 9 True Positives - 1 False- Positives

Medical Resequencing

- Discovery of rare functional variants -
  - Sequencing at the tails of the distribution
- Testing the Common Disease Common Variant (CDCV) hypothesis
  - Candidate genes very feasible
- Whole Genome Sequencing

Genetic Strategy Determined by Effect Size & Allele Frequency



Ardlie, Kruglyak & Seielstad (2002) Nat. Genet. Rev. 3: 299-309
Zondervan & Cardon (2004) Nat. Genet. Rev. 5: 89-100

ABCA1 and HDL-C

	Sequence variants unique to one group			
	Low HDL-C		High HDL-C	
	NS	S	NS	S
			DHS	
ABCA1	14	6	2	5
APOA1	1	0	0	1
LCAT	0	1	1	0
			Canadians	
ABCA1	14	2	2	3
APOA1	0	1	0	0
LCAT	6	1	0	0

–Cohen et al, Science 305, 869-872, 2004

- **Observed excess of rare, nonsynonymous variants in low HDL-C samples at ABCA1**
- **Demonstrated functional relevance in cell culture**

Rare coding variants

- No single variant frequent enough for significant association
- Indications of function
  - Ratio of synonymous to nonsynonymous
  - Predicted function from evolutionary data
  - Wet bench tests

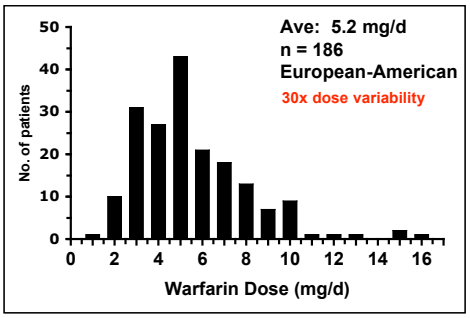
Medical Resequencing

- Testing the Common Disease Common Variant (CDCV) hypothesis
  - Candidate genes very feasible
- What about rare variants (CDRV)?
- Whole genome using tagSNPs feasible but sequencing could be in the future

Warfarin Background

- **Commonly prescribed oral anti-coagulant and acts as an inhibitor of the vitamin K cycle**
- **In 2003, 21.2 million prescriptions were written for warfarin (Coumadin™)**
- **Prescribed following MI, atrial fibrillation, stroke, venous thrombosis, prosthetic heart valve replacement, and following major surgery**
- **Difficult to determine effective dosage**
  - **Narrow therapeutic range**
  - **Large inter-individual variation**

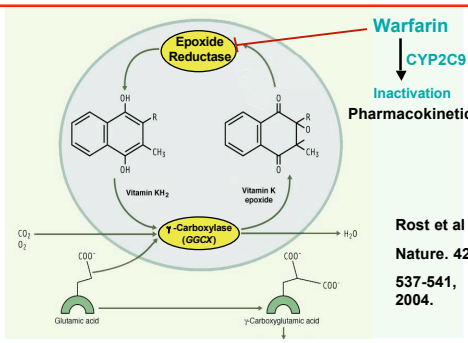
WARF+coumarin



✓Patient/Clinical/Environmental Factors

✓Pharmacokinetic/Pharmacodynamic - Genetic

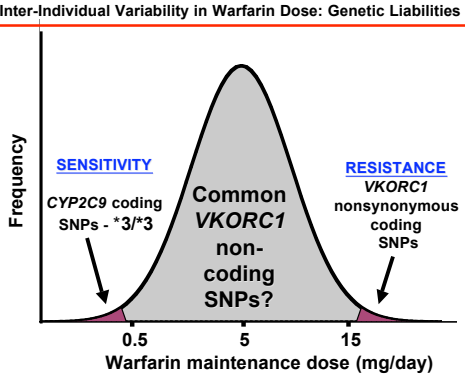
Warfarin inhibits the vitamin K cycle



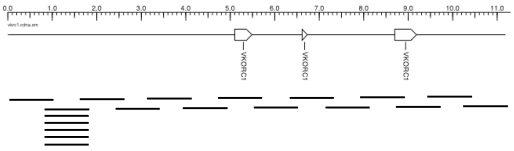
Rost et al Nature. 427: 537-541, 2004.

Vitamin K-dependent clotting factors (FII, FVII, FIX, FX, Protein C/S/Z)

Inter-Individual Variability in Warfarin Dose: Genetic Liabilities



SNP Discovery: Resequencing *VKORC1*



- PCR amplicons --> Resequencing of the complete genomic region
- 5 Kb upstream and each of the 3 exons and intronic segments; ~11 Kb
- Warfarin treated clinical patients (UWMC): 186 European
- Other populations: 96 European, 96 African-Am., 120 Asian

Rieder et al NEJM 352: 2285-2293, 2005

SNP Discovery: Resequencing Results

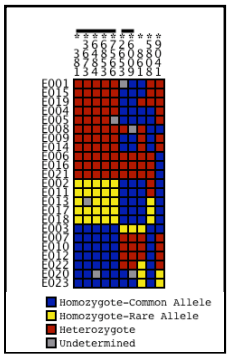
VKORC1 - PGA samples (European, n = 23)
Total: 13 SNPs identified
10 common/3 rare (<5% MAF)

VKORC1 - Clinical Samples (European patients n = 186)
Total: 28 SNPs identified
10 common/18 rare (<5% MAF)

15 - intronic/regulatory
7 - promoter SNPs
2 - 3' UTR SNPs
3 - synonymous SNPs
1 - nonsynonymous
- single heterozygous indiv. - highest warfarin dose = 15.5 mg/d

None of the previously identified *VKORC1* warfarin-resistance SNPs were present (Rost, et al.)

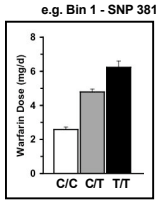
Do common SNPs associate with warfarin dose?



Five Bins to Test

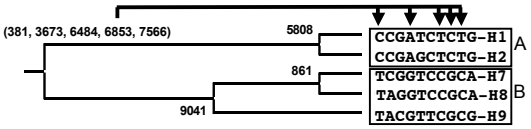
1. 381, 3673, 6484, 6853, 7566
2. 2653, 6009
3. 861
4. 5808
5. 9041

Bin 1 - p < 0.001
Bin 2 - p < 0.02
Bin 3 - p < 0.01
Bin 4 - p < 0.001
Bin 5 - p < 0.001



SNP x SNP interactions - haplotype analysis?

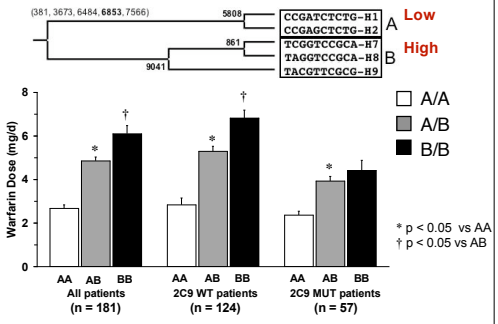
VKORC1 haplotypes cluster into divergent clades



Patients were assigned a clade diplotype:

e.g. Patient 1 - H1/H2 = A/A
Patient 2 - H1/H7 = A/B
Patient 3 - H7/H9 = B/B

VKORC1 clade diplotypes show a strong association with warfarin dose



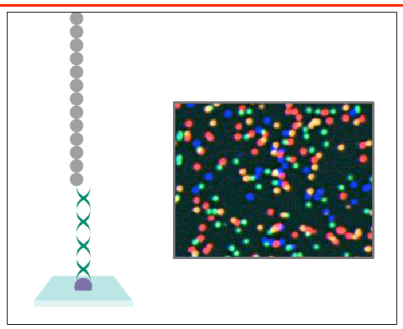
Medical Resequencing

- Discovery of rare functional variants -
  - Sequencing at the tails of the distribution
- Testing the Common Disease Common Variant (CDCV) hypothesis
  - Candidate genes very feasible
- Whole Genome Sequencing

SNP Genotyping -

Is it an intermediate stop on the way to whole-genome sequencing?

Long term sequencing - In situ approaches



Solexa - an example

Sequencing could be the ultimate genotyping tool

- More applications

- Further Technology Development