Findings from a genome-wide study of ovarian cancer susceptibility

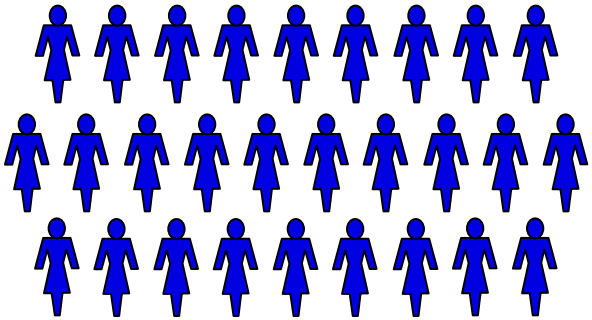
Paul Pharoah
Reader in Cancer Epidemiology
University of Cambridge

Background

- Ovarian cancer has substantial inherited genetic component
- Genes with rare, high-penetrance alleles identified by linkage and positional cloning in the mid-1990's
  - BRCA1, BRCA2, MMR
- These account for about 40% of the genetic component of risk
- A variety of genetic models may underlie the unexplained fraction

Association studies

Study design

	Patients	SNPs
1,819 cases 2,353 controls		Illumina 610K and 550k
4,833 cases 5,237 controls		Illumina 22k iSelect
2,962 cases 5,232 controls		Taqman 1 SNP

Sample sets by stage

Stage 1 data

- Cases
  - 1,819 invasive cases from SEARCH, FOCR, UKOPS, RMH
  - Illumina 610k array
- Controls 1
  - 1,437 individuals from 1958 birth cohort
  - Illumina 550k array
- Controls 2
  - 936 individuals from NSCCG
  - Illumina 550k array

Stage 1 exclusions

- Call rate <80%
- Assign intercontinental ancestry for European, African and Asian populations (using LAMP)
- Exclude 73 samples with < 90% European ancestry

Stage 1 genotyping QA

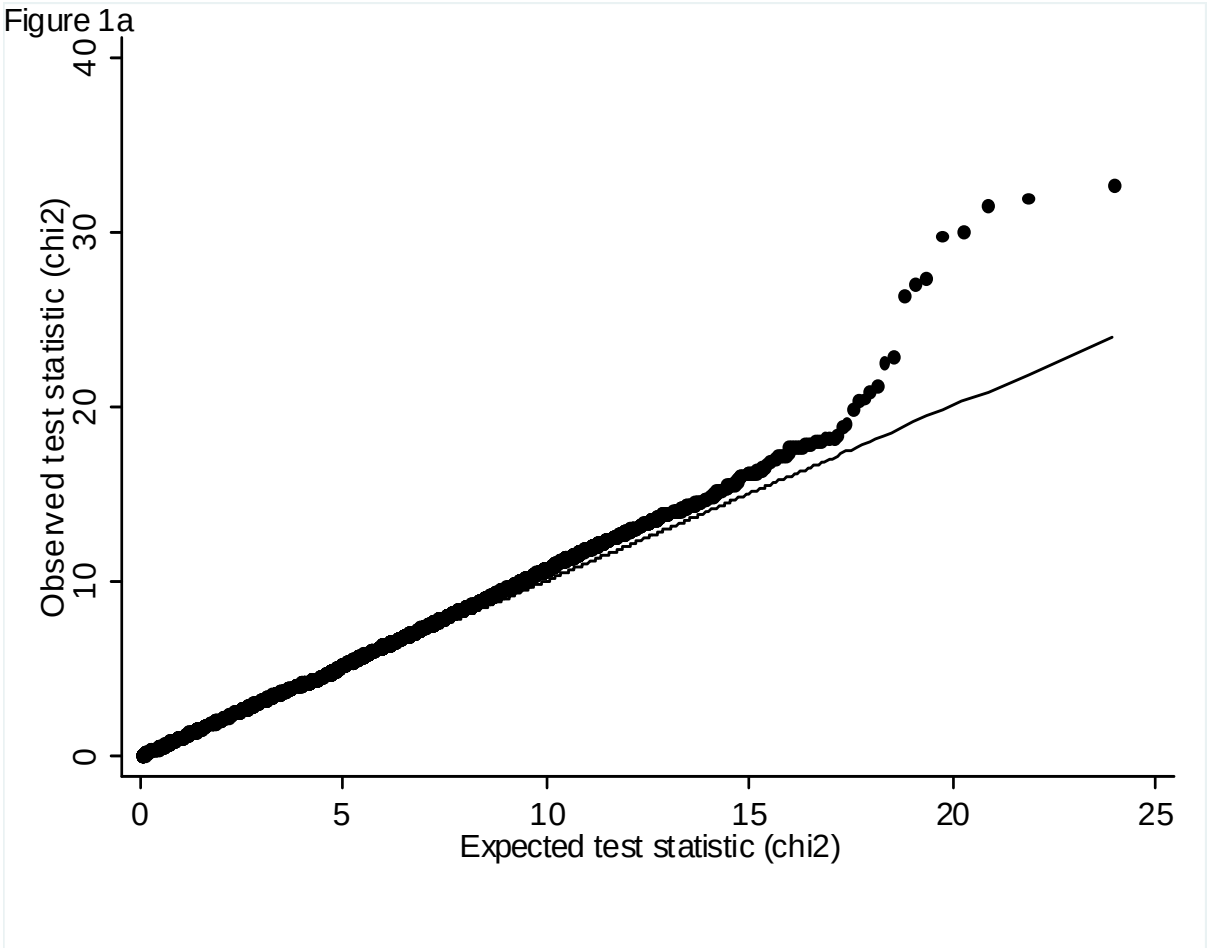
- Each study evaluated separately
- Exclude if
  - MAF < 1%
  - HWE $P<10^{-5}$
  - Call rate <99% and MAF 1-5%
  - Call rate <95% and MAF >5%
- Duplicate concordance 99.9%

Stage 1 data analysis

- Data for cases and controls available for 507k SNPs
- Used HapMap data to impute genotypes at further 2 million SNPs (MAF>5%)
- Comparison genotype frequencies in cases and controls using trend test (1df)

Stage 1 results

Figure 1a



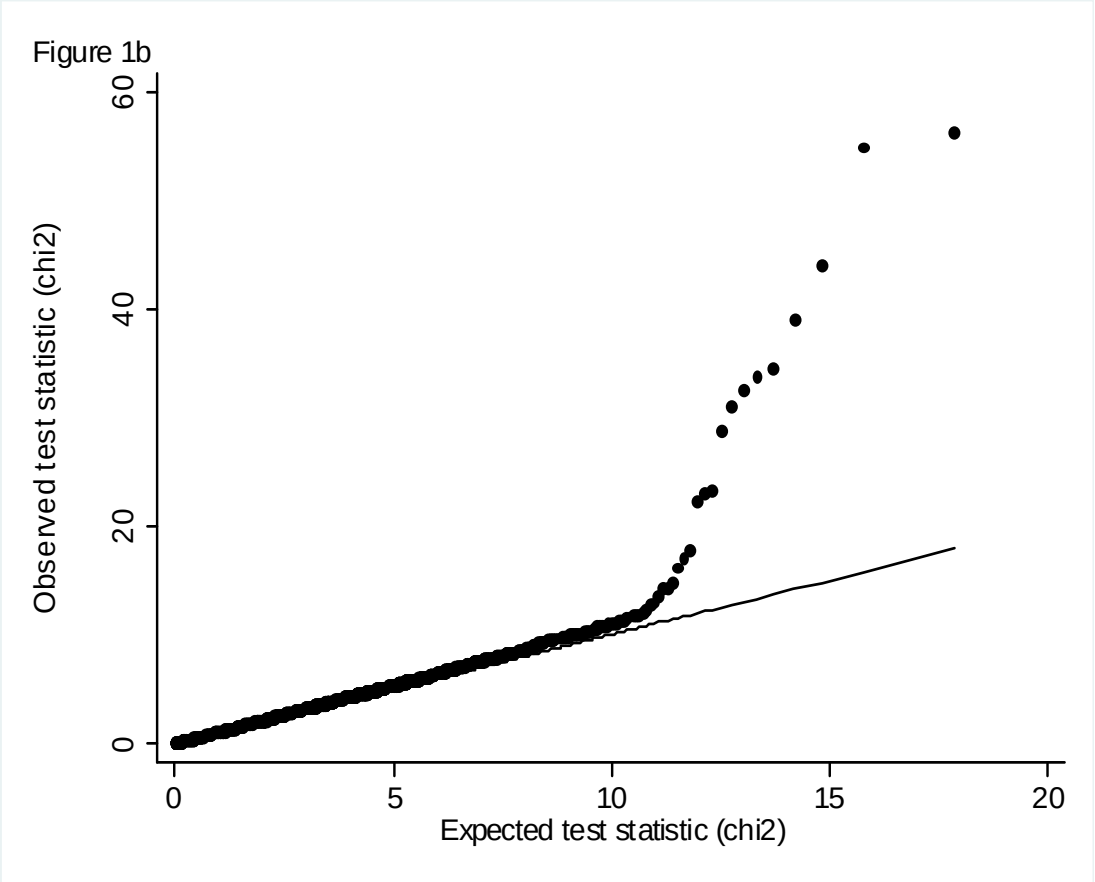
Stage 2

- Top 21,700 SNPs based on test statistic selected
- Results for imputed SNPs down weighted by 20%
- Only best SNP for sets of correlated SNPs (rsq>0.8) selected
- Additional 1,000 SNPS correlated with top 1,000 SNPs included
- 800 ancestry informative markers (AIMs)

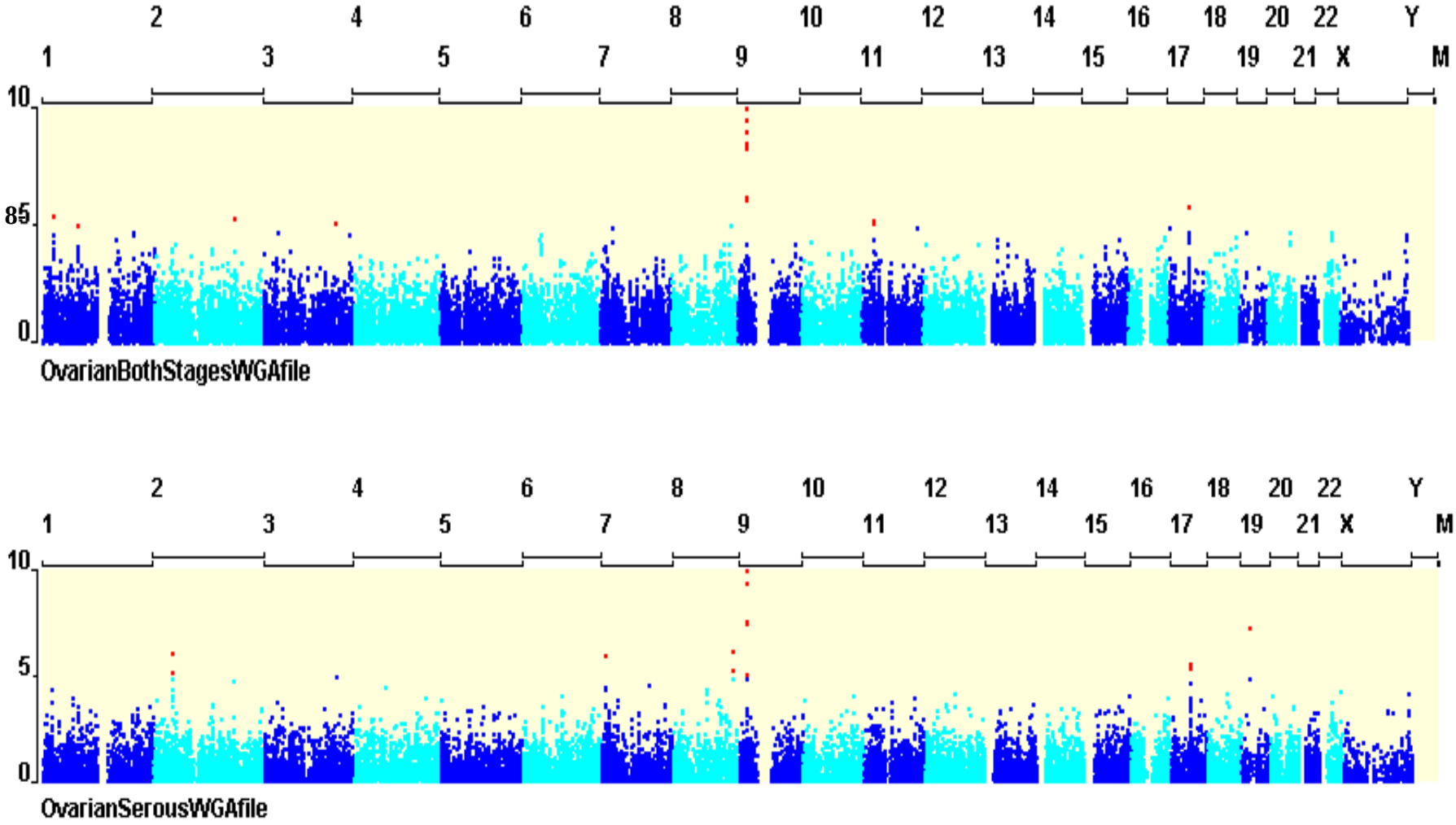
Stage 2 data analysis

- Comparison genotype frequencies in cases and controls using trend test (1df)
- Stratified by study
- Adjusted for population of origin determined using ancestry informative markers
  - calculate principal components (PCs) using AIMs
  - assign intercontinental ancestry score European, African American, East Asian, Mexican and Indian populations
  - first PC and the ancestry scores were included as covariates in subsequent association analyses

Stage 2 results



Stage 1 and 2 combined results



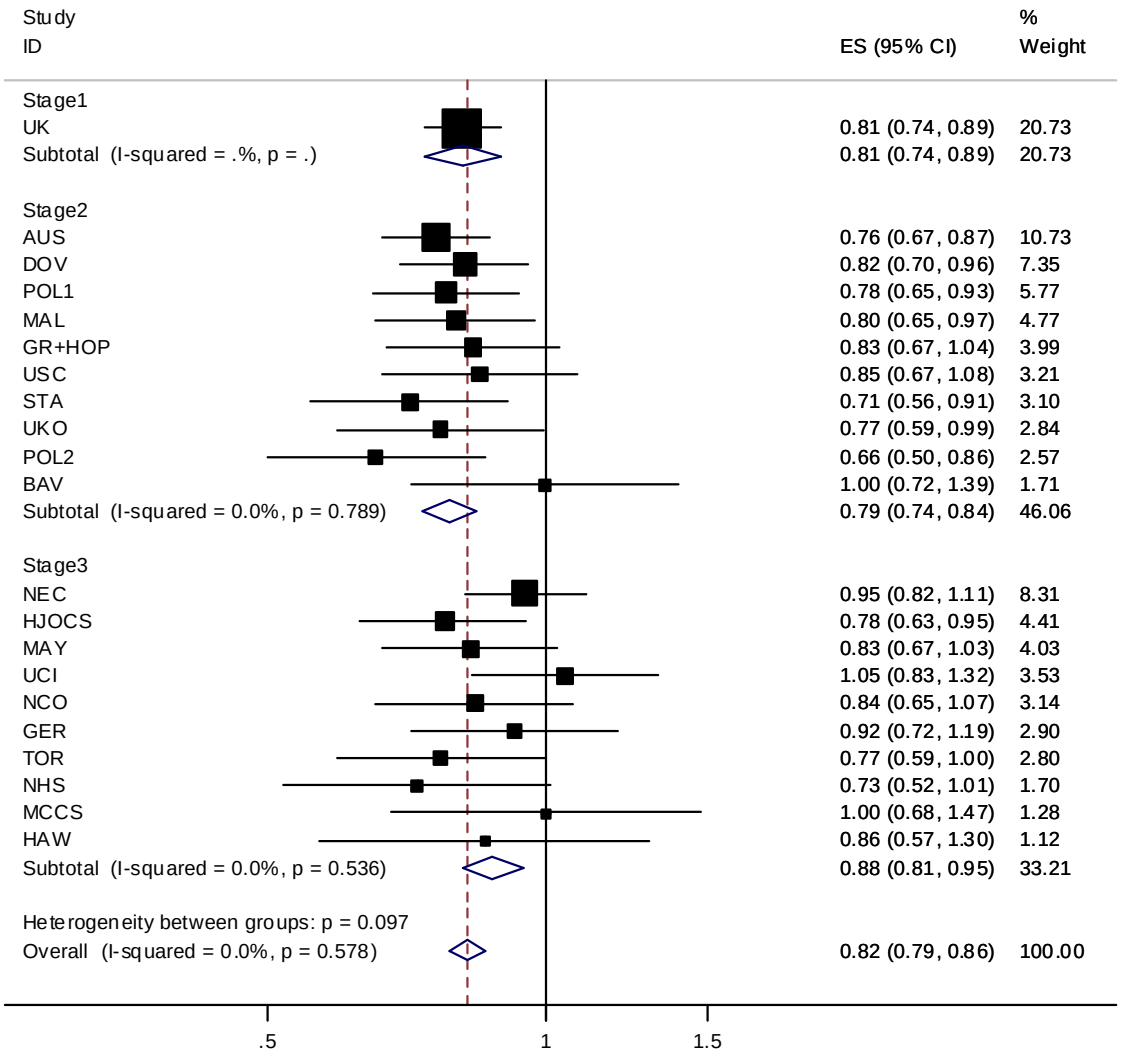
Top Hits

SNP	MAF	LD	OR (95% CI)	P-value
1	0.31	1.0	0.79 (0.75-0.84)	2.47x10-17
2	0.31	1.0	0.79 (0.75-0.84)	2.67x10-17
3	0.19	0.74	0.80 (0.75-0.85)	1.24x10-12
4	0.14	0.59	0.81 (0.76-0.87)	5.88x10-9
5	0.22	0.71	0.82 (0.78-0.87)	1.36x10-10
6	0.41	0.77	0.85 (0.81-0.90)	4.66x10-10
7	0.22	0.70	0.83 (0.78-0.88)	3.61x10-10
8	0.33	0.83	0.83 (0.79-0.88)	4.85x10-12
9	0.4	0.74	0.86 (0.82-0.90)	1.74x10-9
10	0.15	0.57	0.82 (0.77-0.88)	1.27x10-8
11	0.23	0.72	0.82 (0.77-0.87)	8.49x10-12
12	0.41	0.76	0.85 (0.81-0.89)	1.28x10-10

Stage 3

- 10 studies
- 2,962 cases and 5,232 controls
- Top hit genotyped by TAQMAN or as part of GWAS (Illumina)
- Subjects on non-European, self-reported ancestry excluded

Stage 1, 2, 3 combined – all cases

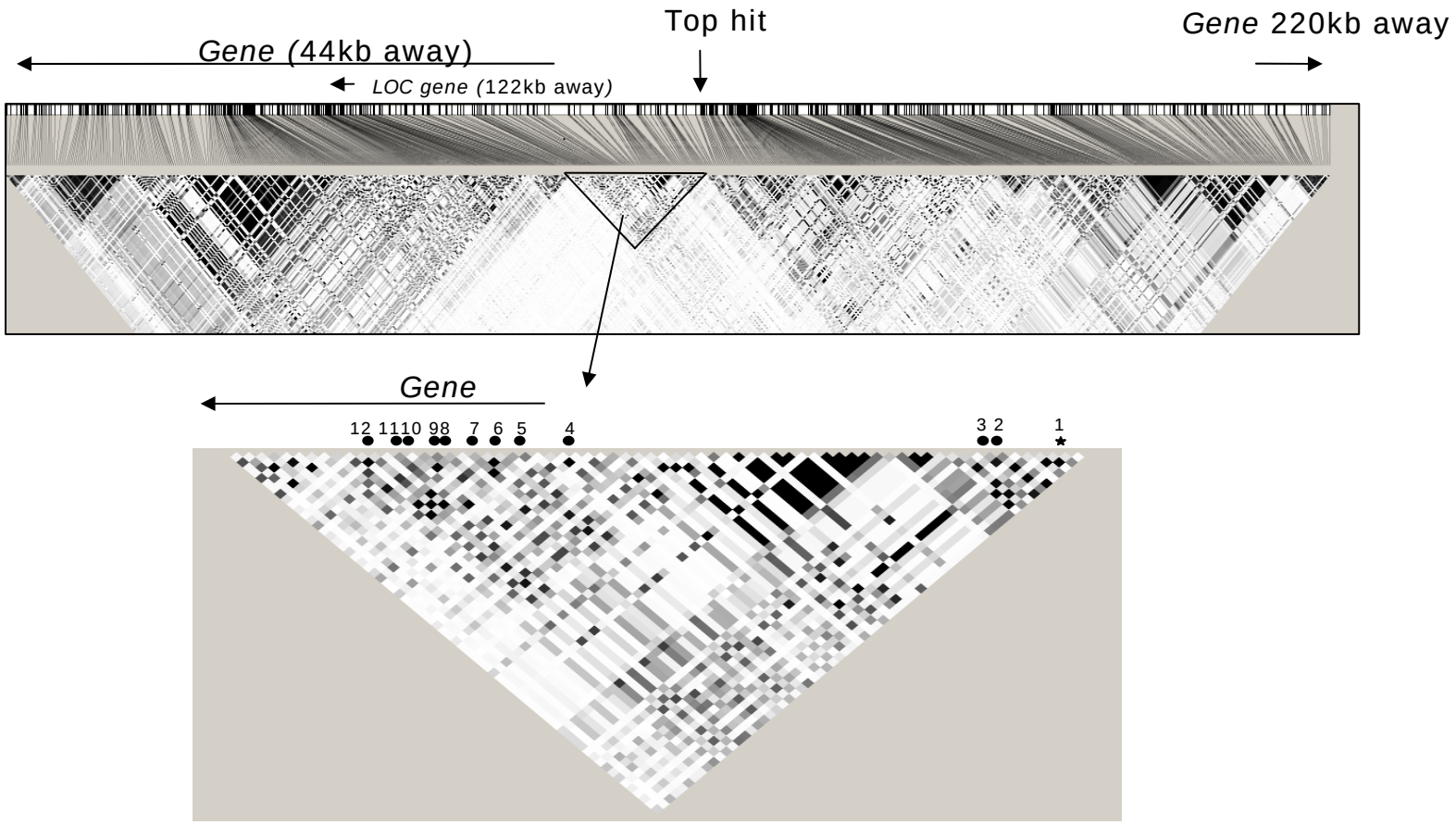


P-trend = 9.9×10^{-20} , 9614 cases/12823 controls

Risks by sub-type

Subtype	No. cases	Odds ratio	P-value
All	9,614	0.83 (0.79-0.86)	9.9×10^{-20}
Serous	5,274	0.78 (0.72-0.84)	5.4×10^{-22}
Endometioid	1,455	0.90 (0.82-0.98)	0.01
Clear cell	708	0.94 (0.84-1.05)	0.34
Mucinous	671	0.99 (0.88-1.12)	0.88

Associated region at 9p



Resequencing of the 9p region and further genotyping in ovarian cancer cases and controls will be needed to clarify the likely causal variant

Nearest gene

- Encodes highly conserved DNA-binding zinc-finger protein
- Extensive transcriptional variability
  - six promoters
  - ~ 90,000 mRNA isoforms, encoding > 2,000 different proteins
- Highly expressed in ovary and testis
  - may play a role in the differentiation of spermatoza and oocytes
- Little evidence of a role for role in cancer development

Next steps – find other loci

- Pooled data from SNP at top 10 loci (n=32) with data from two other GWAS
  - US
  - Decode
- Initial analysis suggests four new loci at genome wide significance
- Heterogenetiy by sub-type
- Further genotyping in OCAC planned

Next steps – fine mapping and function

- Initial analysis of data from 57 subjects in 1000 genome project show that there are 22 common variants correlated with top SNP that could be true causal variant
- Genotyping of all 22 in all OCAC studies planned
- Evaluation of expression in cell lines
- Evaluation of splice variation

Conclusions

- One confirmed susceptibility locus on 9p
  - accounts for~ 0.70% of the polygenic component of ovarian cancer risk
- Sub-type heterogeneity
- Pooled analysis of GWAS with large scale replication genotyping may identify further loci
- Targeted screening may be possible

Acknowledgements

All the study participants and Collaborators from

- AOC/ACS
- BAV
- DOV
- GER
- GR
- HAW
- HOP
- HJOCS
- MAL
- MAY
- NCO
- NEC
- NHS
- POL1
- POL2
- SEA
- STA
- TOR
- UCI
- UKO
- USC

Ovarian Cancer Research Fund

CANCER RESEARCH UK