
Biomarkers: GWAS, Sequencing and Integrative Biology

John D. Rioux, Ph.D.

Associate Professor of Medicine

Canada Research Chair -

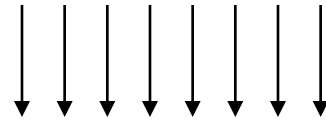
Genetics and Genomic Medicine of Inflammatory Diseases



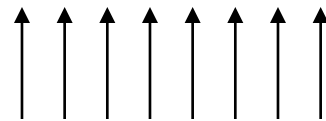
**INSTITUT DE
CARDIOLOGIE
DE MONTRÉAL**

Most disease result from combination of genes,
environment, behavior and bad luck

Genotype



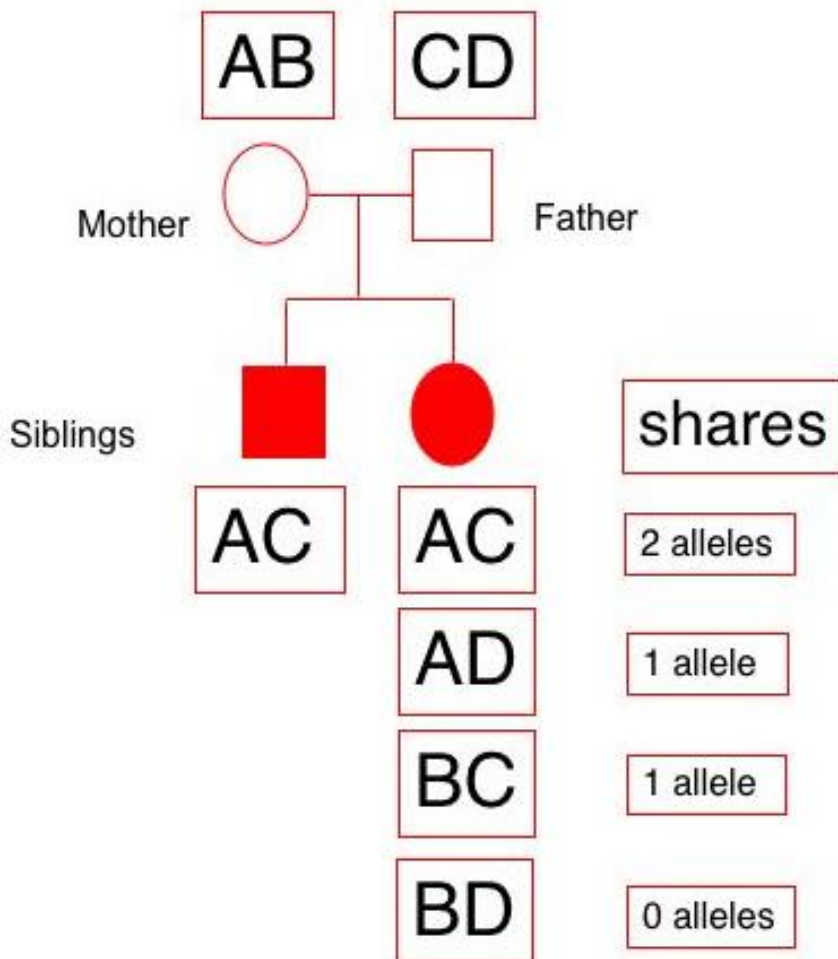
Behavior \Rightarrow **Disease** \Leftarrow Chance



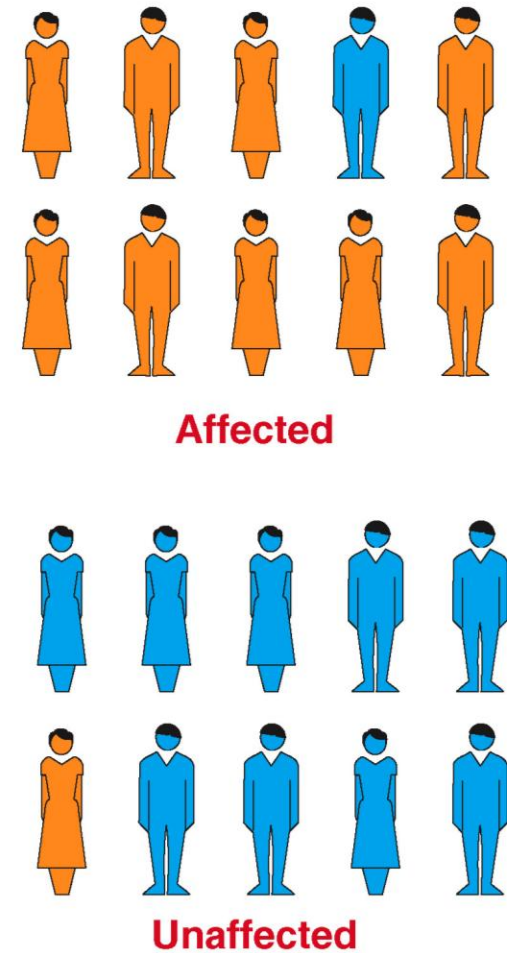
Environment

Two Main Approaches Used for Identification of Genetic Risk Factors

LINKAGE



ASSOCIATION



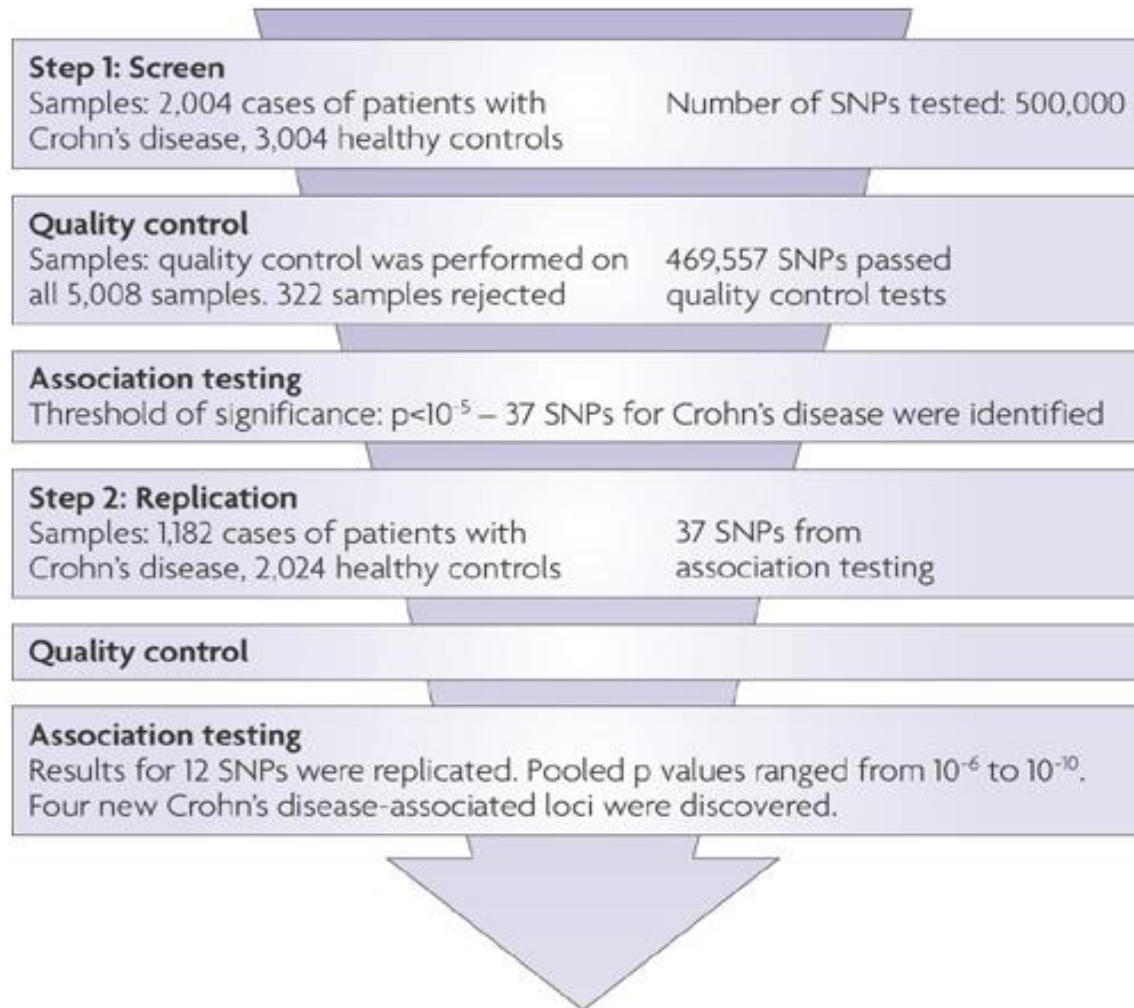
The candidate gene approach



The Genome-wide Association (GWA) approach

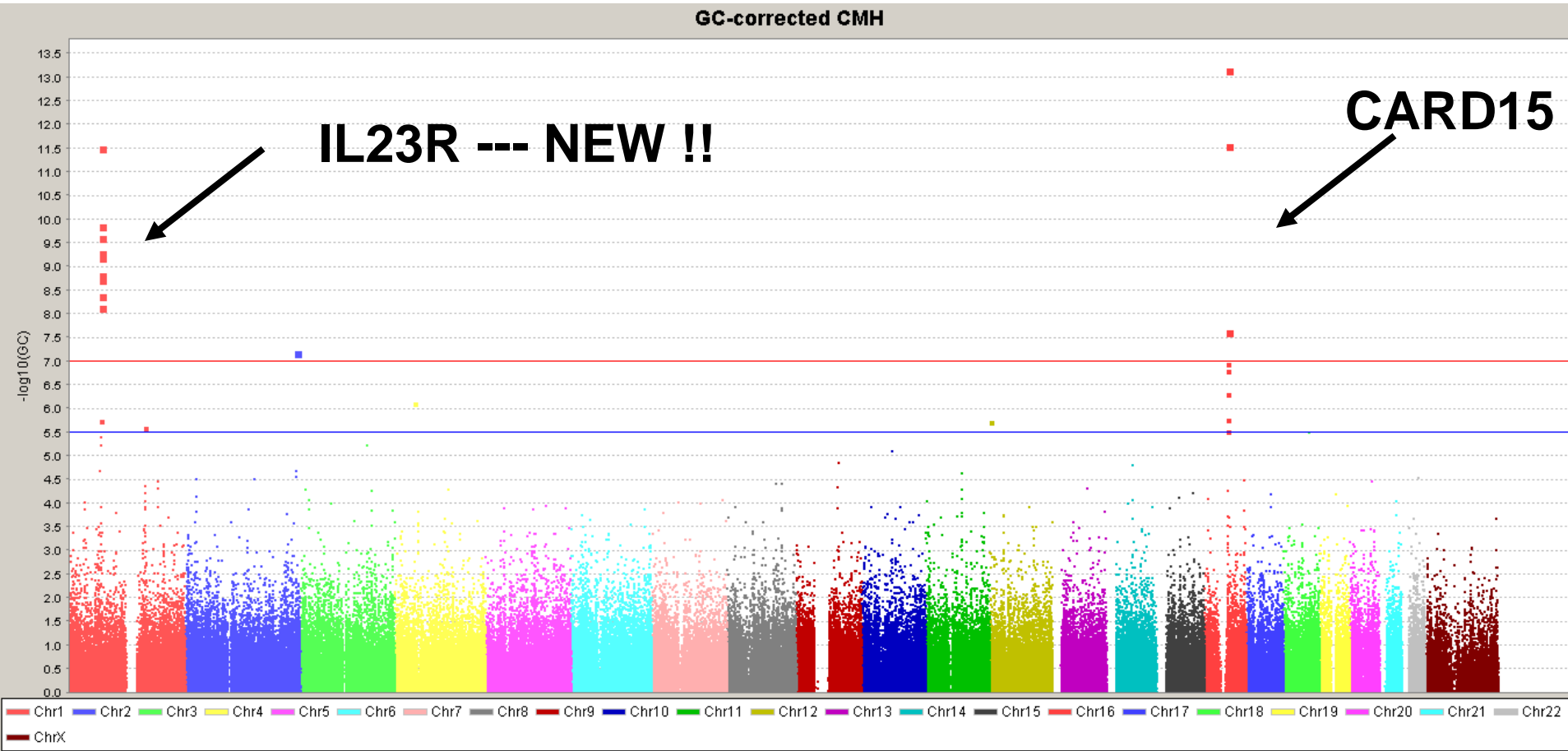


Basic Design of GWA Studies



Validated Genetic Risk Factors

Typical results from a GWA Study



1st Crohn's Disease GWA Identifies Protective and Predisposing Alleles in IL23R

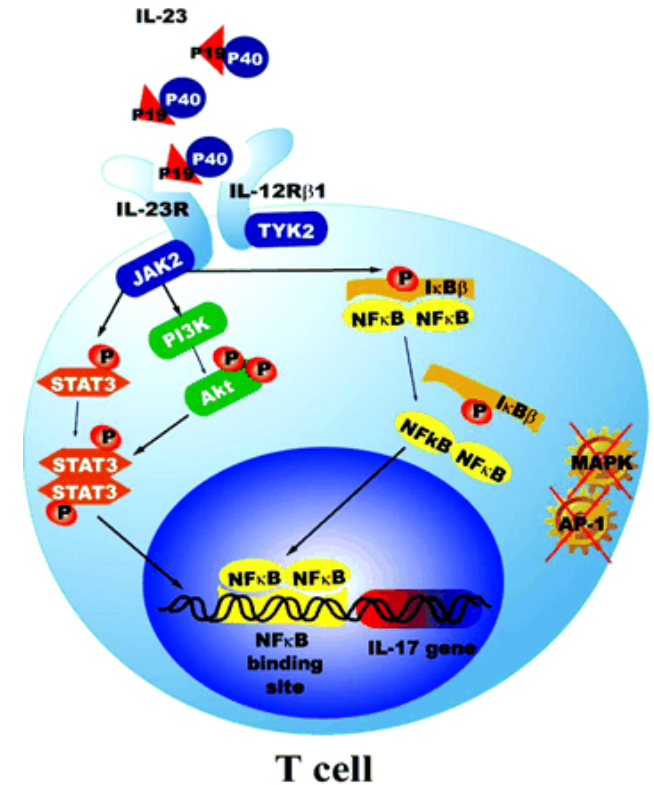
Arg381Gln:

Gln is protective

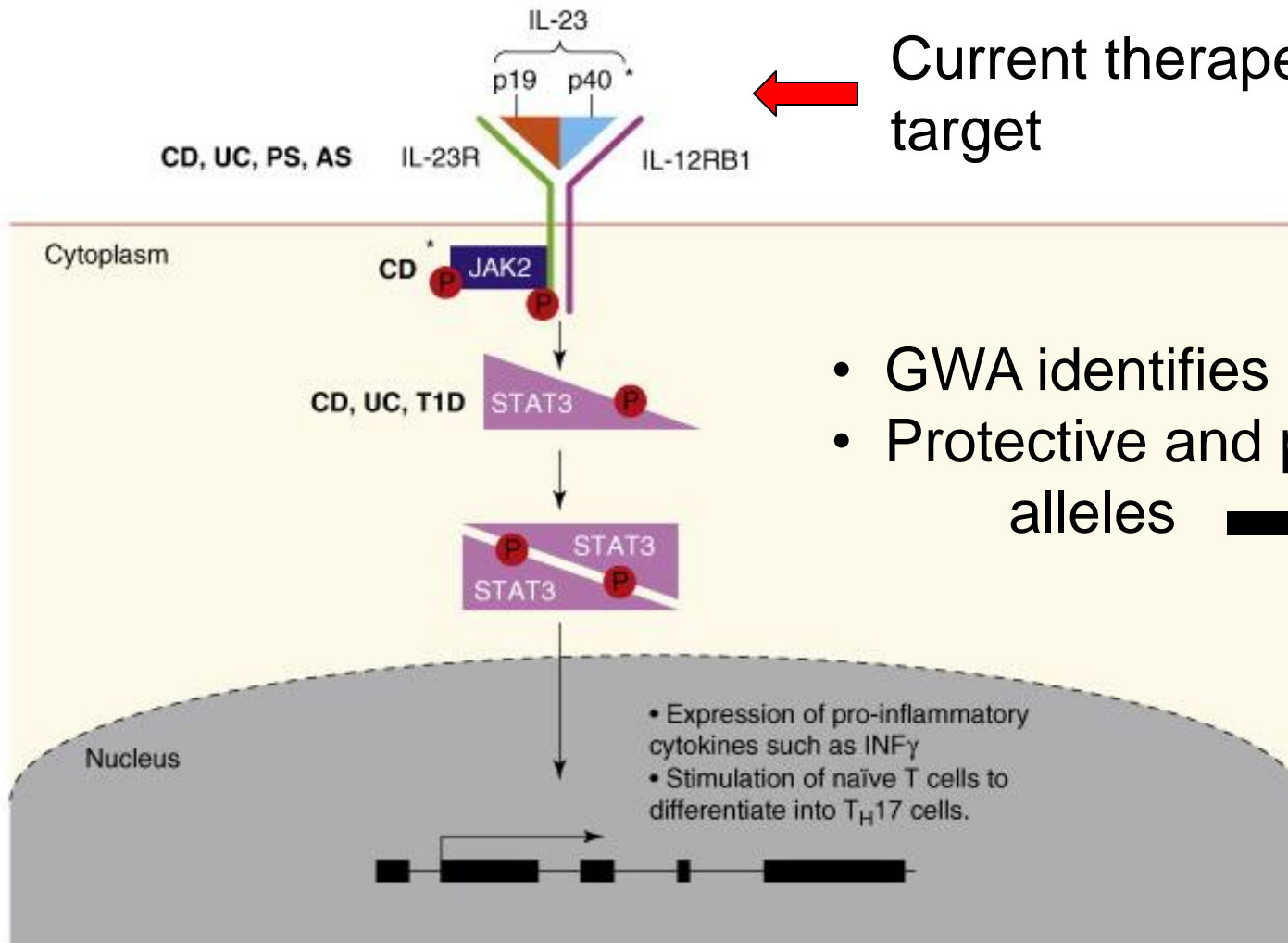
~2% in cases

~7% in controls

Non-coding/regulatory variants are found to increase risk to disease



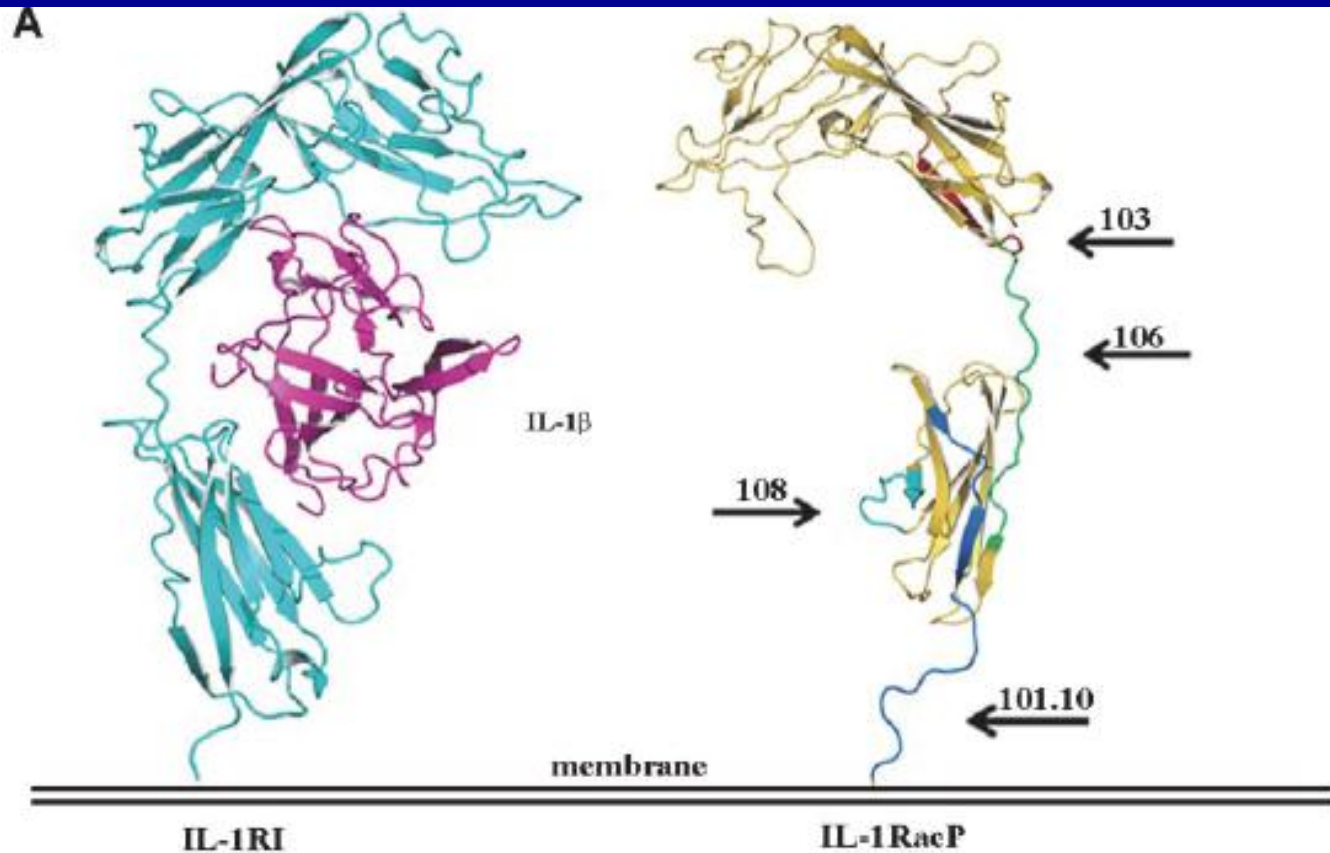
Multiple Genes in IL23R Pathway Involved Chronic Inflammatory Diseases



Current therapeutic target

- GWA identifies key pathways
- Protective and predisposing alleles → PGx

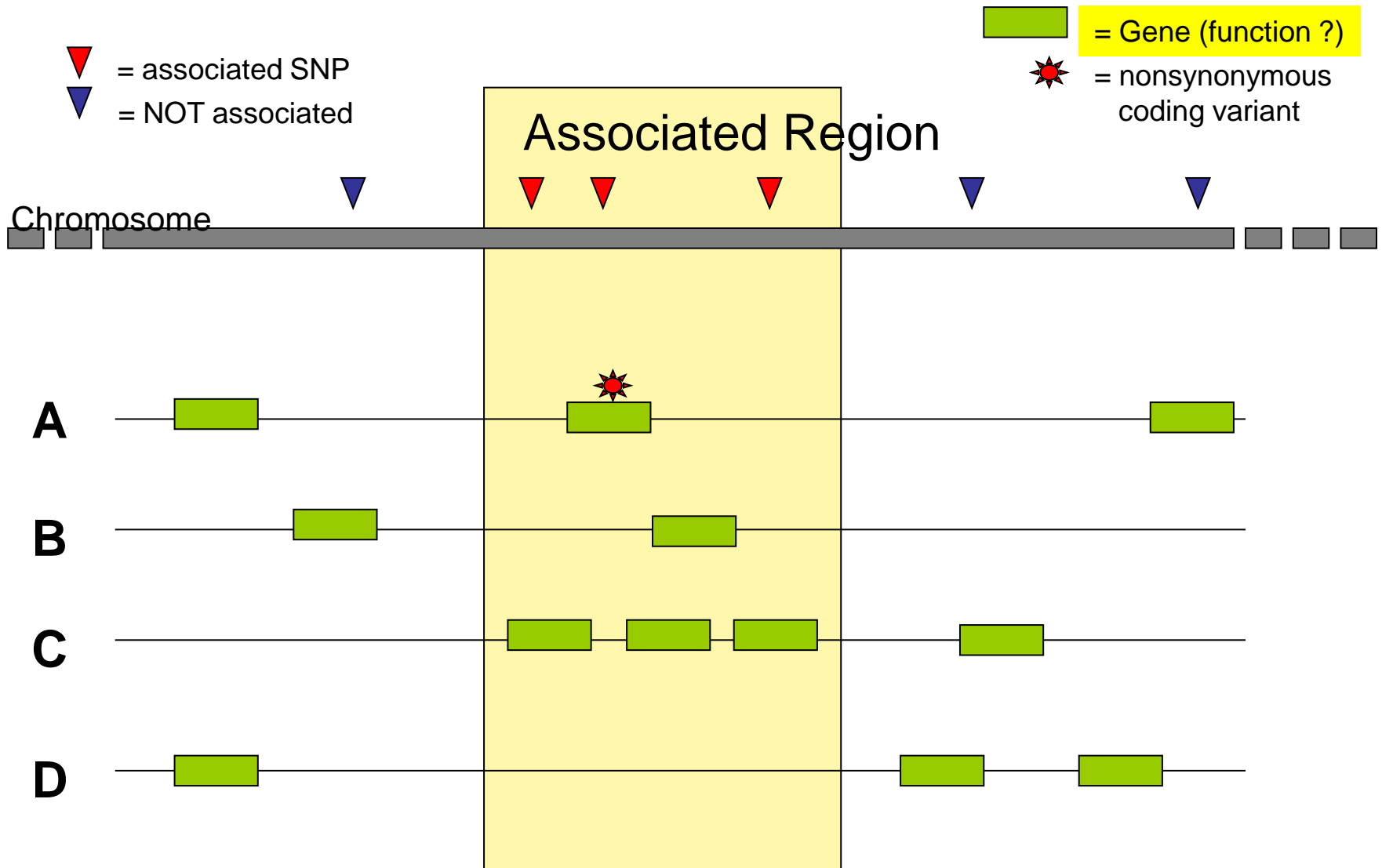
Pathways Identified by GWAS – Targets for Novel Therapies



B Sequence of the extracellular/transmembrane portion of IL-1RacP:
targeted loops/regions are color indicated

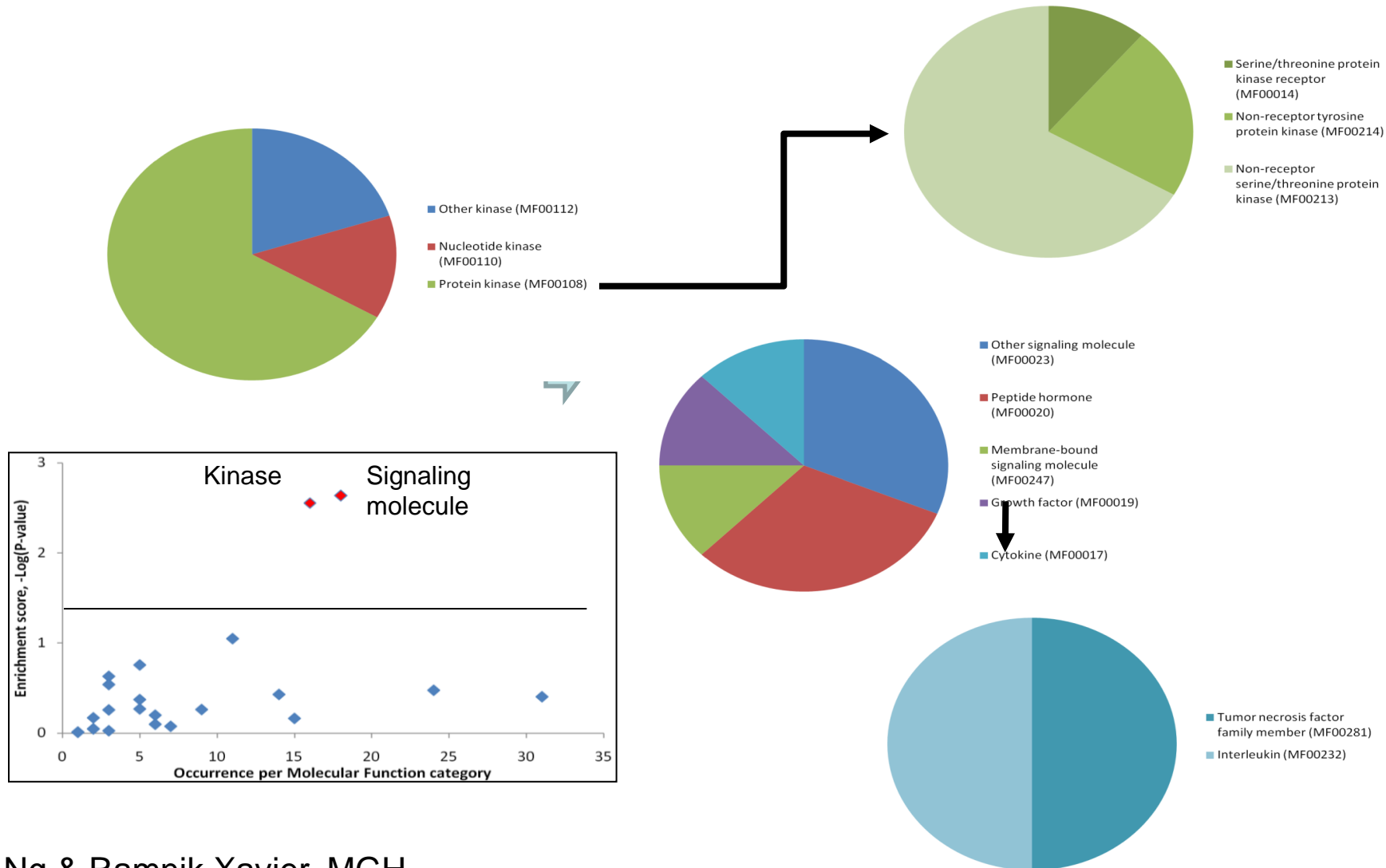
1	MTLLWCVVS	LYFYGILQSDA	SERCDDWGLD	TMRQIQVFED	EPARIKCPLF	EHFLKFNYST	60
61	AHSAGLTLIW	YWTRQDRDLE	EPINFRLPEN	RISKEKDVW	FRPTLLNDTG	NYTCMLRNTT	120
121	YCSKVAFPLE	VVQKDFNS	PMKLPVHKLY	IEYGIQRITC	PNVDGYFPSS	VKPTITWYMG	180
181	CYKIQNENNV	IPEGMNLNLSFL	IALISNNGNY	TCVVYPENG	RTFHLTRTLT	VKVVGSPKNA	240
241	VPPVVIHSEND	HVVYEKEPGE	ELLIPCTVYF	SFLMDSRNEV	WQVQVQVQVQ	WQVQVQVQVQ	300
301	QIHSRDEFE	ETGAILSTG	LEGTDLRDSH	LEHARDAMGE	ILMAMMSEGV	LDLDEHTEFE	360

Not All GWA Results Are Equal



Bioinformatic Approaches Can Group Risk Factors Into Functional Groups

Molecular Function Enrichment



Need to Identify CASUAL Genetic Variation

« *Next Generation* » Sequencing - Increased Capacity = Enabling Technology

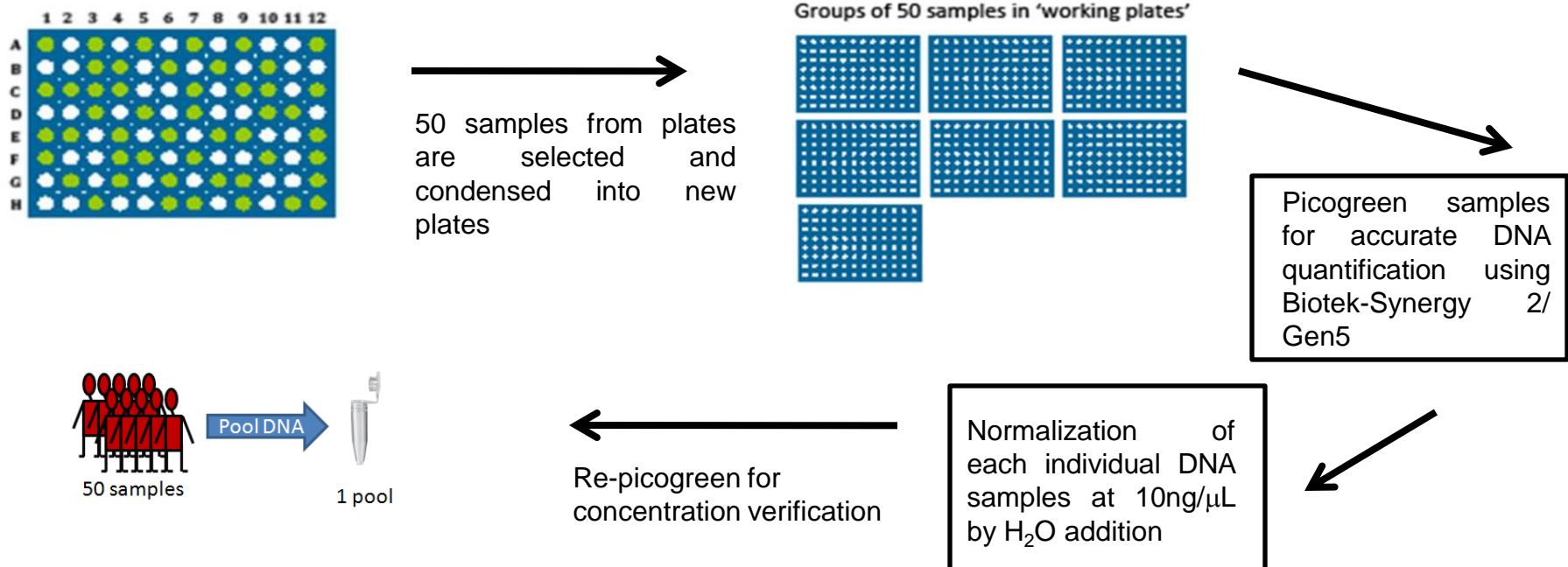
Example of technological progress ...

Year	Instrument	Read Length	# reads per run	Total # bases per run
2005	1st	~100 bp	100,000	20M
2007	2nd	~200 bp	400,000	100M
2010	3rd	~500 bp	1M	500M

Sequencing Strategy for Sequencing Risk Factors Identified by GWA Studies

Samples: 500 Cases + 500 controls

Enrichment: Targeted amplification of genomic regions (PCR)



Common and Rare Variants Identified In Targeted Sequencing of GWA Loci

708 high quality variants discovered in ~50 genes
***** Must Identify which are CAUSAL**

alleles	gene	type	rs#	base_chg	annot	f_nr_case	f_nr_con	chisqstat	assoc	lrt_stat	Genotype_FA	Genotype_FU
GC	Corf109	s	NA	CGG21CGC	Arg21Arg	0.0032	0.0082		1.52	1.581		
CG	Corf109	s	rs2271018	CCC73CCG	Pro73Pro	0.2377	0.217		0.85	0.8513		
CT	Corf109	s	rs2271017	AGC93AGT	Ser93Ser	0.0314	0.0483		2.6	2.619	0.04741	0.04928
TA	Corf109	ns	rs41269923	ATC100AAC	Ile100Asn	0.0117	0.0046		2.35	0	0.01433	0.005747
GT	Corf109	s	NA	GGG160GGT	Gly160Gly	0.0014	0		0.99	1.382		
CT	Corf109	s	rs34596202	ATC170ATT	Ile170Ile	0.0046	0.002		0.68	0.7048		
AT	Corf109	ns	rs41313912	TAT333TTT	Tyr333Phe	0.0188	0.003		8.08	8.989	0.01724	0.004335
CT	Corf109	ns	NA	CGT376TGT	Arg376Cys	0.0028	0		2	2.773	0.002865	0
GC	Corf109	ns	NA	AGG397ACG	Arg397Thr	0.0015	0.0014		0	0.0008132		
AC	Corf109	ns	rs45547233	AGA501AGC	Arg501Ser	0.073	0.0757		0.03	0.03808		
AC	Corf109	ns	NA	CAC506CCC	His506Pro	0	0.0015		0.95	1.165		
GC	Corf109	s	NA	GGG520GGC	Gly520Gly	0.1828	0.1649		0.78	0.7858		
GC	Corf109	s	NA	CGG521CGC	Arg521Arg	0.4702	0.4584		0.19	0.1973		
CT	Corf109	ns	rs296520	CGC538TGC	Arg538Cys	0.3809	0.3426		2.22	2.226		
CT	Corf109	ns	NA	CGC554TGC	Arg554Cys	0.004	0.0013		0.92	0.9624		
CA	Corf109	s	rs296521	CGC602CGA	Arg602Arg	0.555	0.534		0.62	0.6207		
CT	Corf109	s	NA	CTG606TTG	Leu606Leu	0.0017	0		1.11	1.427		
CA	Corf109	ns	NA	GAC607GAA	Asp607Glu	0.0018	0.0056		1.41	1.486		
GA	Corf109	ns	NA	CGG641CAG	Arg641Gln	0.0014	0		1	1.397		

unpublished

Whole Exome/Genome Sequencing Tool for Genetic Risk Factor Identification

Targeted capture and massively parallel sequencing of 12 human exomes

Sarah B. Ng¹, Emily H. Turner¹, Peggy D. Robertson¹, Steven D. Flygare¹, Abigail W. Bigham², Choli Lee¹,
Tristan Shaffer¹, Michelle Wong¹, Arindam Bhattacharjee⁴, Evan E. Eichler^{1,3}, Michael Bamshad²,
Deborah A. Nickerson¹ & Jay Shendure¹

Nature Vol 461 | 10 September 2009

		FSS24895	FSS10208	FSS10066	FSS24895 FSS10208 FSS10066 FSS22194	Any 3 of 4 FSS24895 FSS10208 FSS10066 FSS22194
Number of genes in which each affected has at least one...	Non-synonymous cSNP, splice site variant or coding indel (NS/SS/I)	4,510	3,284	2,765	2,479	3,768
	NS/SS/I not in dbSNP	513	128	71	53	119
	NS/SS/I not in eight HapMap exomes	799	168	53	21	160
	NS/SS/I neither in dbSNP nor eight HapMap exomes	360	38	8	1 (MYH3)	22
	...And predicted to be damaging	160	10	2	1 (MYH3)	3

Figure 2 | Direct identification of the causal gene for a monogenic disorder by exome sequencing. Boxes list the number of genes with one or more non-

High Performance Platform for the Discovery, development and validation of biomarkers



Cohorts



Typing known genetic variation



Deep re-sequencing
> causal variants

BIOMARKERS

MHI & UdeM High Performance Platform for Discovery, Development and Validation of Biomarkers



Cohorts



Typing known genetic variation

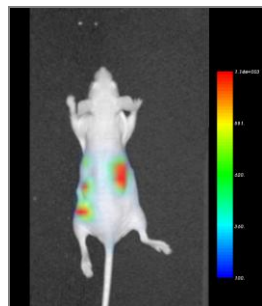


Deep re-sequencing
> causal variants



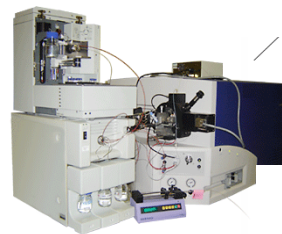
R&D

GLP



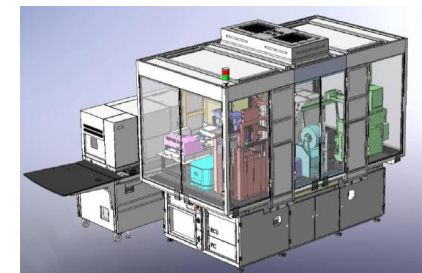
Molecular Imaging

4.7 tesla MRI (mouse, rat, rabbit)
PET-CT (rabbit); PET-CT (human)



Metabolomics

GC/MS/MS
LCMS-MS/MSQTOF

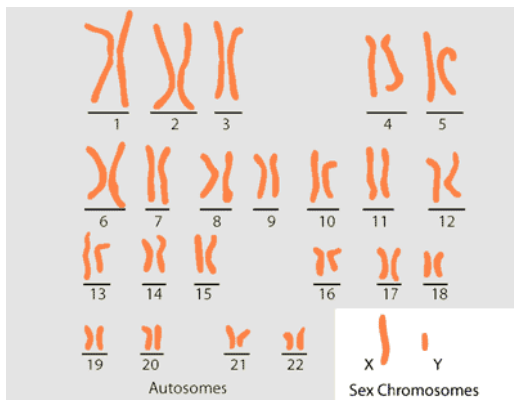
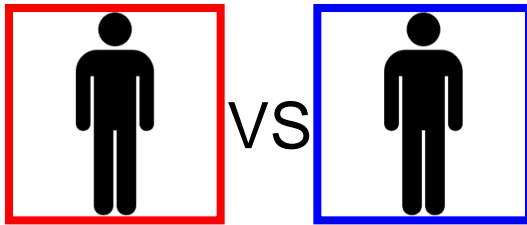


Integrative Biology

HCS (Opera)
Circulating BioM

GWA, Sequencing & Integrative Biology

Discovery of Risk Factors and Biomarkers



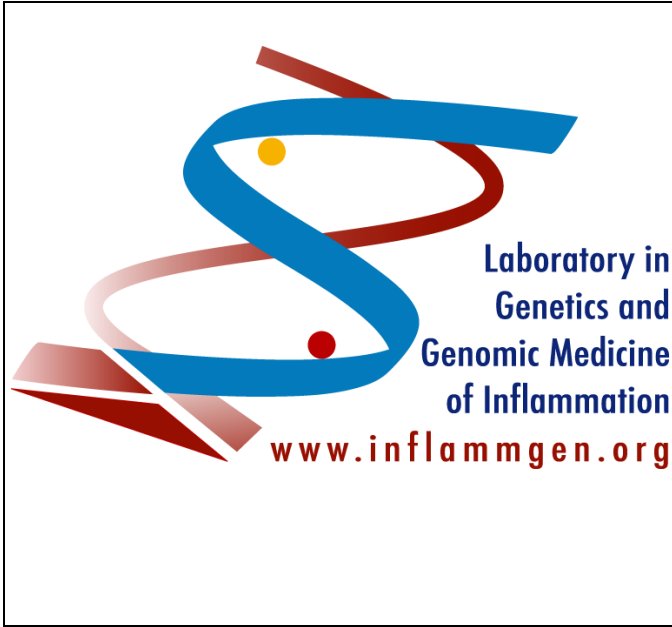
The Human Genome

300,000 - 1,000,000 SNPs
(Genetic tests) per individual



*GWA: a Revolution in
Gene Discovery*

- Risk Prediction
- Molecular classifiers of Dx
- Stratifiers for clinical trials
- Novel drug targets



Melissa Beaudoin
 Gabrielle Boucher
 Marcia Budarf, PhD
 Mariana Chagas, M.Sc.
 Guy Charron, PhD
 Geneviève David, B.Sc
 Christian Eyendja, MD
 Sylvain Foisy, PhD
 Philippe Goyette, PhD
 Catherine Labbé, MSc
 Frédéric Latour, BSc
 Marie-Pierre Lévesque, BSc
 Diana Miclaus, BA

A special thanks to:

Mark Daly (Broad Institute; MGH)
 Ramnik Xavier (CCIB, MGH)
 NIDDK IBD Genetics Consortium



Supported by:



Canadian Institutes of Health Research

Fonds de la recherche en santé

