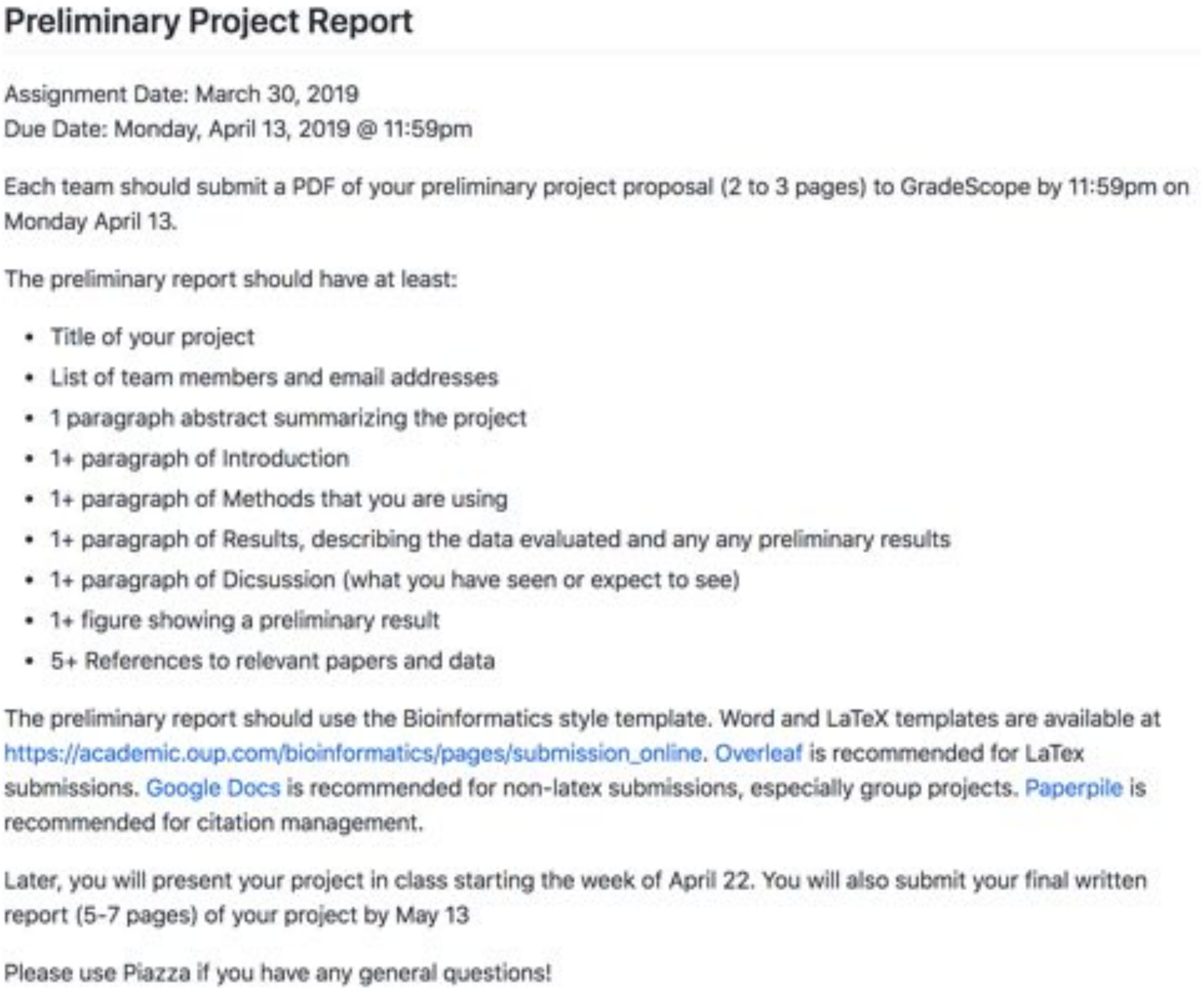
Lecture 21. Cancer Genetics

Michael Schatz

April 13, 2020

JHU 600.749: Applied Comparative Genomics



Preliminary Project Report

Assignment Date: March 30, 2019
Due Date: Monday, April 13, 2019 @ 11:59pm

Each team should submit a PDF of your preliminary project proposal (2 to 3 pages) to GradeScope by 11:59pm on Monday April 13.

The preliminary report should have at least:

- Title of your project
- List of team members and email addresses
- 1 paragraph abstract summarizing the project
- 1+ paragraph of Introduction
- 1+ paragraph of Methods that you are using
- 1+ paragraph of Results, describing the data evaluated and any any preliminary results
- 1+ paragraph of Dicsussion (what you have seen or expect to see)
- 1+ figure showing a preliminary result
- 5+ References to relevant papers and data

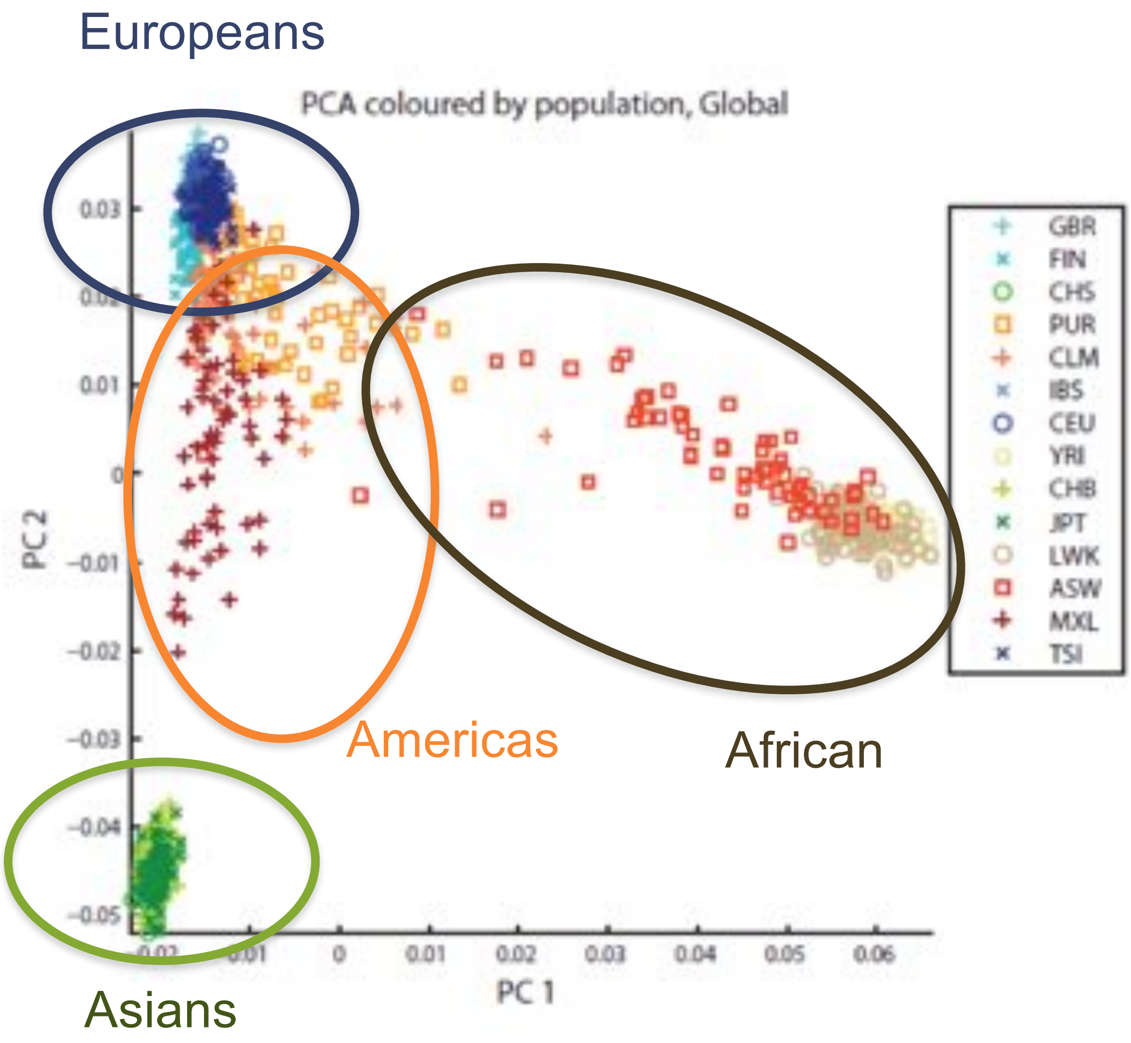
The preliminary report should use the Bioinformatics style template. Word and LaTeX templates are available at https://academic.oup.com/bioinformatics/pages/submission_online. Overleaf is recommended for LaTex submissions. Google Docs is recommended for non-latex submissions, especially group projects. Paperpile is recommended for citation management.

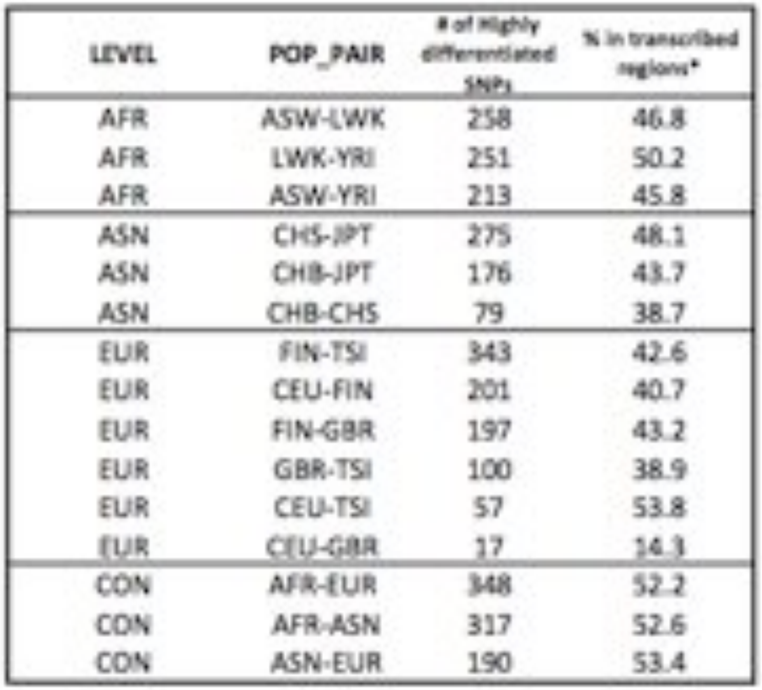
Later, you will present your project in class starting the week of April 22. You will also submit your final written report (5-7 pages) of your project by May 13

Please use Piazza if you have any general questions!

Part 1: Review

Variation across populations

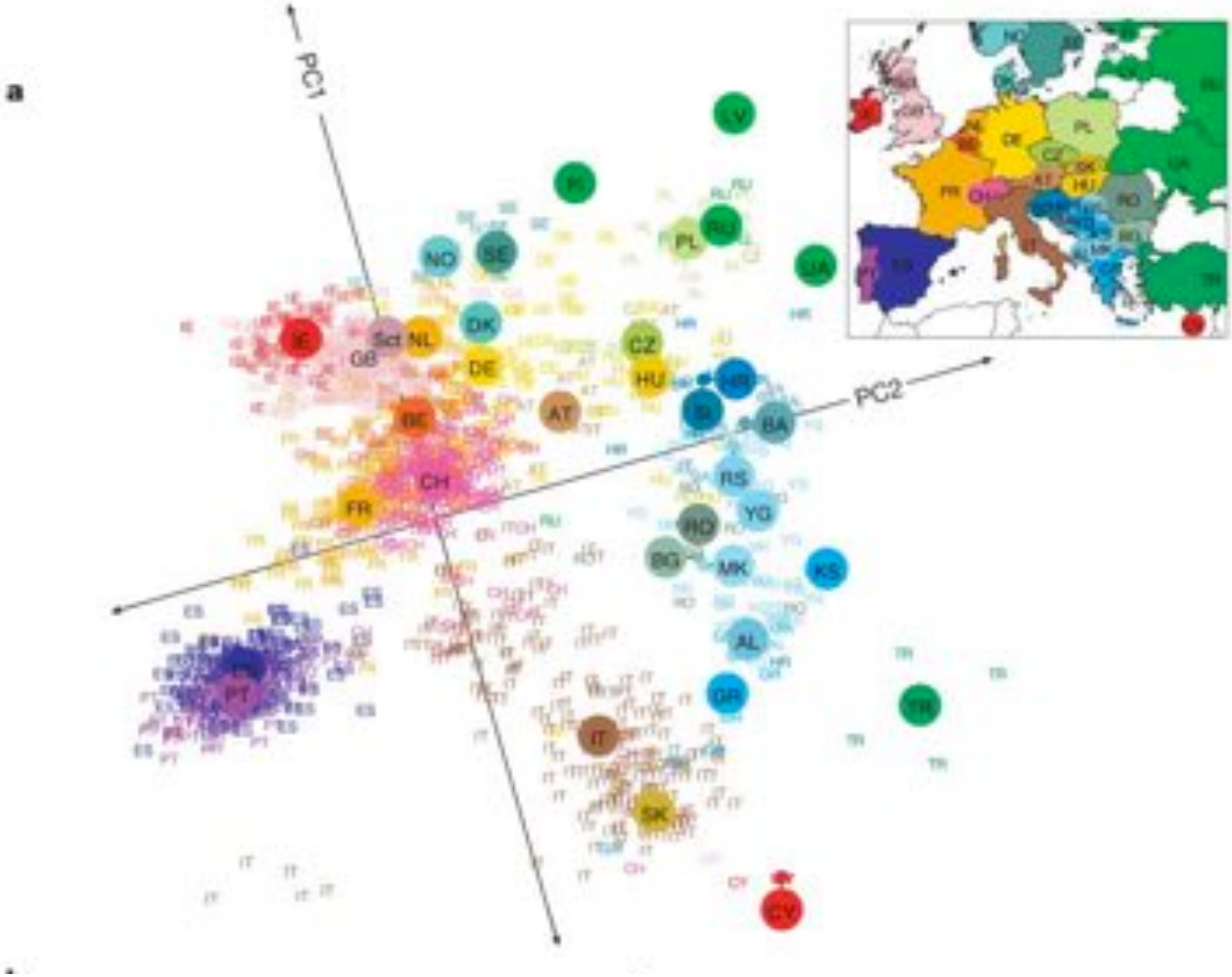




LEVEL	POP_PAIR	# of Highly differentiated SNPs	% in transcribed regions*
AFR	ASW-LWK	258	46.8
AFR	LWK-YRI	251	50.2
AFR	ASW-YRI	213	45.8
ASN	CHS-JPT	275	48.1
ASN	CHB-JPT	176	43.7
ASN	CHB-CHS	79	38.7
EUR	FIN-TSI	343	42.6
EUR	CEU-FIN	201	40.7
EUR	FIN-GBR	197	43.2
EUR	GBR-TSI	100	38.9
EUR	CEU-TSI	57	53.8
EUR	CEU-GBR	17	14.3
CON	AFR-EUR	348	52.2
CON	AFR-ASN	317	52.6
CON	ASN-EUR	190	53.4

Table S12A **Summary of sites showing high levels of population differentiation**

- Not a single variant 100% unique to a given population
- 17% of low-frequency variants (.5-5% pop. freq) observed in a single ancestry group
- 50% of rare variants (<.5%) observed in a single population

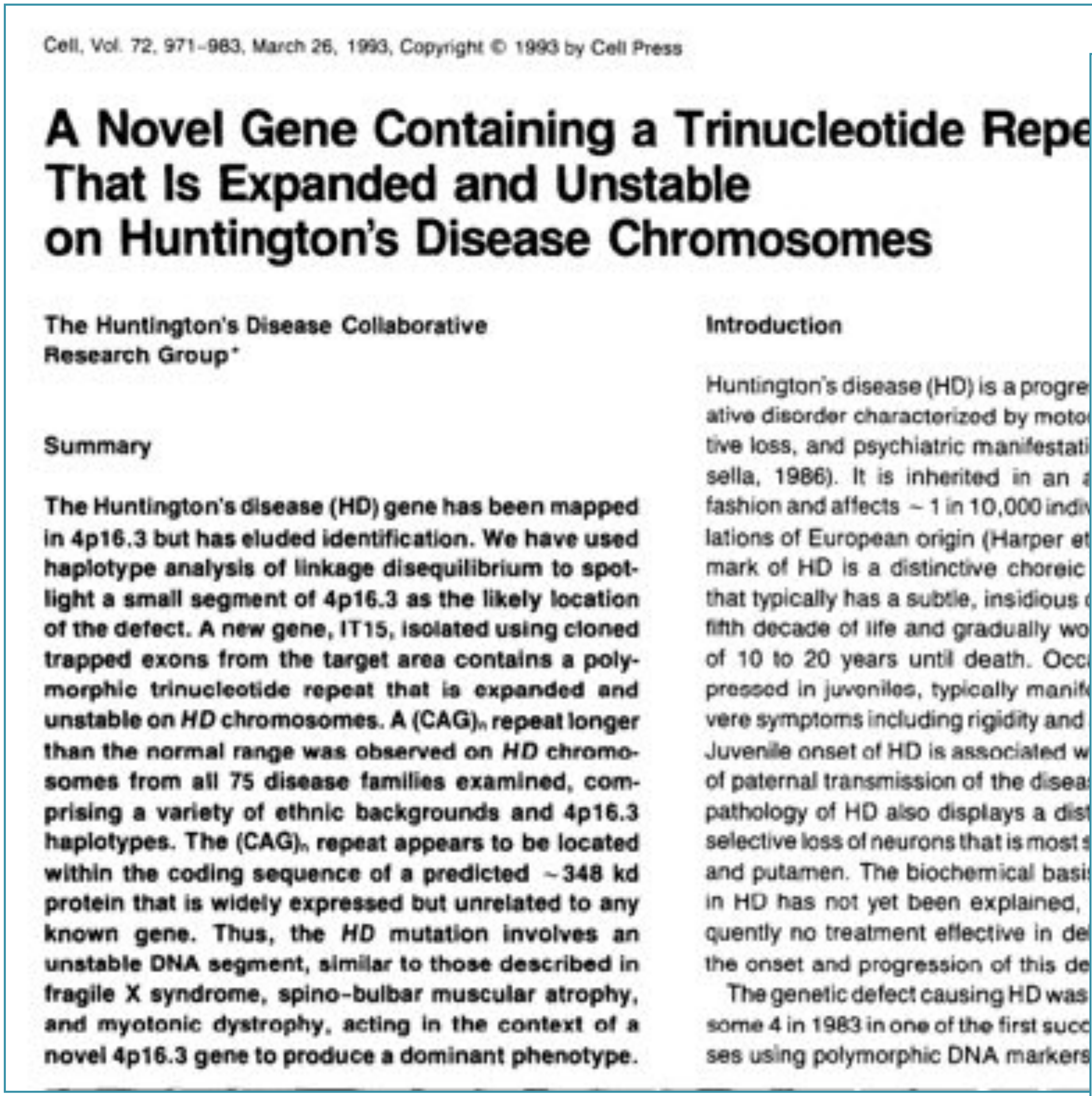


Genes mirror geography within Europe

Novembre et al (2008) Nature. doi: 10.1038/nature07331

Part 2:
Inherited Diseases

Huntington's Disease



Cell, Vol. 72, 971–983, March 26, 1993, Copyright © 1993 by Cell Press

A Novel Gene Containing a Trinucleotide Repe
That Is Expanded and Unstable
on Huntington's Disease Chromosomes

The Huntington's Disease Collaborative
Research Group*

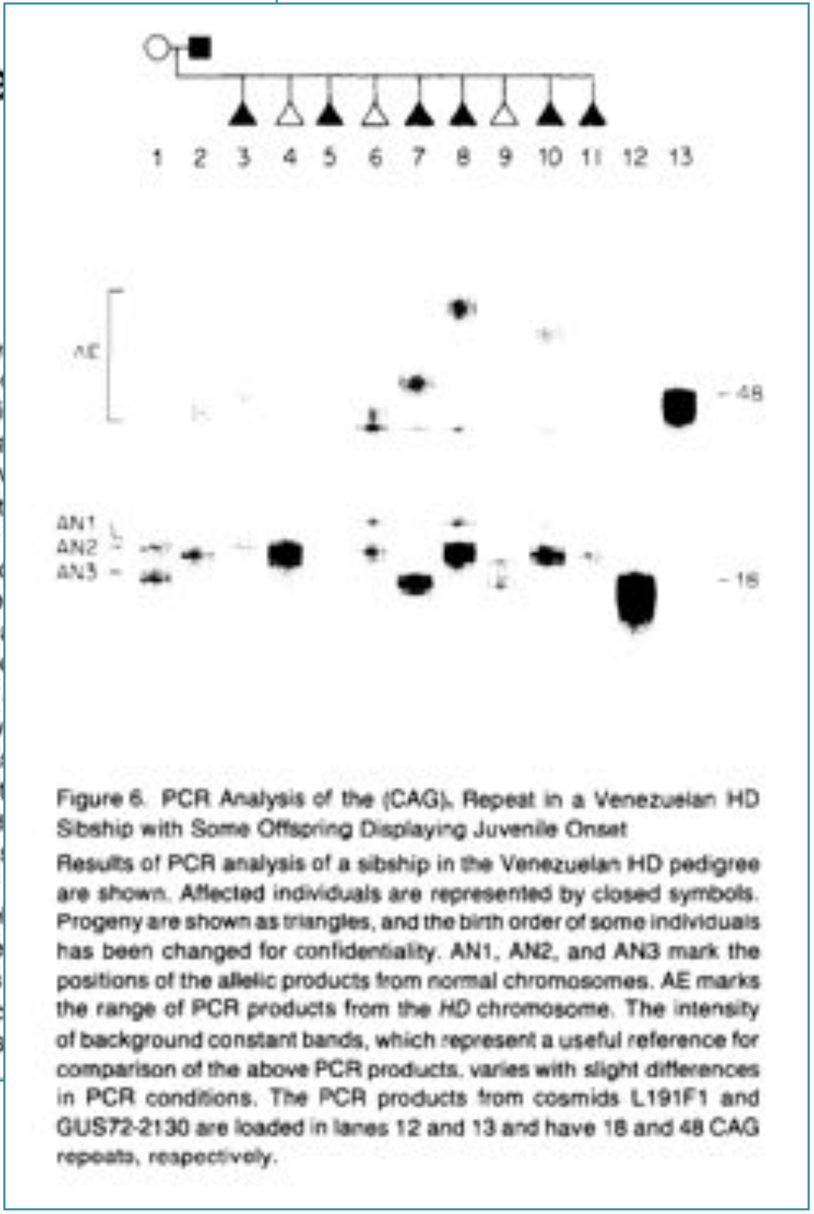
Summary

The Huntington's disease (HD) gene has been mapped in 4p16.3 but has eluded identification. We have used haplotype analysis of linkage disequilibrium to spotlight a small segment of 4p16.3 as the likely location of the defect. A new gene, IT15, isolated using cloned trapped exons from the target area contains a polymorphic trinucleotide repeat that is expanded and unstable on *HD* chromosomes. A $(CAG)_n$ repeat longer than the normal range was observed on *HD* chromosomes from all 75 disease families examined, comprising a variety of ethnic backgrounds and 4p16.3 haplotypes. The $(CAG)_n$ repeat appears to be located within the coding sequence of a predicted ~348 kd protein that is widely expressed but unrelated to any known gene. Thus, the *HD* mutation involves an unstable DNA segment, similar to those described in fragile X syndrome, spino-bulbar muscular atrophy, and myotonic dystrophy, acting in the context of a novel 4p16.3 gene to produce a dominant phenotype.

Introduction

Huntington's disease (HD) is a progre
ative disorder characterized by moto
tive loss, and psychiatric manifestati
sella, 1986). It is inherited in an a
fashion and affects ~1 in 10,000 indiv
lations of European origin (Harper et
mark of HD is a distinctive choreic
that typically has a subtle, insidious c
fifth decade of life and gradually wo
of 10 to 20 years until death. Occ
pressed in juveniles, typically manif
vere symptoms including rigidity and
Juvenile onset of HD is associated w
of paternal transmission of the disea
pathology of HD also displays a dist
selective loss of neurons that is most s
and putamen. The biochemical basi
in HD has not yet been explained,
quently no treatment effective in de
the onset and progression of this de

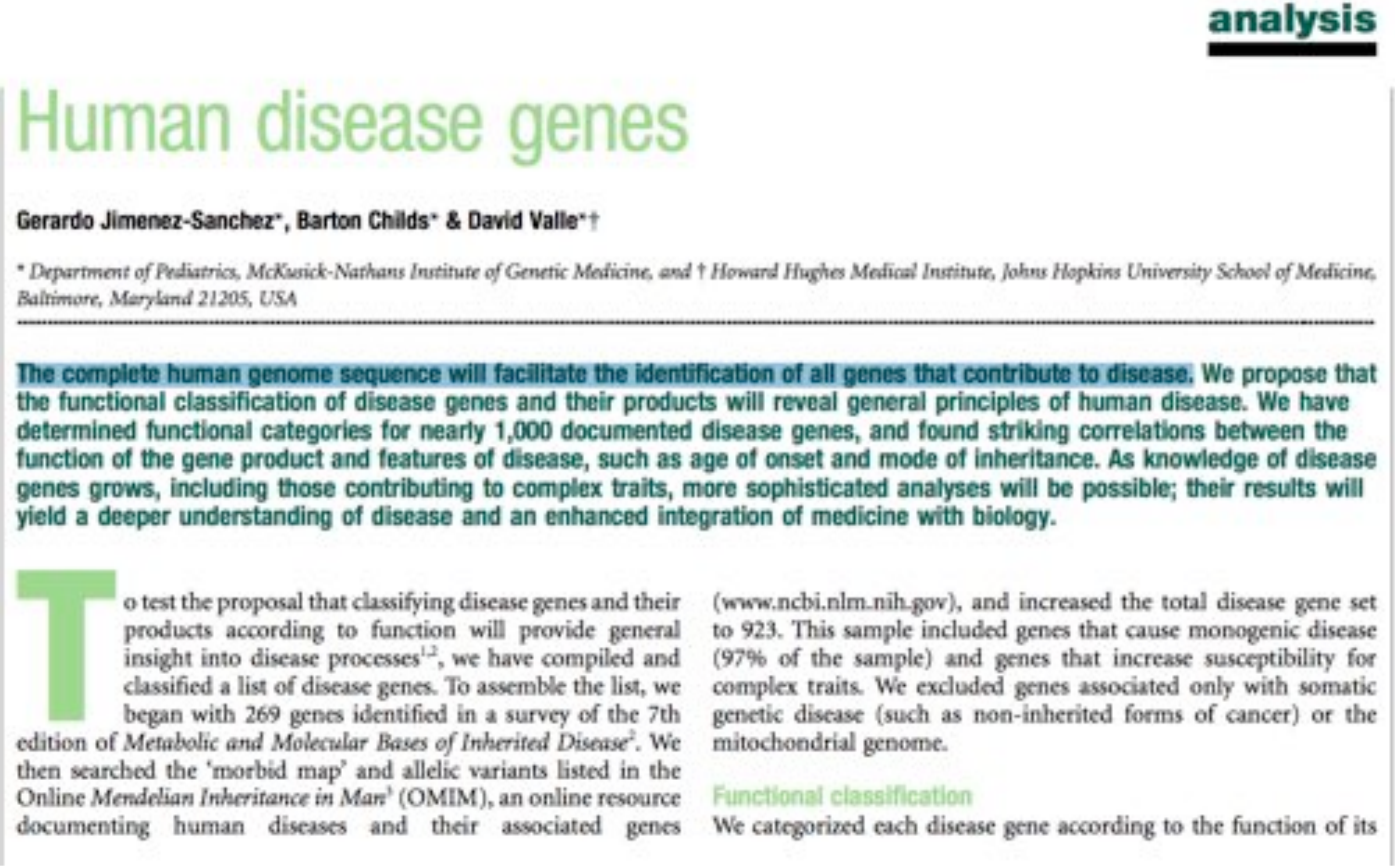
The genetic defect causing HD was
some 4 in 1983 in one of the first succ
ses using polymorphic DNA markers



1
2
3
4
5
6
7
8
9
10
11
12
13
AE
48
AN1
AN2
AN3
16

Figure 6. PCR Analysis of the $(CAG)_n$ Repeat in a Venezuelan HD Sibship with Some Offspring Displaying Juvenile Onset

Results of PCR analysis of a sibship in the Venezuelan HD pedigree are shown. Affected individuals are represented by closed symbols. Progeny are shown as triangles, and the birth order of some individuals has been changed for confidentiality. AN1, AN2, and AN3 mark the positions of the allelic products from normal chromosomes. AE marks the range of PCR products from the *HD* chromosome. The intensity of background constant bands, which represent a useful reference for comparison of the above PCR products, varies with slight differences in PCR conditions. The PCR products from cosmids L191F1 and GUS72-2130 are loaded in lanes 12 and 13 and have 18 and 48 CAG repeats, respectively.



analysis

Human disease genes

Gerardo Jimenez-Sanchez[*], Barton Childs[*] & David Valle[*†]

[*] Department of Pediatrics, McKusick-Nathans Institute of Genetic Medicine, and † Howard Hughes Medical Institute, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

The complete human genome sequence will facilitate the identification of all genes that contribute to disease. We propose that the functional classification of disease genes and their products will reveal general principles of human disease. We have determined functional categories for nearly 1,000 documented disease genes, and found striking correlations between the function of the gene product and features of disease, such as age of onset and mode of inheritance. As knowledge of disease genes grows, including those contributing to complex traits, more sophisticated analyses will be possible; their results will yield a deeper understanding of disease and an enhanced integration of medicine with biology.

To test the proposal that classifying disease genes and their products according to function will provide general insight into disease processes[1,2], we have compiled and classified a list of disease genes. To assemble the list, we began with 269 genes identified in a survey of the 7th edition of *Metabolic and Molecular Bases of Inherited Disease*[2]. We then searched the 'morbid map' and allelic variants listed in the Online *Mendelian Inheritance in Man*[3] (OMIM), an online resource documenting human diseases and their associated genes (www.ncbi.nlm.nih.gov), and increased the total disease gene set to 923. This sample included genes that cause monogenic disease (97% of the sample) and genes that increase susceptibility for complex traits. We excluded genes associated only with somatic genetic disease (such as non-inherited forms of cancer) or the mitochondrial genome.

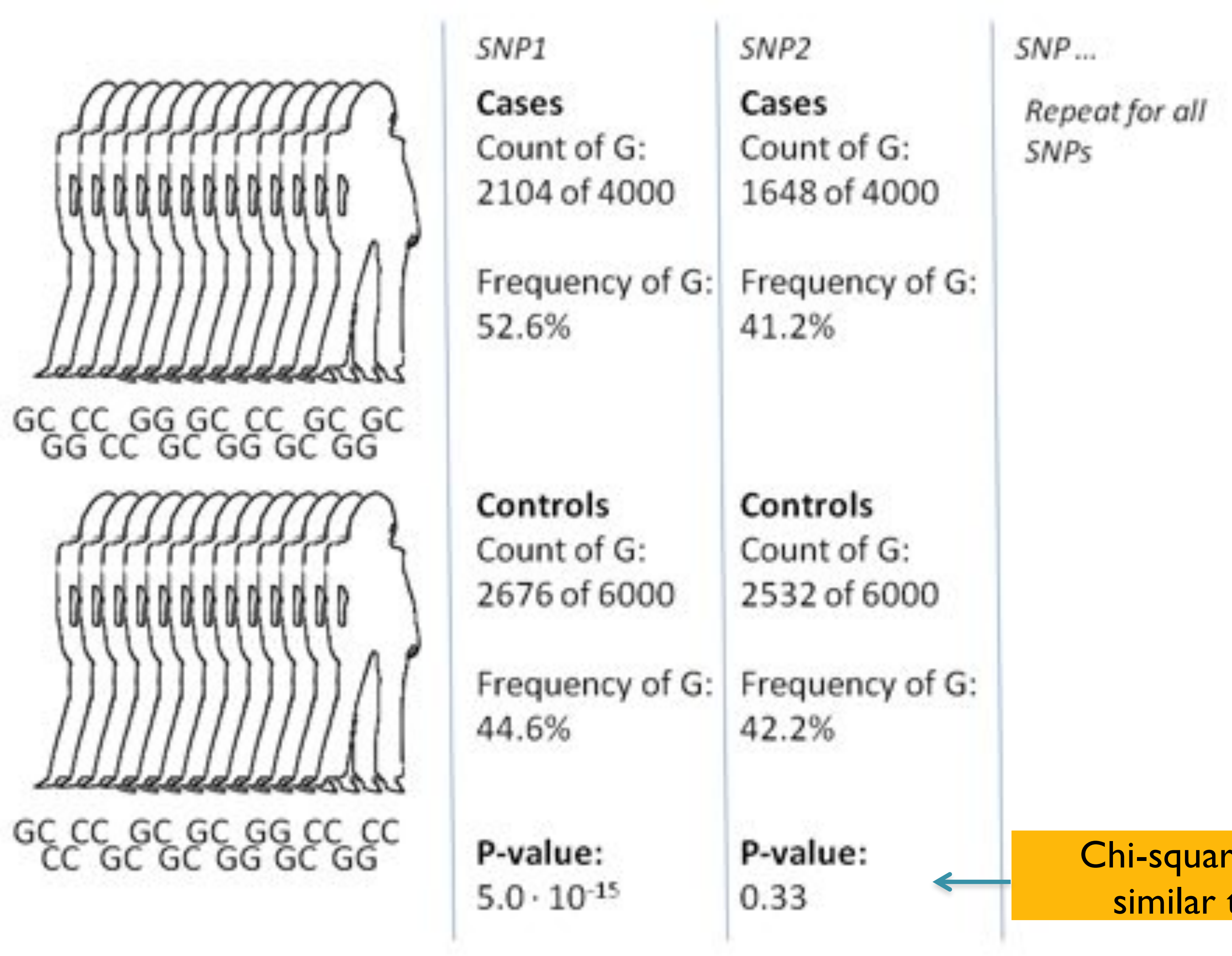
Functional classification

We categorized each disease gene according to the function of its

Human disease genes

Jimenez-Sanchez, G., Childs, B. & Valle, D. (2001) Nature 409, 853–855

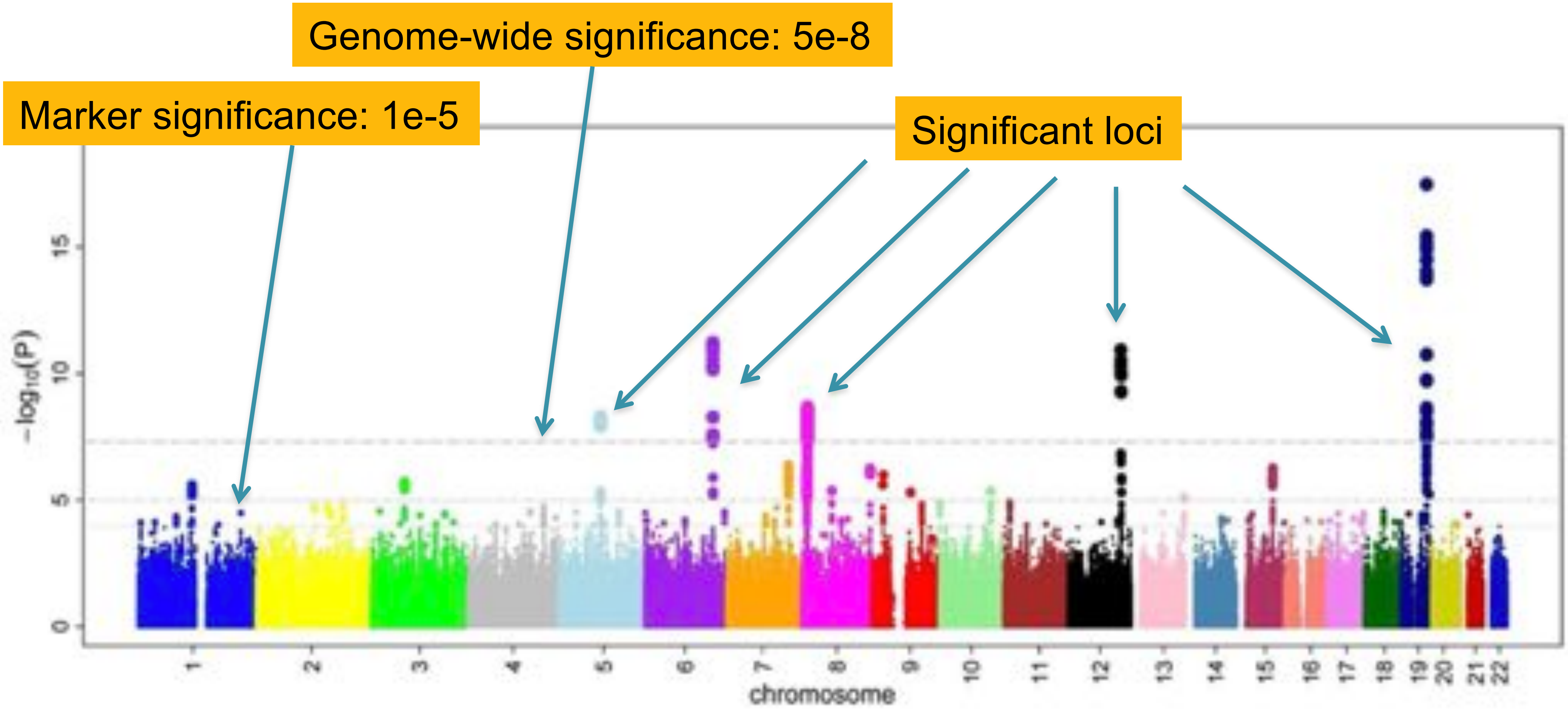
Genome Wide Association (GWAS)



SNP1	*SNP2*	*SNP ...*
Cases Count of G: 2104 of 4000	**Cases** Count of G: 1648 of 4000	*Repeat for all SNPs*
Frequency of G: 52.6%	Frequency of G: 41.2%	
Controls Count of G: 2676 of 6000	**Controls** Count of G: 2532 of 6000	
Frequency of G: 44.6%	Frequency of G: 42.2%	
P-value: $5.0 \cdot 10^{-15}$	**P-value:** 0.33	

Chi-squared or similar test

Manhattan Plot



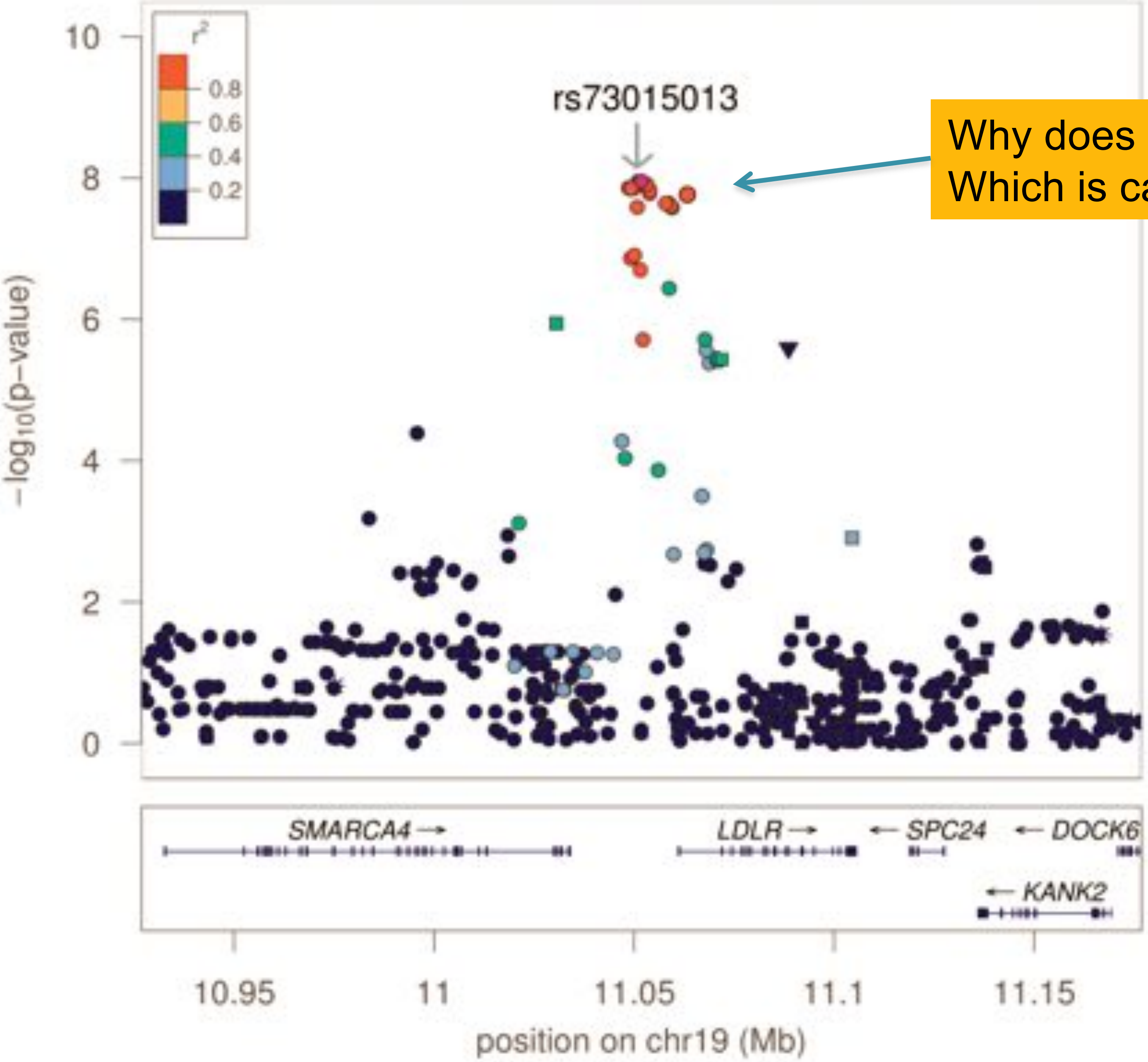
Genome-wide significance: 5e-8

Marker significance: 1e-5

Significant loci

Four Novel Loci (19q13, 6q24, 12q24, and 5q14) Influence the Microcirculation In Vivo
Ikram et al (2010) PLOS Genetics. doi: 10.1371/journal.pgen.1001184

Regional Association Plot



Why does it look like this?
Which is causal?

GWAS Catalog

As of 2020-03-08, the GWAS Catalog contains 4493 publications and 179364 associations.

http://www.ebi.ac.uk/gwas/diagram

GWAS In Crisis

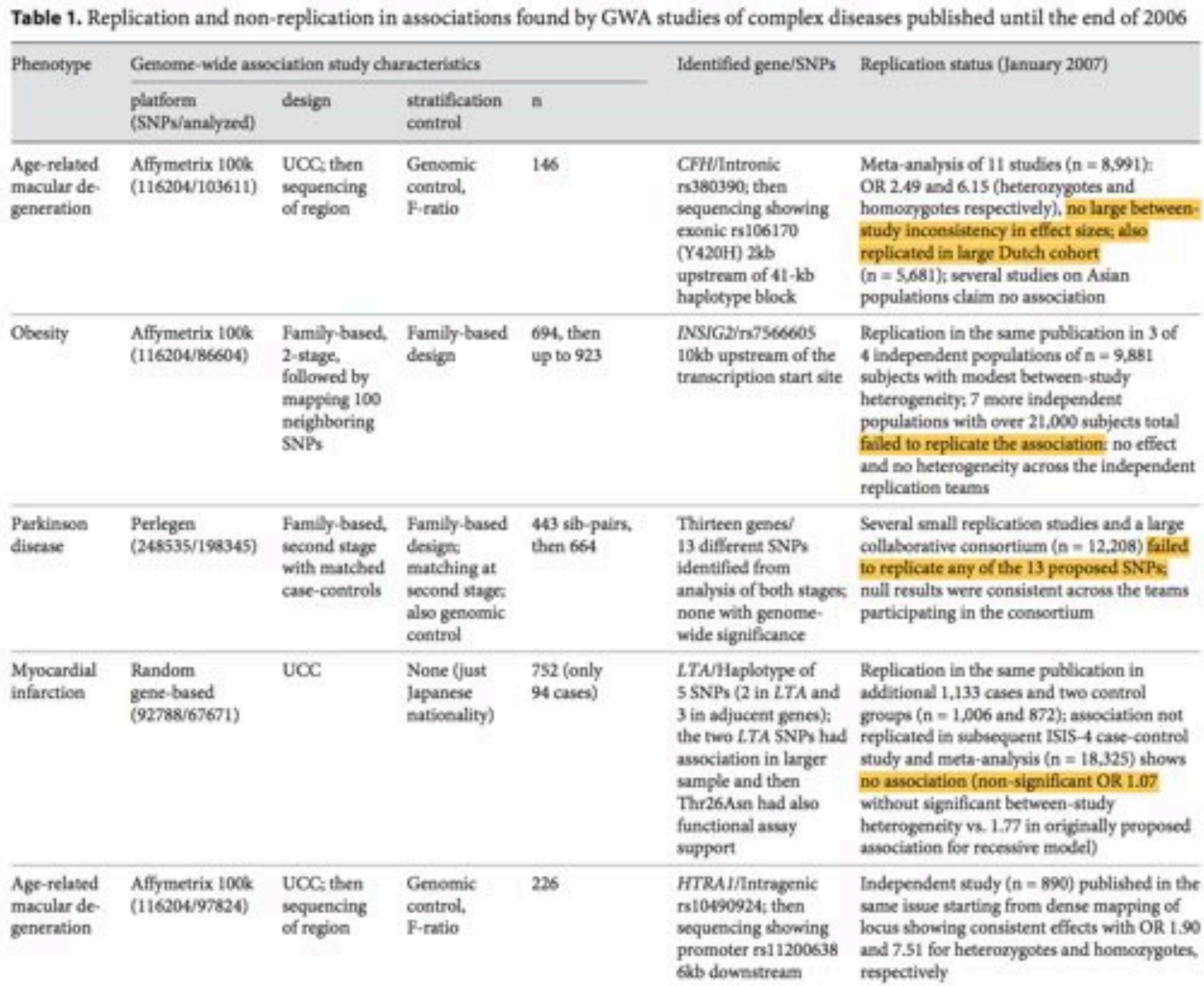


Table 1. Replication and non-replication in associations found by GWA studies of complex diseases published until the end of 2006

Phenotype	Genome-wide association study characteristics				Identified gene/SNPs	Replication status (January 2007)
	platform (SNPs/analyzed)	design	stratification control	n		
Age-related macular degeneration	Affymetrix 100k (116204/103611)	UCC; then sequencing of region	Genomic control, F-ratio	146	*CFH*/Intronic rs380390; then sequencing showing exonic rs106170 (Y420H) 2kb upstream of 41-kb haplotype block	Meta-analysis of 11 studies (n = 8,991): OR 2.49 and 6.15 (heterozygotes and homozygotes respectively), no large between-study inconsistency in effect sizes; also replicated in large Dutch cohort (n = 5,681); several studies on Asian populations claim no association
Obesity	Affymetrix 100k (116204/86604)	Family-based, 2-stage, followed by mapping 100 neighboring SNPs	Family-based design	694, then up to 923	*INSIG2*/rs7566605 10kb upstream of the transcription start site	Replication in the same publication in 3 of 4 independent populations of n = 9,881 subjects with modest between-study heterogeneity; 7 more independent populations with over 21,000 subjects total failed to replicate the association; no effect and no heterogeneity across the independent replication teams
Parkinson disease	Perlegen (248535/198345)	Family-based, second stage with matched case-controls	Family-based design; matching at second stage; also genomic control	443 sib-pairs, then 664	Thirteen genes/ 13 different SNPs identified from analysis of both stages; none with genome-wide significance	Several small replication studies and a large collaborative consortium (n = 12,208) failed to replicate any of the 13 proposed SNPs; null results were consistent across the teams participating in the consortium
Myocardial infarction	Random gene-based (92788/67671)	UCC	None (just Japanese nationality)	752 (only 94 cases)	*LTA*/Haplotype of 5 SNPs (2 in *LTA* and 3 in adjucent genes); the two *LTA* SNPs had association in larger sample and then Thr26Asn had also functional assay support	Replication in the same publication in additional 1,133 cases and two control groups (n = 1,006 and 872); association not replicated in subsequent ISIS-4 case-control study and meta-analysis (n = 18,325) shows no association (non-significant OR 1.07 without significant between-study heterogeneity vs. 1.77 in originally proposed association for recessive model)
Age-related macular degeneration	Affymetrix 100k (116204/97824)	UCC; then sequencing of region	Genomic control, F-ratio	226	*HTRA1*/Intragenic rs10490924; then sequencing showing promoter rs11200638 6kb downstream	Independent study (n = 890) published in the same issue starting from dense mapping of locus showing consistent effects with OR 1.90 and 7.51 for heterozygotes and homozygotes, respectively

Non-Replication and Inconsistency in the Genome-Wide Association Setting
Ioannidis (2007) Hum Hered 2007;64:203–213 https://doi.org/10.1159/000103512

Missing Heritability

NEWS FEATURE PERSONAL GENOMES NATURE|Vol 456|6 November 2008

The case of the missing heritability

When scientists opened up the human genome, they expected to find the genetic components of common traits and diseases. But they were nowhere to be seen. **Brendan Maher** shines a light on six places where the missing loot could be stashed away.

If you want to predict how tall your children might one day be, a good bet would be to look in the mirror, and at your mate. Studies going back almost a century have estimated that height is 80–90% heritable. So if 29 centimetres separate the tallest 5% of a population from the shortest, then genetics would account for as many as 27 of them[1].

This year, three groups of researchers[2-4] scoured the genomes of huge populations (the largest study[4] looked at more than 30,000 people) for genetic variants associated with the height differences. More than 40 turned up.

But there was a problem: the variants had tiny effects. Altogether, they accounted for little more than 5% of height's heritability — just 6 centimetres by the calculations above. Even though these genome-wide association studies (GWAS) turned up dozens of variants, they did "very little of the prediction that you would do just by asking people how tall their parents are", says Joel Hirschhorn at the Broad Institute in Cambridge, Massachusetts, who led one of the studies[4].

Height isn't the only trait in which genes have gone missing, nor is it the most important. Studies looking at similarities between identical and fraternal twins estimate heritability at more than 90% for autism[5] and more than 80% for schizophrenia[6]. And genetics makes a major contribution to disorders such as obesity, diabetes and heart disease. GWAS, one of the most celebrated techniques of the past five years, promised to deliver many of the genes involved (see "Where's the reward?", page 20). And to some extent they have, identifying more than 400 genetic variants that contribute to a variety of traits and common diseases. But even when dozens of genes have been linked to a trait, both the individual and cumulative effects are disappointingly small and nowhere near enough to explain earlier estimates of heritability. "It is the big topic in the genetics of common disease right now," says Francis Collins, former head of the National Human Genome Research Institute (NHGRI) in Bethesda, Maryland. The unexpected results left researchers at a point "where we all had to scratch our heads and say, 'Huh?'", he says.

Although flummoxed by this missing heritability, geneticists remain optimistic that they can find more of it. "These are very early days, and there are things that are doable in the next year or two that may well explain another sizeable chunk of heritability," says Hirschhorn. So where might it be hiding?

18

© 2008 Macmillan Publishers Limited. All rights reserved

"Three groups of researchers scoured the genomes of huge populations (>30,000 people) for genetic variants associated with the height differences. More than 40 turned up. ***But there was a problem: the variants had tiny effects.*** Altogether, they accounted for little more than 5% of height's heritability"

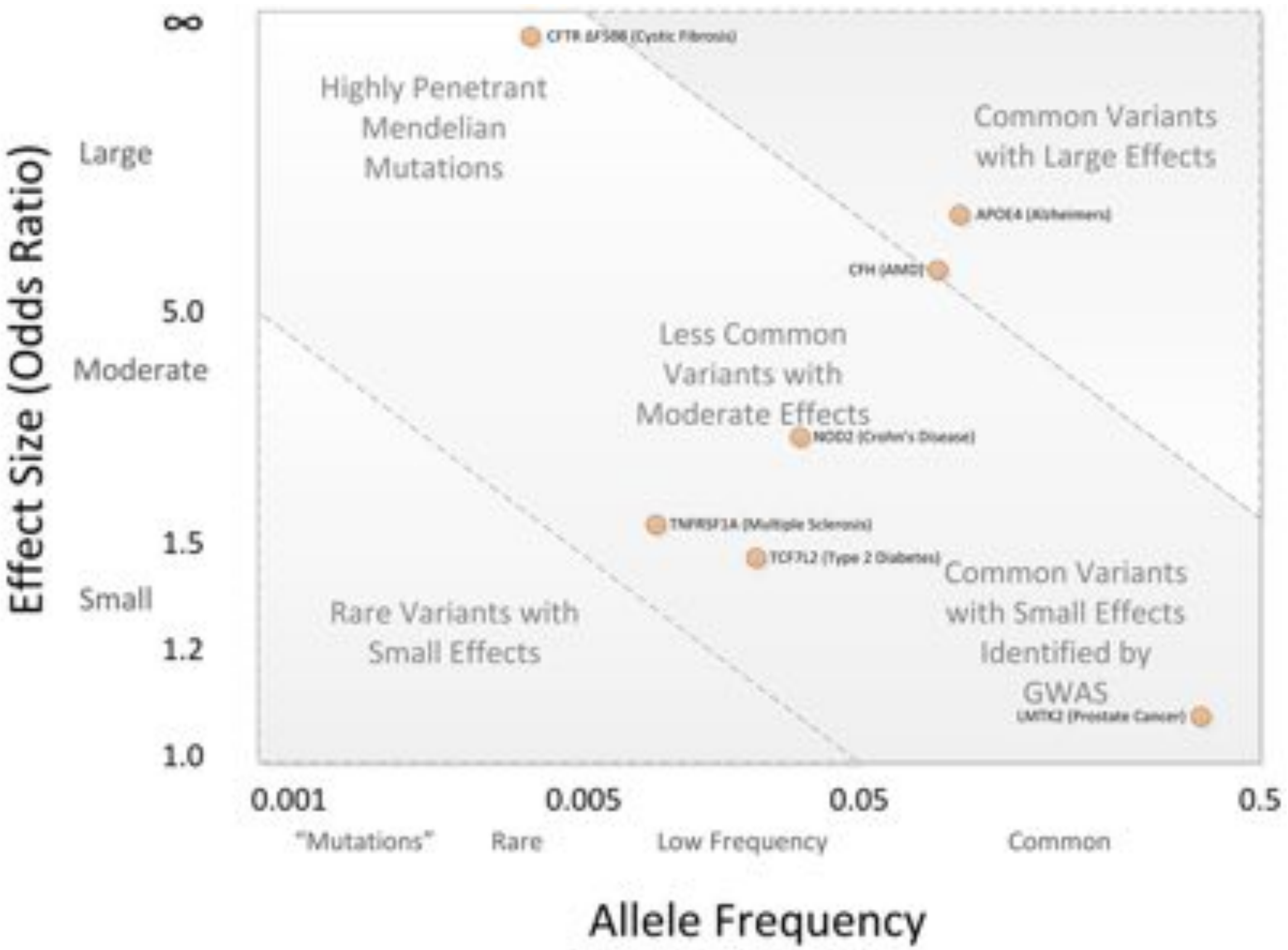
- ***Rare, moderately penetrant or common, weakly penetrant variants?***

- ***CNVs and SVs?***

- ***Epistasis (multiple genes working together)?***

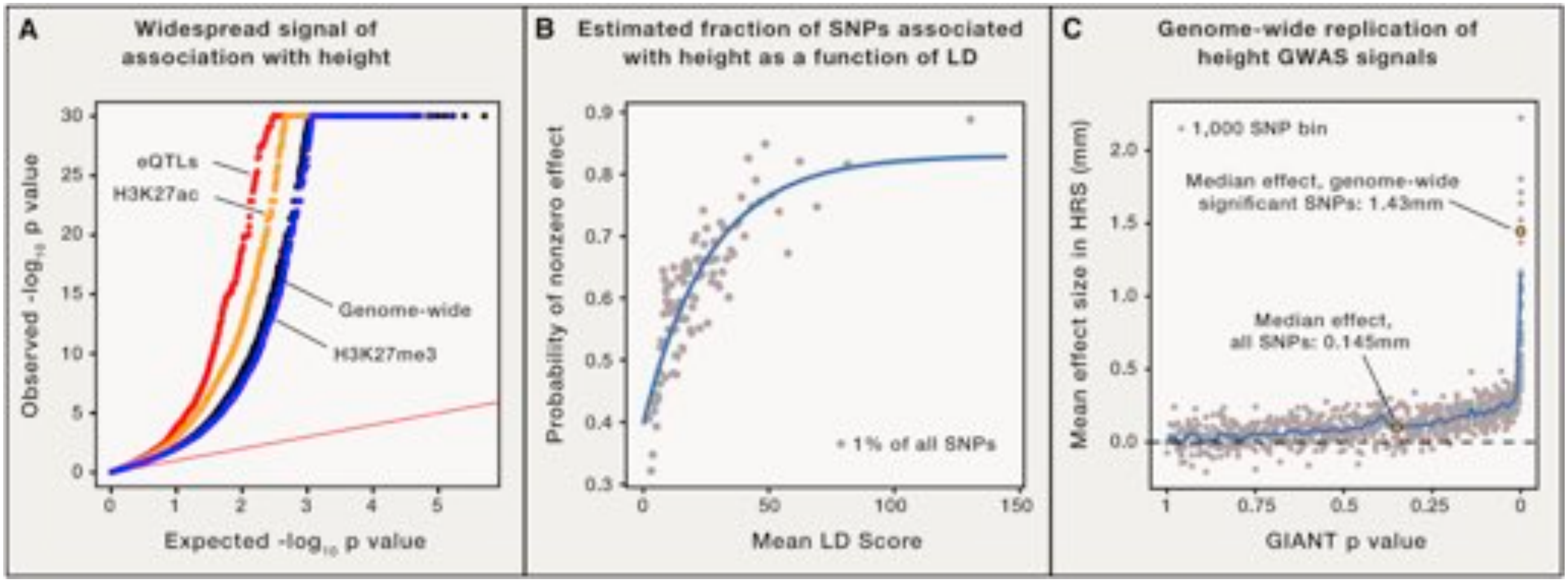
- ***Epigenetic effects, especially in utero?***

Penetrance & Allele Frequency



Penetrance: The proportion of individuals with a specific genotype who manifest the genotype at the phenotypic level.

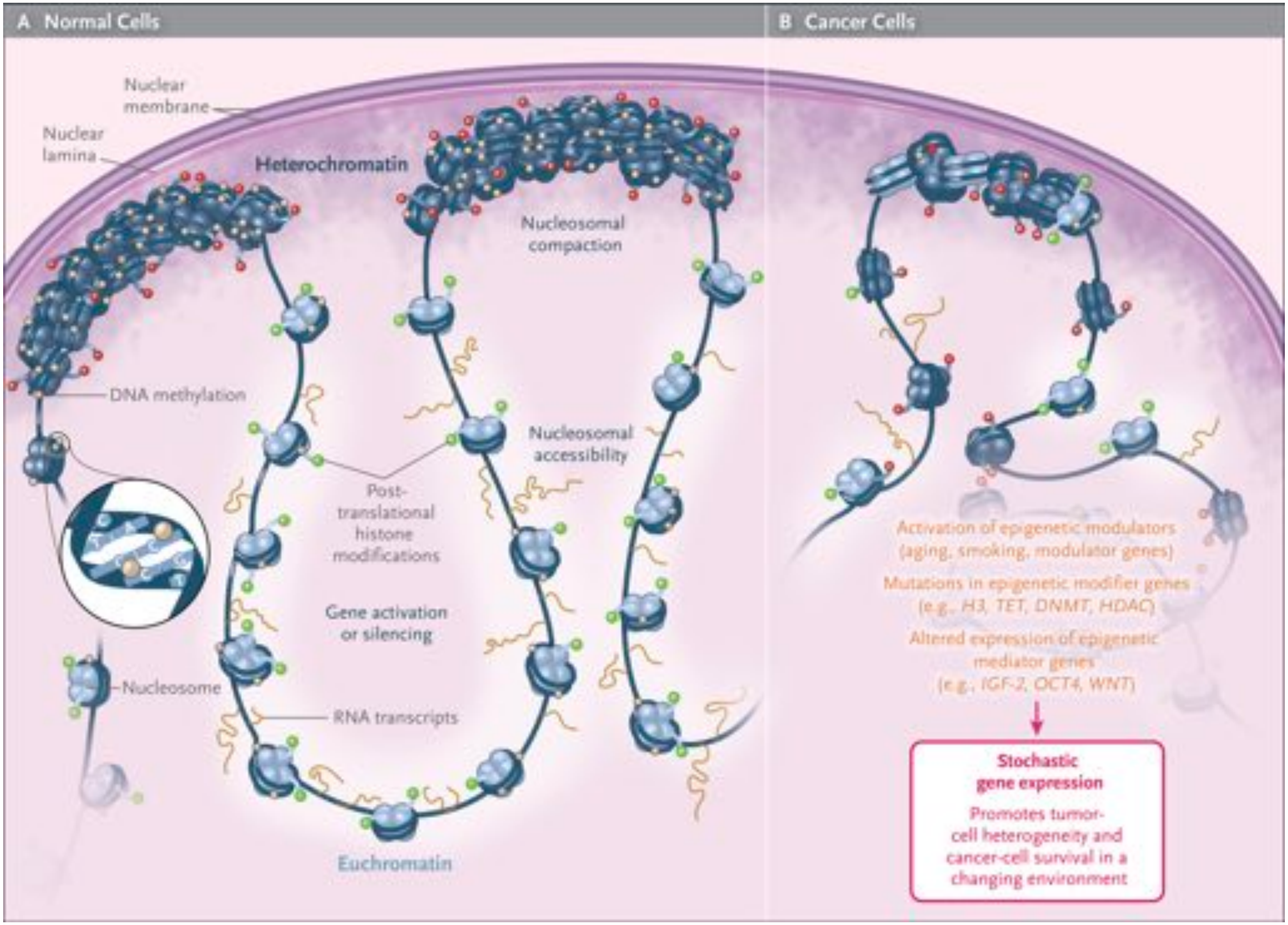
Omnigenics



A central goal of genetics is to understand the links between genetic variation and disease. Intuitively, one might expect disease-causing variants to cluster into key pathways that drive disease etiology. But for complex traits, association signals tend to be spread across most of the genome—including near many genes without an obvious connection to disease. **We propose that gene regulatory networks are sufficiently interconnected such that all genes expressed in disease-relevant cells are liable to affect the functions of core disease-related genes and that most heritability can be explained by effects on genes outside core pathways. We refer to this hypothesis as an “omnigenic” model.**

An Expanded View of Complex Traits: From Polygenic to Omnigenic
Boyle, Li, Pritchard (2017) Cell. https://doi.org/10.1016/j.cell.2017.05.038

Epigenetic Factors


A Normal Cells
B Cancer Cells
Nuclear membrane
Nuclear lamina
Heterochromatin
Nucleosomal compaction
DNA methylation
Nucleosomal accessibility
Post-translational histone modifications
Gene activation or silencing
Nucleosome
RNA transcripts
Euchromatin
Activation of epigenetic modulators (aging, smoking, modulator genes)
Mutations in epigenetic modifier genes (e.g., H3, TET, DNMT, HDAC)
Altered expression of epigenetic mediator genes (e.g., IGF-2, OCT4, WNT)
Stochastic gene expression
Promotes tumor-cell heterogeneity and cancer-cell survival in a changing environment

The Key Role of Epigenetics in Human Disease Prevention and Mitigation
Feinberg (2018) NEJM. doi: 10.1056/NEJMra1402513

Needles in stacks of needles

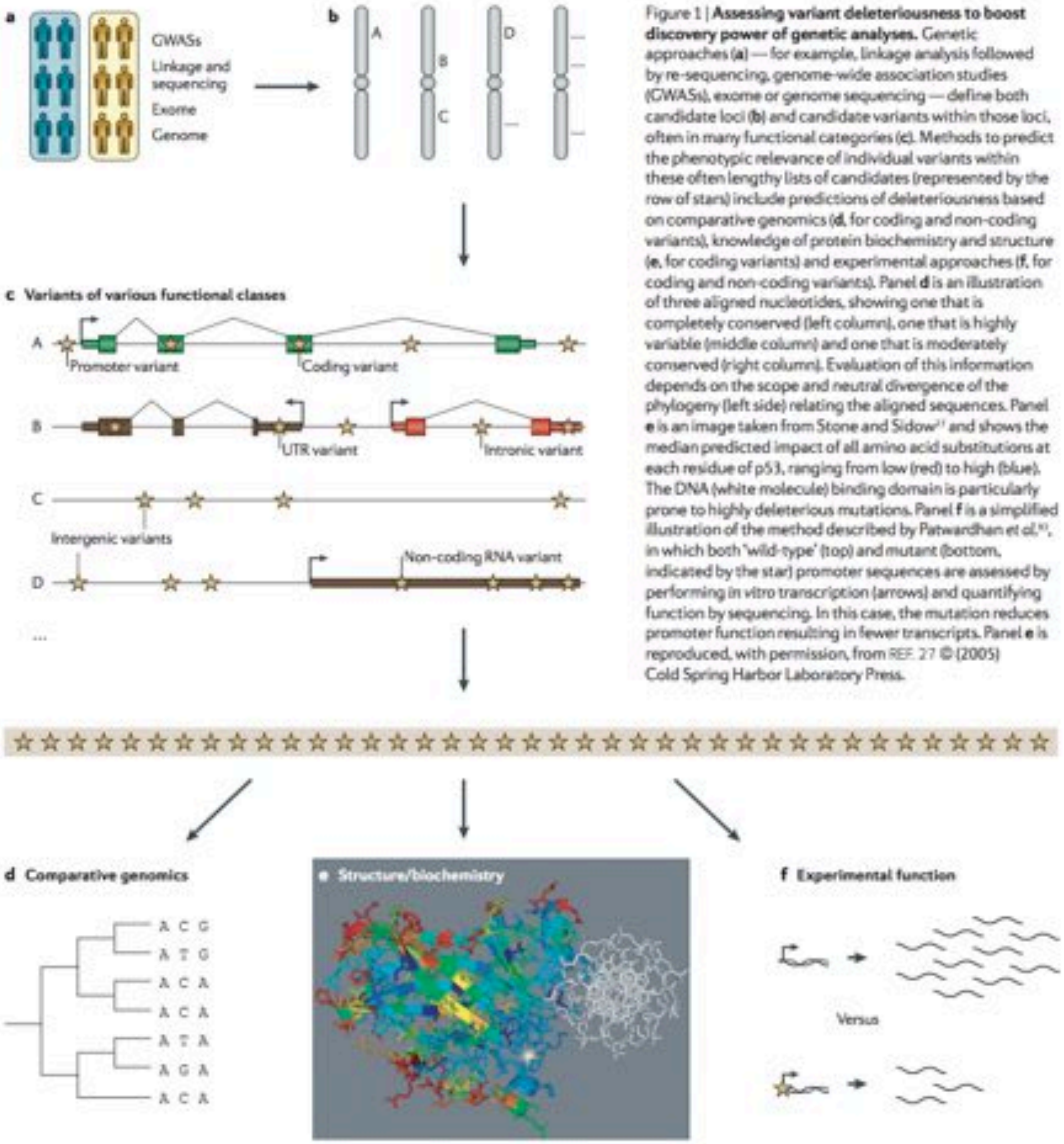
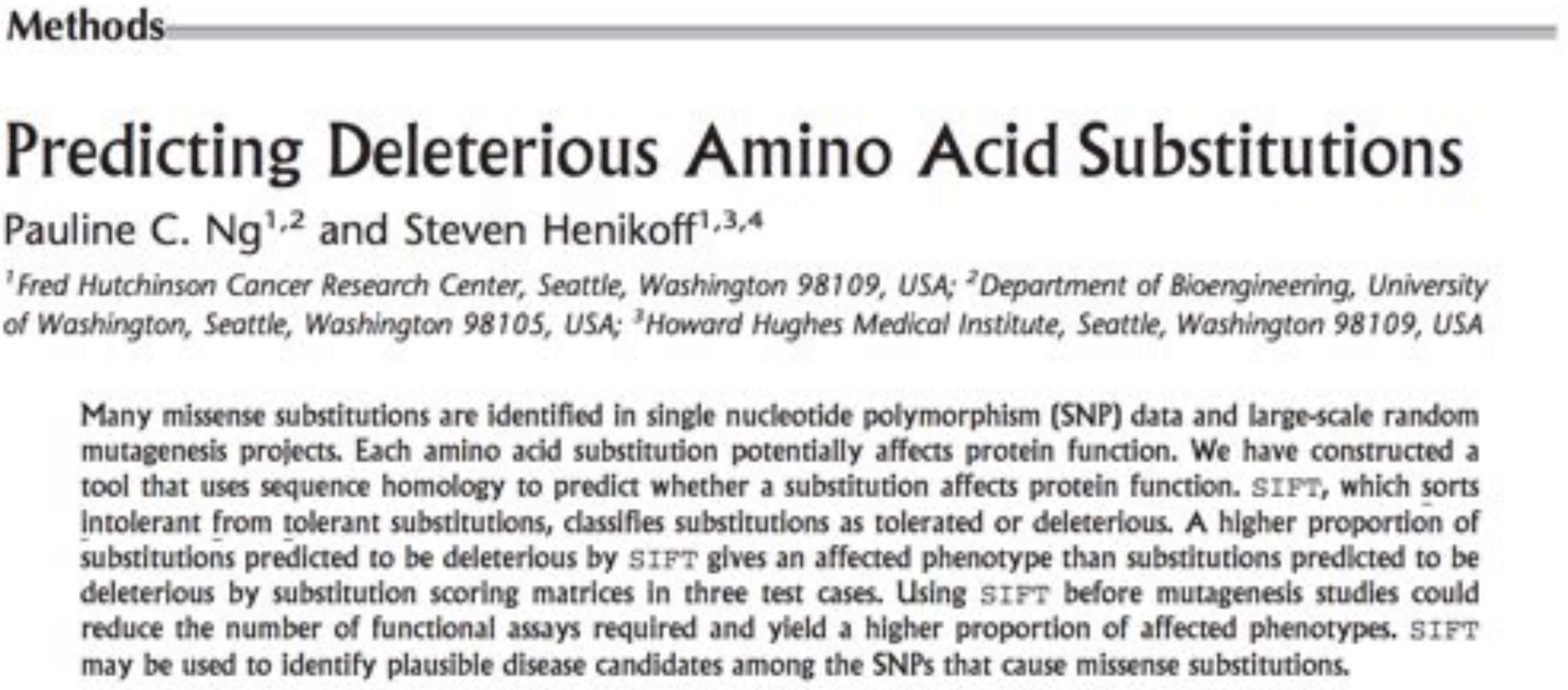


Figure 1 | **Assessing variant deleteriousness to boost discovery power of genetic analyses.** Genetic approaches (**a**) — for example, linkage analysis followed by re-sequencing, genome-wide association studies (GWASs), exome or genome sequencing — define both candidate loci (**b**) and candidate variants within those loci, often in many functional categories (**c**). Methods to predict the phenotypic relevance of individual variants within these often lengthy lists of candidates (represented by the row of stars) include predictions of deleteriousness based on comparative genomics (**d**, for coding and non-coding variants), knowledge of protein biochemistry and structure (**e**, for coding variants) and experimental approaches (**f**, for coding and non-coding variants). Panel **d** is an illustration of three aligned nucleotides, showing one that is completely conserved (left column), one that is highly variable (middle column) and one that is moderately conserved (right column). Evaluation of this information depends on the scope and neutral divergence of the phylogeny (left side) relating the aligned sequences. Panel **e** is an image taken from Stone and Sidow[27] and shows the median predicted impact of all amino acid substitutions at each residue of p53, ranging from low (red) to high (blue). The DNA (white molecule) binding domain is particularly prone to highly deleterious mutations. Panel **f** is a simplified illustration of the method described by Patwardhan *et al.*[41], in which both 'wild-type' (top) and mutant (bottom, indicated by the star) promoter sequences are assessed by performing in vitro transcription (arrows) and quantifying function by sequencing. In this case, the mutation reduces promoter function resulting in fewer transcripts. Panel **e** is reproduced, with permission, from REF. 27 © (2005) Cold Spring Harbor Laboratory Press.

Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data

Cooper & Shendure (2011) Nature Reviews Genetics.



Methods

Predicting Deleterious Amino Acid Substitutions

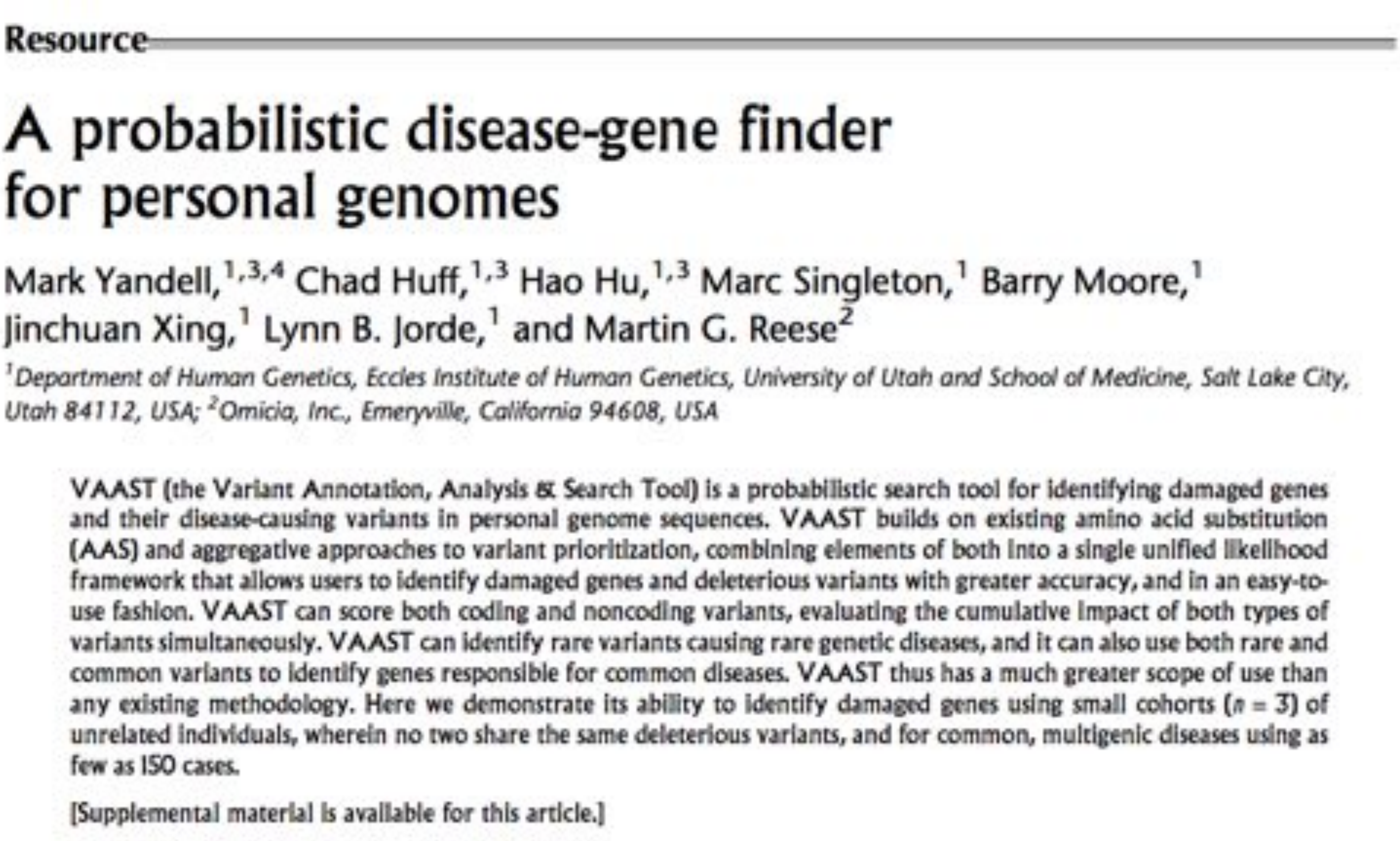
Pauline C. Ng[1,2] and Steven Henikoff[1,3,4]

[1]Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA; [2]Department of Bioengineering, University of Washington, Seattle, Washington 98105, USA; [3]Howard Hughes Medical Institute, Seattle, Washington 98109, USA

Many missense substitutions are identified in single nucleotide polymorphism (SNP) data and large-scale random mutagenesis projects. Each amino acid substitution potentially affects protein function. We have constructed a tool that uses sequence homology to predict whether a substitution affects protein function. SIFT, which sorts intolerant from tolerant substitutions, classifies substitutions as tolerated or deleterious. A higher proportion of substitutions predicted to be deleterious by SIFT gives an affected phenotype than substitutions predicted to be deleterious by substitution scoring matrices in three test cases. Using SIFT before mutagenesis studies could reduce the number of functional assays required and yield a higher proportion of affected phenotypes. SIFT may be used to identify plausible disease candidates among the SNPs that cause missense substitutions.

SIFT Key Idea: Substituting one amino acid for another with another with very similar biochemical properties is probably less significant that a more dissimilar substitution. Learn those similarities by comparing orthologs across species

Genome Research. 2001 May;11(5):863-74.



Resource

A probabilistic disease-gene finder for personal genomes

Mark Yandell,[1,3,4] Chad Huff,[1,3] Hao Hu,[1,3] Marc Singleton,[1] Barry Moore,[1] Jinchuan Xing,[1] Lynn B. Jorde,[1] and Martin G. Reese[2]

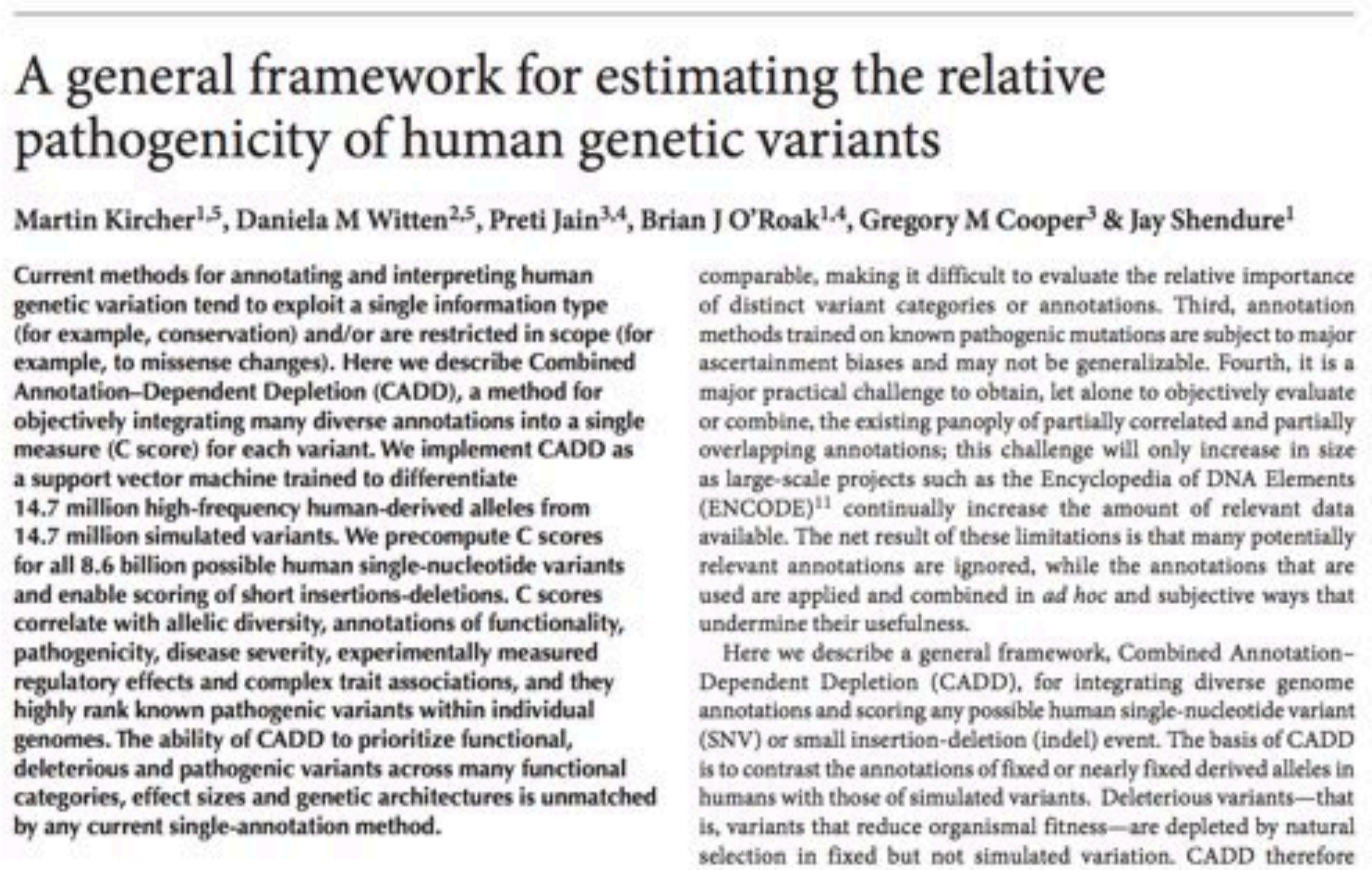
[1]Department of Human Genetics, Eccles Institute of Human Genetics, University of Utah and School of Medicine, Salt Lake City, Utah 84112, USA; [2]Omicia, Inc., Emeryville, California 94608, USA

VAAST (the Variant Annotation, Analysis & Search Tool) is a probabilistic search tool for identifying damaged genes and their disease-causing variants in personal genome sequences. VAAST builds on existing amino acid substitution (AAS) and aggregative approaches to variant prioritization, combining elements of both into a single unified likelihood framework that allows users to identify damaged genes and deleterious variants with greater accuracy, and in an easy-to-use fashion. VAAST can score both coding and noncoding variants, evaluating the cumulative impact of both types of variants simultaneously. VAAST can identify rare variants causing rare genetic diseases, and it can also use both rare and common variants to identify genes responsible for common diseases. VAAST thus has a much greater scope of use than any existing methodology. Here we demonstrate its ability to identify damaged genes using small cohorts ($n = 3$) of unrelated individuals, wherein no two share the same deleterious variants, and for common, multigenic diseases using as few as 150 cases.

[Supplemental material is available for this article.]

VAAST Key Idea: Evaluate amino acid substitutions in evolution AND allele frequencies in 1000 genomes project

Genome Research 2011. doi:10.1101/gr.123158.111



A general framework for estimating the relative pathogenicity of human genetic variants

Martin Kircher[1,5], Daniela M Witten[2,5], Preti Jain[3,4], Brian J O'Roak[1,4], Gregory M Cooper[3] & Jay Shendure[1]

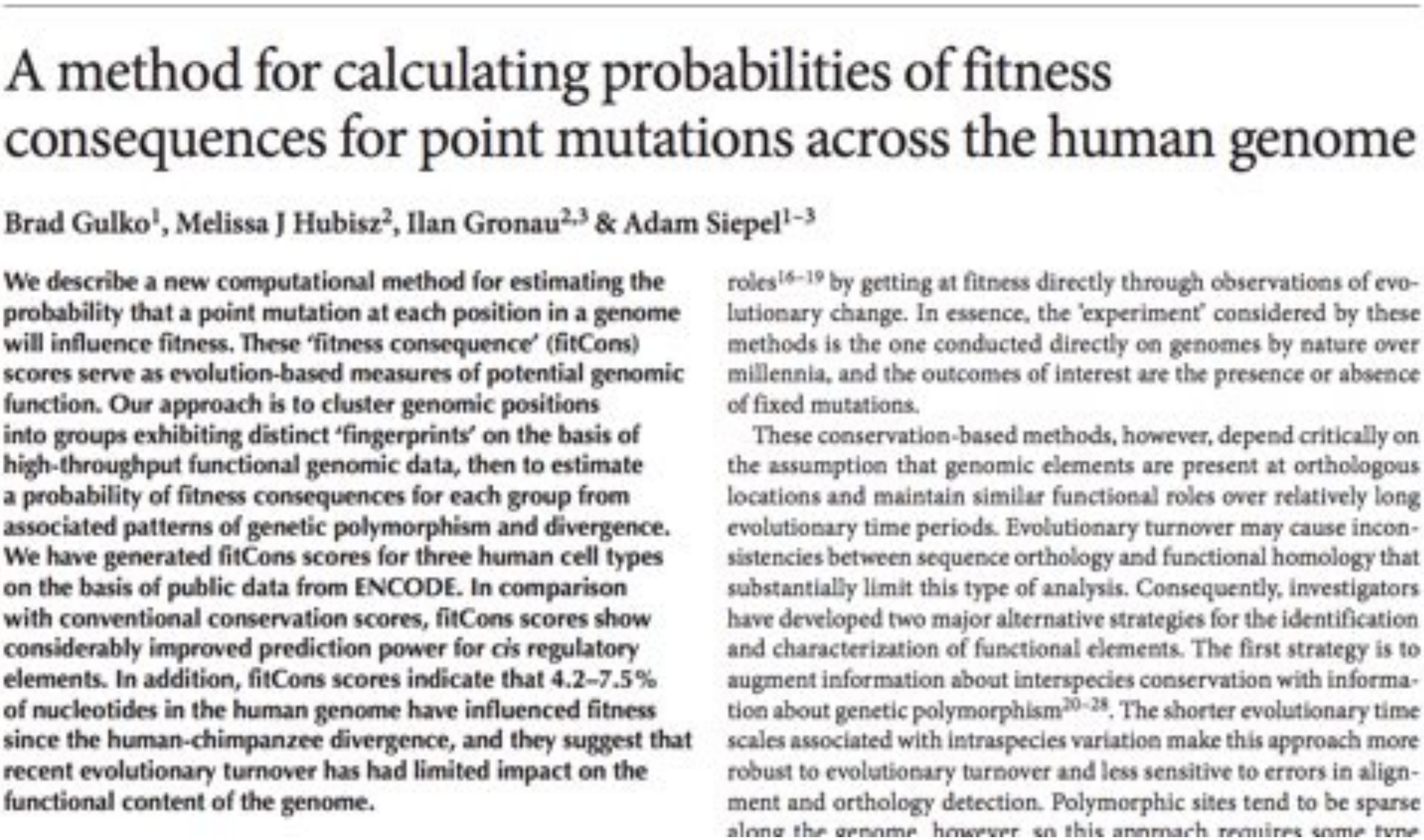
Current methods for annotating and interpreting human genetic variation tend to exploit a single information type (for example, conservation) and/or are restricted in scope (for example, to missense changes). Here we describe Combined Annotation–Dependent Depletion (CADD), a method for objectively integrating many diverse annotations into a single measure (C score) for each variant. We implement CADD as a support vector machine trained to differentiate 14.7 million high-frequency human-derived alleles from 14.7 million simulated variants. We precompute C scores for all 8.6 billion possible human single-nucleotide variants and enable scoring of short insertions-deletions. C scores correlate with allelic diversity, annotations of functionality, pathogenicity, disease severity, experimentally measured regulatory effects and complex trait associations, and they highly rank known pathogenic variants within individual genomes. The ability of CADD to prioritize functional, deleterious and pathogenic variants across many functional categories, effect sizes and genetic architectures is unmatched by any current single-annotation method.

comparable, making it difficult to evaluate the relative importance of distinct variant categories or annotations. Third, annotation methods trained on known pathogenic mutations are subject to major ascertainment biases and may not be generalizable. Fourth, it is a major practical challenge to obtain, let alone to objectively evaluate or combine, the existing panoply of partially correlated and partially overlapping annotations; this challenge will only increase in size as large-scale projects such as the Encyclopedia of DNA Elements (ENCODE)[11] continually increase the amount of relevant data available. The net result of these limitations is that many potentially relevant annotations are ignored, while the annotations that are used are applied and combined in *ad hoc* and subjective ways that undermine their usefulness.

Here we describe a general framework, Combined Annotation–Dependent Depletion (CADD), for integrating diverse genome annotations and scoring any possible human single-nucleotide variant (SNV) or small insertion-deletion (indel) event. The basis of CADD is to contrast the annotations of fixed or nearly fixed derived alleles in humans with those of simulated variants. Deleterious variants—that is, variants that reduce organismal fitness—are depleted by natural selection in fixed but not simulated variation. CADD therefore

CADD Key Idea: Evaluate amino acid substitutions AND allele frequencies in 1000 genomes project AND ENCODE regions AND … (63 annotations total :)

Nature Genetics 46, 310–315 (2014) doi:10.1038/ng.2892



A method for calculating probabilities of fitness consequences for point mutations across the human genome

Brad Gulko[1], Melissa J Hubisz[2], Ilan Gronau[2,3] & Adam Siepel[1–3]

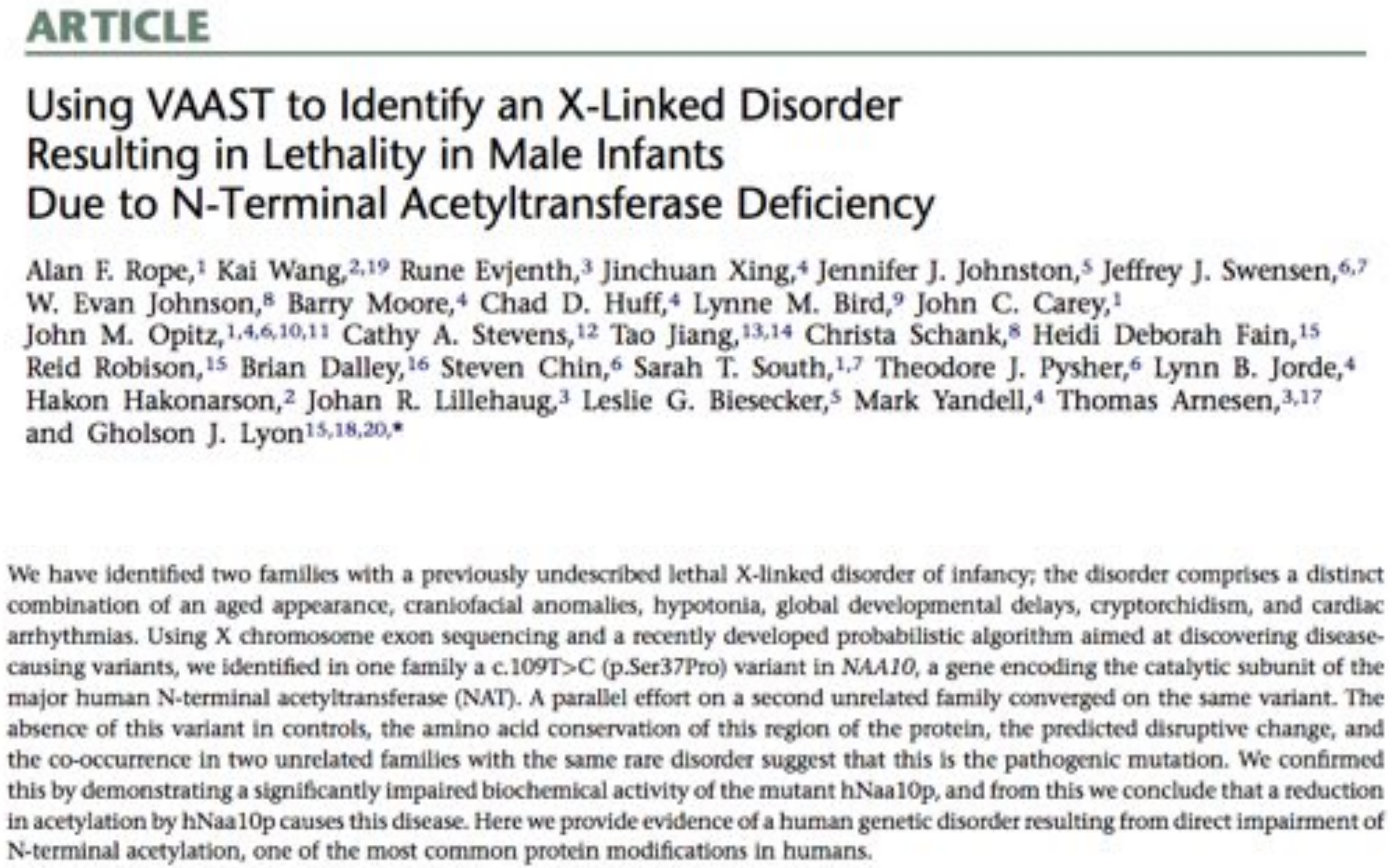
We describe a new computational method for estimating the probability that a point mutation at each position in a genome will influence fitness. These 'fitness consequence' (fitCons) scores serve as evolution-based measures of potential genomic function. Our approach is to cluster genomic positions into groups exhibiting distinct 'fingerprints' on the basis of high-throughput functional genomic data, then to estimate a probability of fitness consequences for each group from associated patterns of genetic polymorphism and divergence. We have generated fitCons scores for three human cell types on the basis of public data from ENCODE. In comparison with conventional conservation scores, fitCons scores show considerably improved prediction power for *cis* regulatory elements. In addition, fitCons scores indicate that 4.2–7.5% of nucleotides in the human genome have influenced fitness since the human-chimpanzee divergence, and they suggest that recent evolutionary turnover has had limited impact on the functional content of the genome.

roles[16–19] by getting at fitness directly through observations of evolutionary change. In essence, the 'experiment' considered by these methods is the one conducted directly on genomes by nature over millennia, and the outcomes of interest are the presence or absence of fixed mutations.

These conservation-based methods, however, depend critically on the assumption that genomic elements are present at orthologous locations and maintain similar functional roles over relatively long evolutionary time periods. Evolutionary turnover may cause inconsistencies between sequence orthology and functional homology that substantially limit this type of analysis. Consequently, investigators have developed two major alternative strategies for the identification and characterization of functional elements. The first strategy is to augment information about interspecies conservation with information about genetic polymorphism[20–28]. The shorter evolutionary time scales associated with intraspecies variation make this approach more robust to evolutionary turnover and less sensitive to errors in alignment and orthology detection. Polymorphic sites tend to be sparse along the genome, however, so this approach requires some type

fitCons Key Idea: Evaluate amino acid substitutions AND allele frequencies in 1000 genomes project AND aggregate by ENCODE regions

Nature Genetics 47, 276–283 (2015) doi:10.1038/ng.3196



ARTICLE

Using VAAST to Identify an X-Linked Disorder Resulting in Lethality in Male Infants Due to N-Terminal Acetyltransferase Deficiency

Alan F. Rope,[1] Kai Wang,[2,19] Rune Evjenth,[3] Jinchuan Xing,[4] Jennifer J. Johnston,[5] Jeffrey J. Swensen,[6,7] W. Evan Johnson,[8] Barry Moore,[4] Chad D. Huff,[4] Lynne M. Bird,[9] John C. Carey,[1] John M. Opitz,[1,4,6,10,11] Cathy A. Stevens,[12] Tao Jiang,[13,14] Christa Schank,[8] Heidi Deborah Fain,[15] Reid Robison,[15] Brian Dalley,[16] Steven Chin,[6] Sarah T. South,[1,7] Theodore J. Pysher,[6] Lynn B. Jorde,[4] Hakon Hakonarson,[2] Johan R. Lillehaug,[3] Leslie G. Biesecker,[5] Mark Yandell,[4] Thomas Arnesen,[3,17] and Gholson J. Lyon[15,18,20,*]

We have identified two families with a previously undescribed lethal X-linked disorder of infancy; the disorder comprises a distinct combination of an aged appearance, craniofacial anomalies, hypotonia, global developmental delays, cryptorchidism, and cardiac arrhythmias. Using X chromosome exon sequencing and a recently developed probabilistic algorithm aimed at discovering disease-causing variants, we identified in one family a c.109T>C (p.Ser37Pro) variant in *NAA10*, a gene encoding the catalytic subunit of the major human N-terminal acetyltransferase (NAT). A parallel effort on a second unrelated family converged on the same variant. The absence of this variant in controls, the amino acid conservation of this region of the protein, the predicted disruptive change, and the co-occurrence in two unrelated families with the same rare disorder suggest that this is the pathogenic mutation. We confirmed this by demonstrating a significantly impaired biochemical activity of the mutant hNaa10p, and from this we conclude that a reduction in acetylation by hNaa10p causes this disease. Here we provide evidence of a human genetic disorder resulting from direct impairment of N-terminal acetylation, one of the most common protein modifications in humans.

Am J Hum Genet. 2011 Aug 12;89(2):345. doi: 10.1016/j.ajhg.2011.05.017

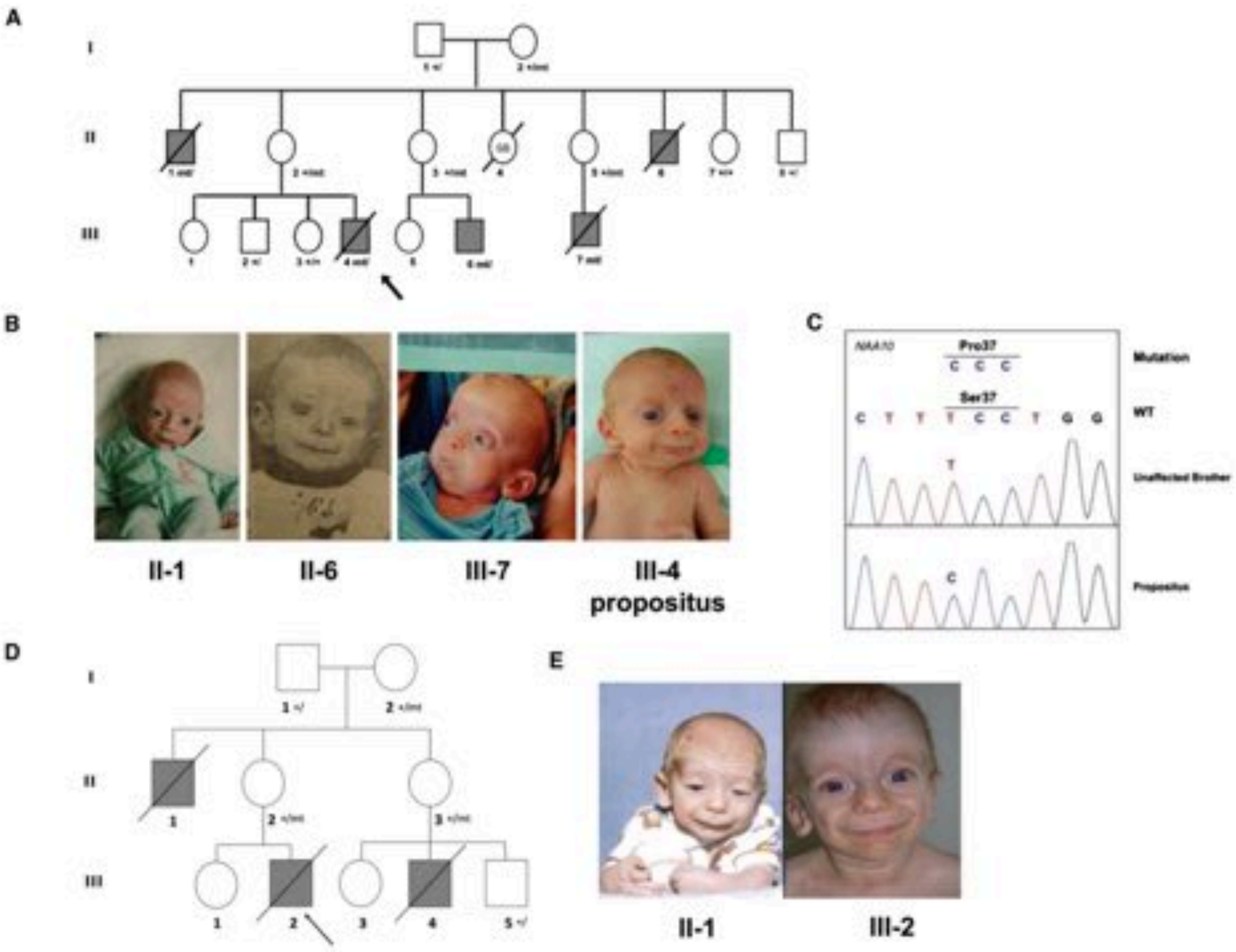


Figure 2. Pedigree Drawing and Pictures of Families 1 and 2

(A) Pedigree drawing for family 1. The most recent deceased individual, III-4, is the most well-studied subject in the family and is indicated by an arrow. Genotypes are marked for those in which DNA was available and tested. The following abbreviations are used: SB, stillborn; +, normal variant; mt, rare mutant variant.

(B) Pictures of four affected and deceased boys in this family, showing the aged appearance.

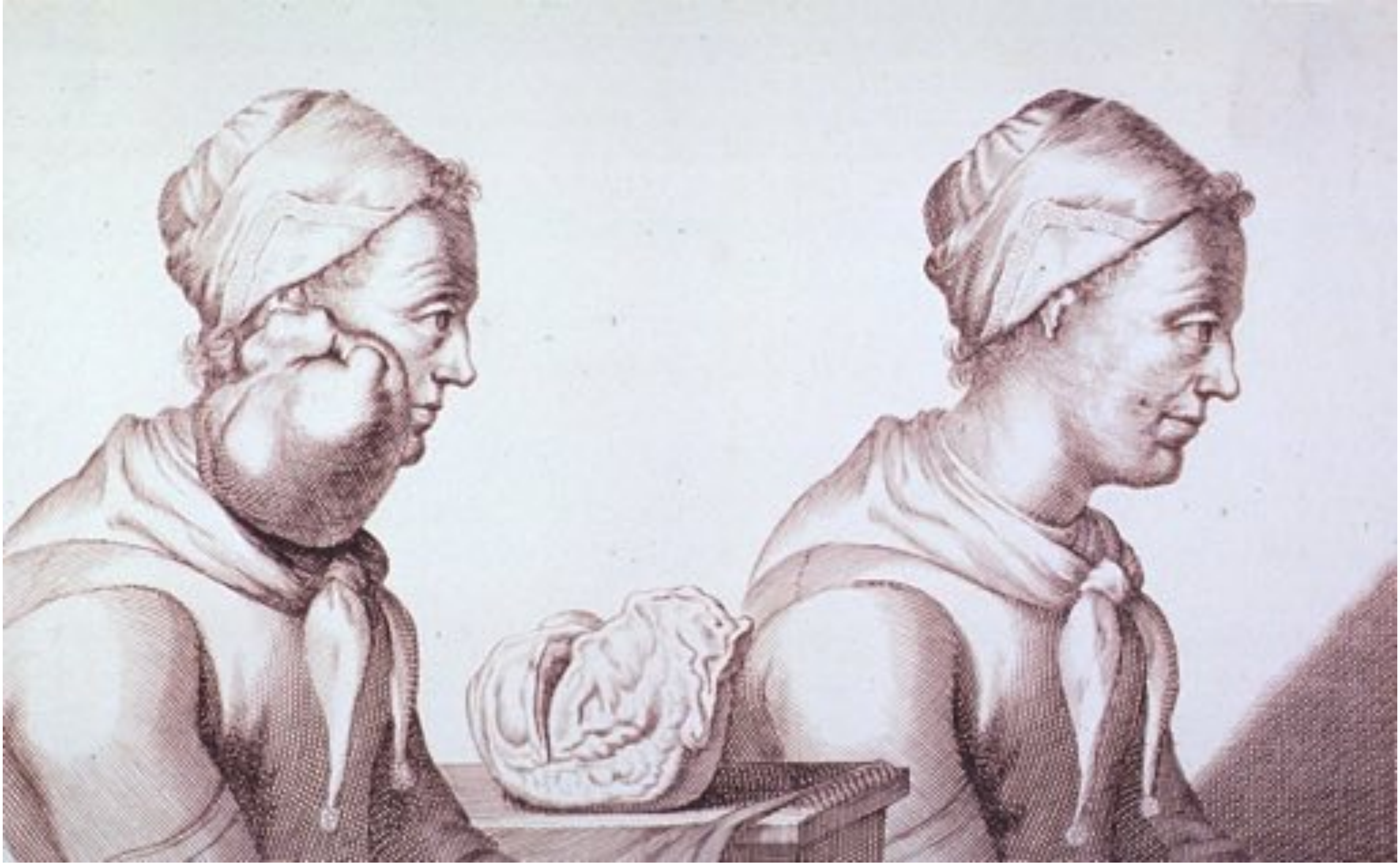
(C) Sanger sequencing results of *NAA10* in individual III-4 from family 1.

(D) Pedigree for family 2. Individual III-2 is the most well-studied subject in the family and is indicated by an arrow.

(E) Picture of individuals II-1 and III-2 in family 2 at ~1 year of age.

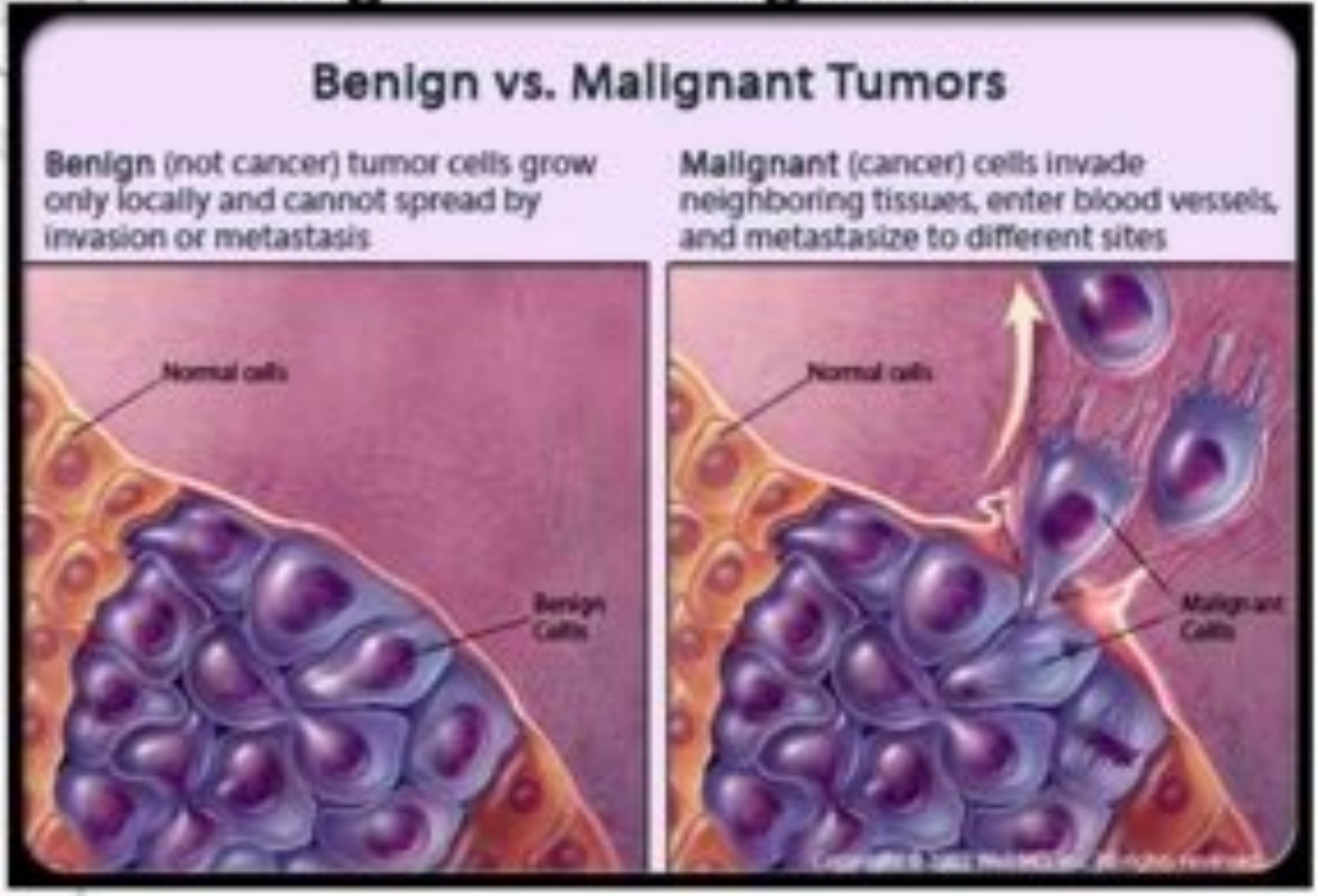
Part 3:

Cancer Genetics & Genomics



A tumor removed by surgery in 1689.

Benign vs. Malignant



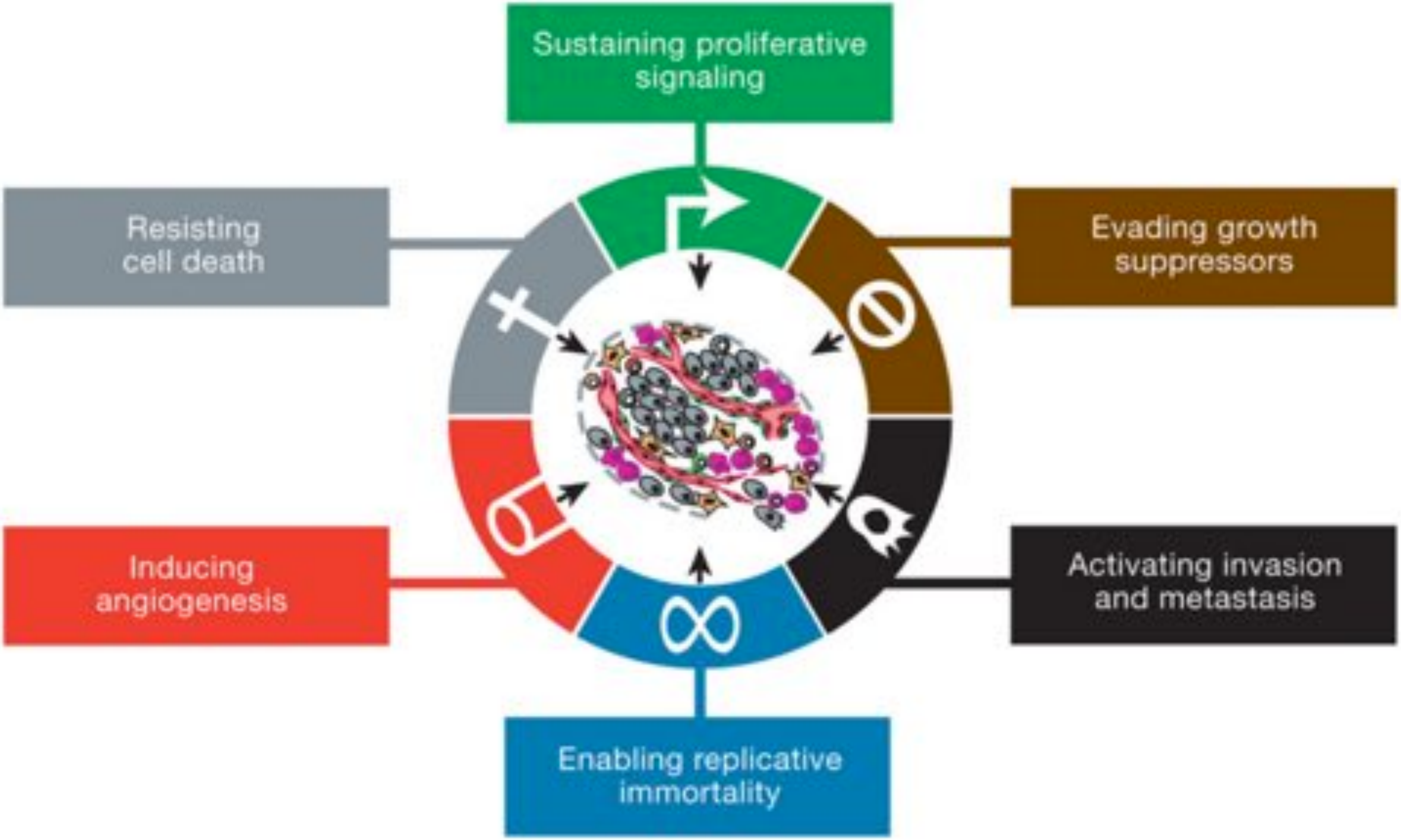
Benign vs. Malignant Tumors

Benign (not cancer) tumor cells grow only locally and cannot spread by invasion or metastasis

Malignant (cancer) cells invade neighboring tissues, enter blood vessels, and metastasize to different sites

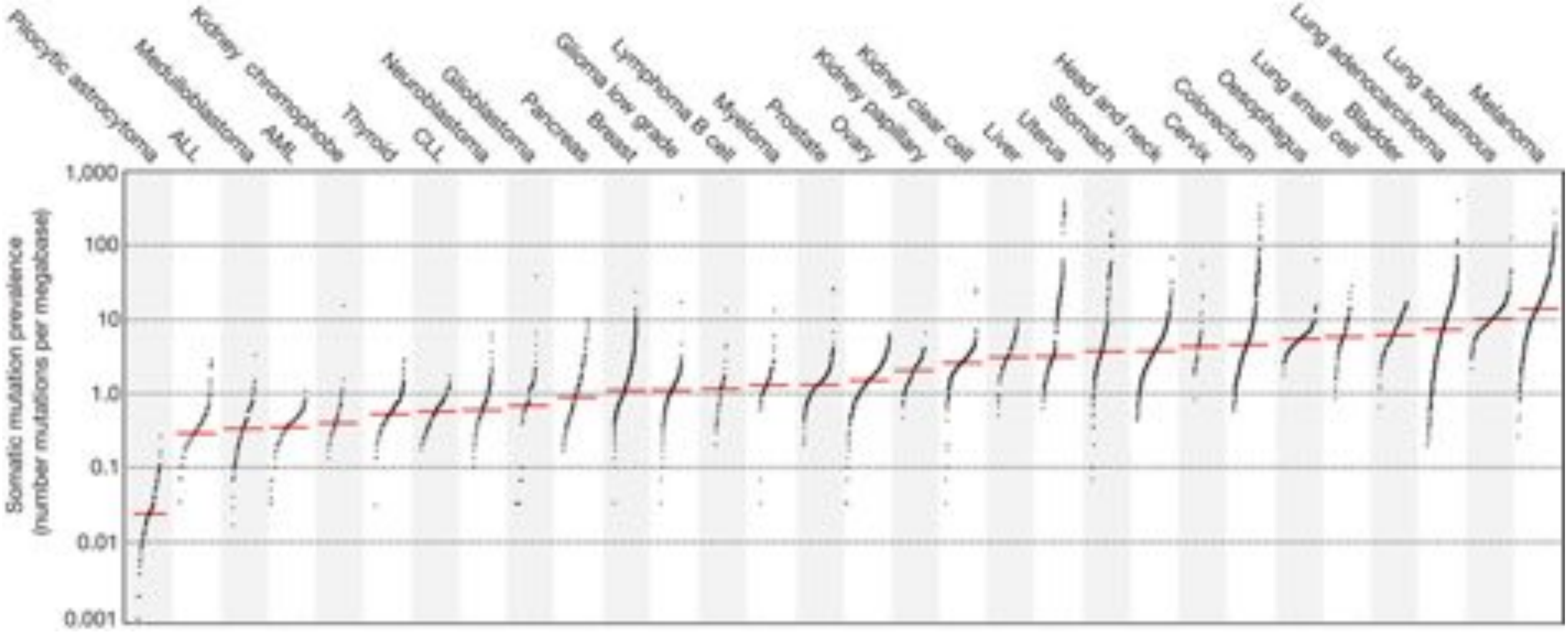
Copyright © 2002 WebMD Inc. All rights reserved.

The Six Hallmarks of Cancer


Sustaining proliferative signaling
Evading growth suppressors
Activating invasion and metastasis
Enabling replicative immortality
Inducing angiogenesis
Resisting cell death

Hallmarks of Cancer
Hanahan and Weinberg (2000) Cell. http://doi.org/10.1016/S0092-8674(00)81683-9

Somatic Mutations In Cancer

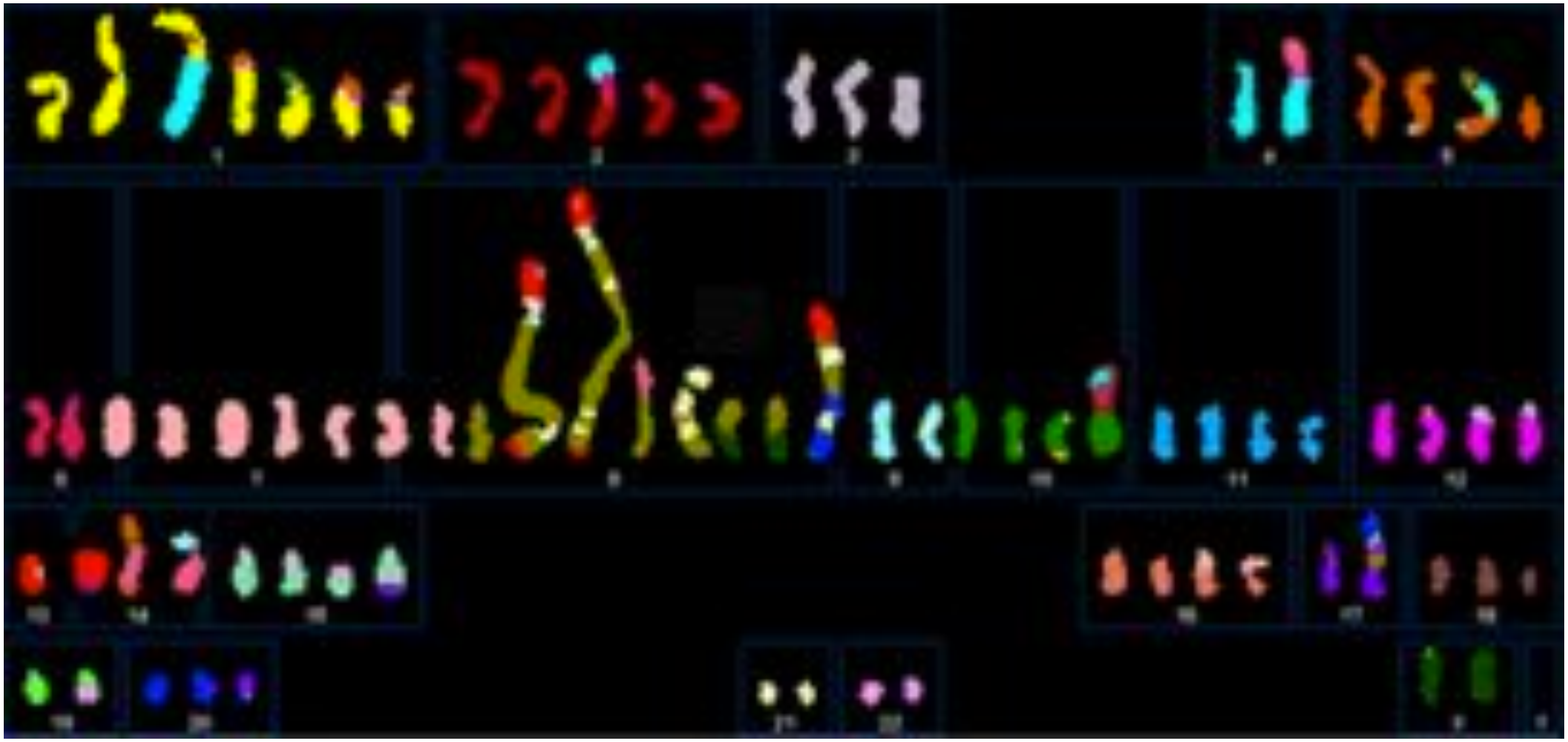


Signatures of mutational processes in human cancer

Alexandrov et al (2013) *Nature*. doi:10.1038/nature12477

SK-BR-3

Most commonly used Her2-amplified breast cancer cell line



(Davidson et al, 2000)

80+ chromosomes,
Many are a patchwork of fragments of other chromosomes

A firestorm in cancer

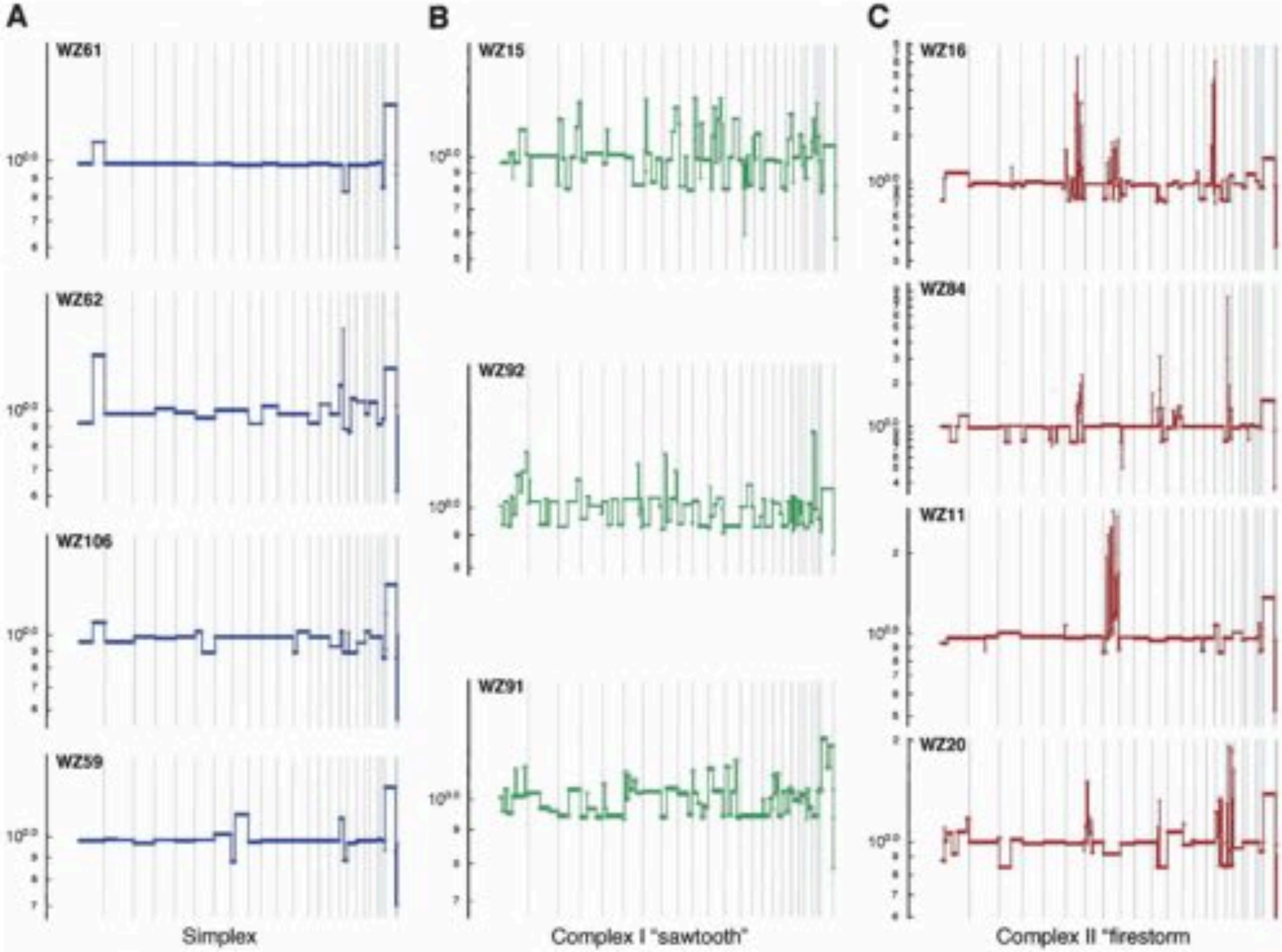
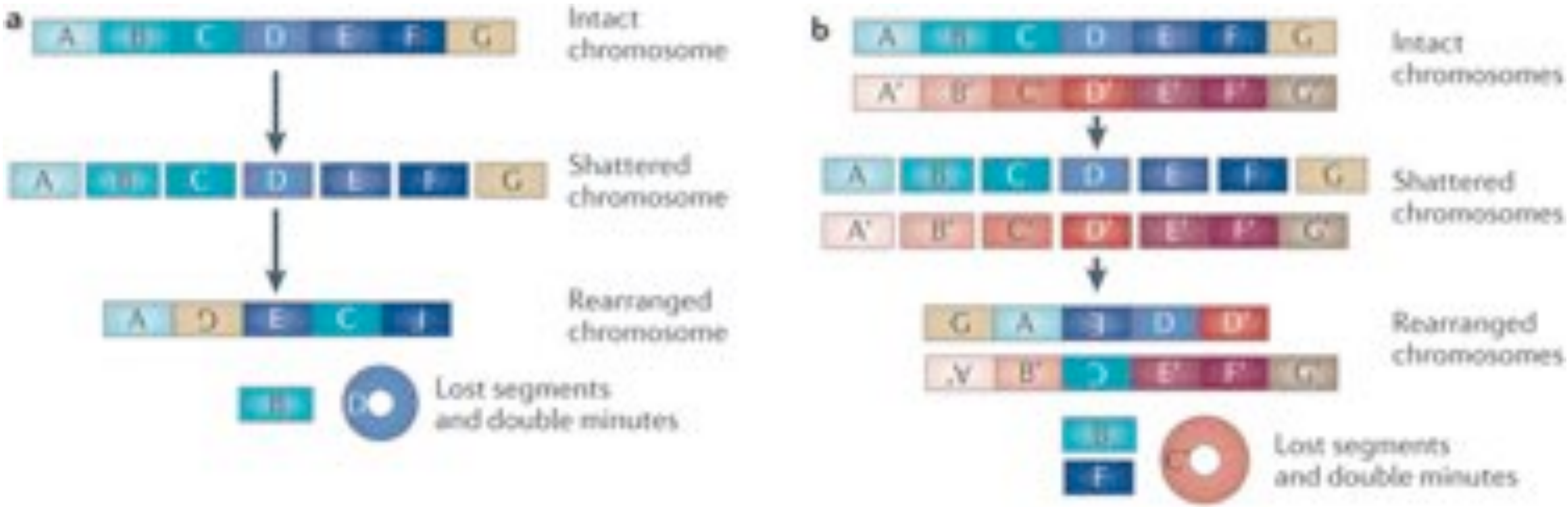


Figure 2. Major types of tumor genomic profiles. Segmentation profiles for individual tumors representing each category: (*A*) simplex; (*B*) complex type I or sawtooth; (*C*) complex type II or firestorm. Scored events consist of a minimum of six consecutive probes in the same state. The *y*-axis displays the geometric mean value of two experiments on a log scale. Note that the scale of the amplifications in *C* is compressed relative to *A* and *B* owing to the high levels of amplification in firestorms. Chromosomes 1–22 plus X and Y are displayed in order from *left* to *right* according to probe position.

Novel patterns of genome rearrangement and their association with survival in breast cancer
Hicks et al (2006) *Genome Research. Doi: 10.1101/gr.5460106*

Aberrations in cancer genomes



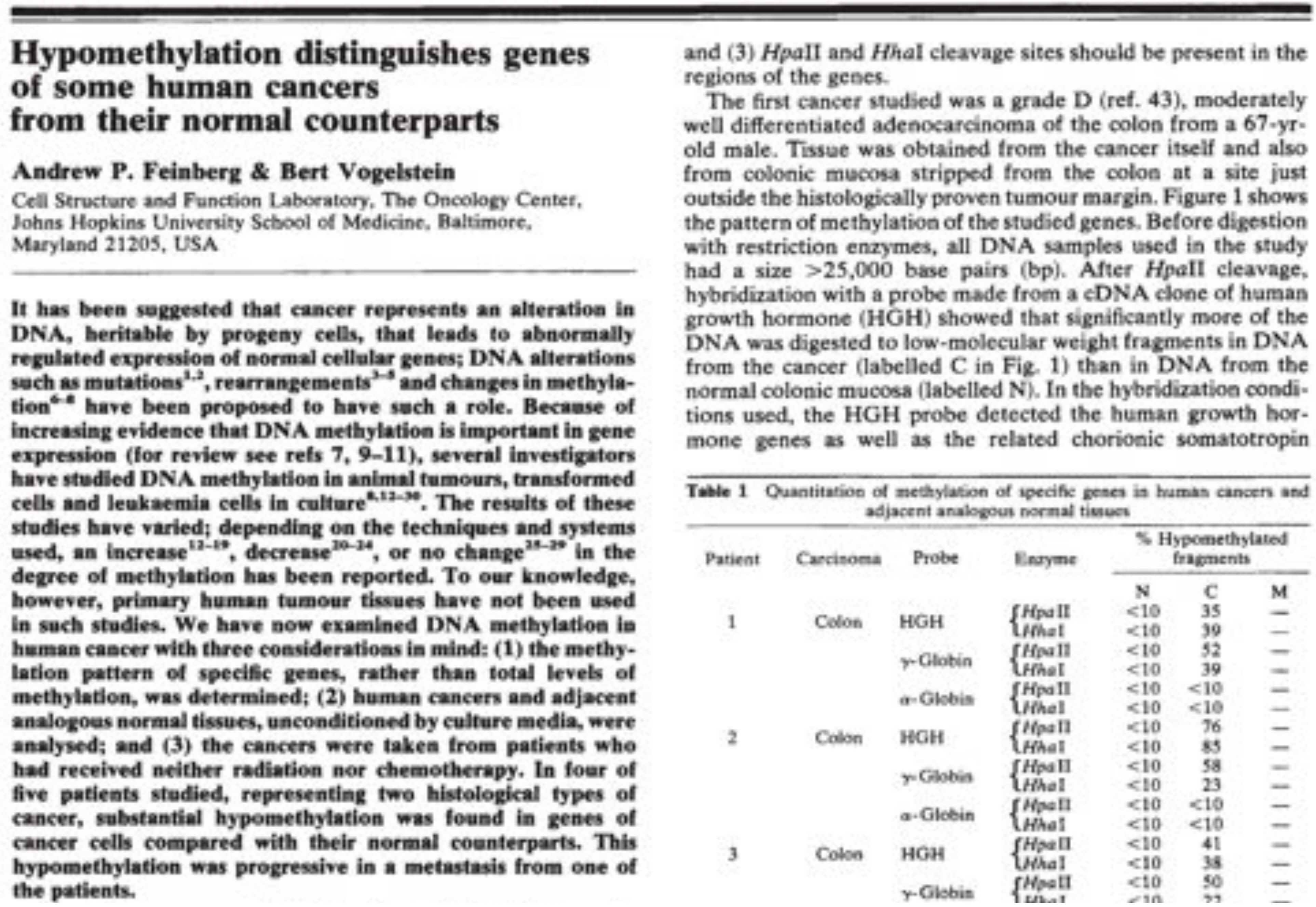
Chromothripsis, which literally means 'chromosome shattering', is a phenomenon that has recently been reported to occur in cells harbouring complex genomic rearrangements (CGRs). Has 3 defining characteristics:

(1) Occurrence of remarkable numbers of rearrangements in localized chromosomal regions;

(2) Low number of copy number states (generally between one or two) across the rearranged region;

(3) Alternation in the chromothriptic areas of regions where heterozygosity is preserved with regions presenting loss of heterozygosity (LOH).

Chromothripsis and cancer: causes and consequences of chromosome shattering
Forment et al (2012) Nature Reviews Cancer. doi:10.1038/nrc3352



Hypomethylation distinguishes genes of some human cancers from their normal counterparts

Andrew P. Feinberg & Bert Vogelstein

Cell Structure and Function Laboratory, The Oncology Center,
Johns Hopkins University School of Medicine, Baltimore,
Maryland 21205, USA

It has been suggested that cancer represents an alteration in DNA, heritable by progeny cells, that leads to abnormally regulated expression of normal cellular genes; DNA alterations such as mutations[1,2], rearrangements[3–5] and changes in methylation[6–8] have been proposed to have such a role. Because of increasing evidence that DNA methylation is important in gene expression (for review see refs 7, 9–11), several investigators have studied DNA methylation in animal tumours, transformed cells and leukaemia cells in culture[8,12–30]. The results of these studies have varied; depending on the techniques and systems used, an increase[12–19], decrease[20–24], or no change[25–29] in the degree of methylation has been reported. To our knowledge, however, primary human tumour tissues have not been used in such studies. We have now examined DNA methylation in human cancer with three considerations in mind: (1) the methylation pattern of specific genes, rather than total levels of methylation, was determined; (2) human cancers and adjacent analogous normal tissues, unconditioned by culture media, were analysed; and (3) the cancers were taken from patients who had received neither radiation nor chemotherapy. In four of five patients studied, representing two histological types of cancer, substantial hypomethylation was found in genes of cancer cells compared with their normal counterparts. This hypomethylation was progressive in a metastasis from one of the patients.

and (3) *Hpa*II and *Hha*I cleavage sites should be present in the regions of the genes.

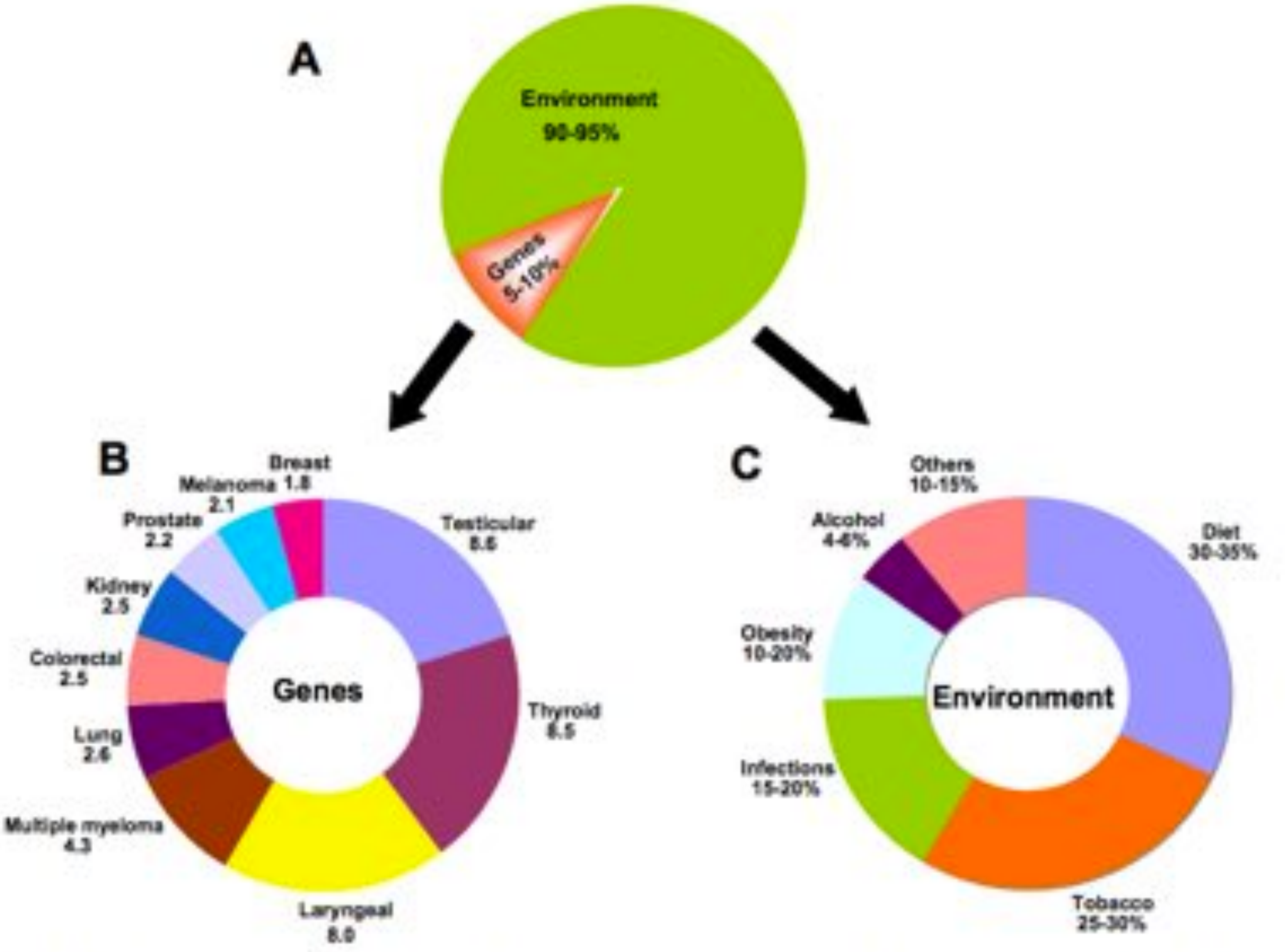
The first cancer studied was a grade D (ref. 43), moderately well differentiated adenocarcinoma of the colon from a 67-yr-old male. Tissue was obtained from the cancer itself and also from colonic mucosa stripped from the colon at a site just outside the histologically proven tumour margin. Figure 1 shows the pattern of methylation of the studied genes. Before digestion with restriction enzymes, all DNA samples used in the study had a size >25,000 base pairs (bp). After *Hpa*II cleavage, hybridization with a probe made from a cDNA clone of human growth hormone (HGH) showed that significantly more of the DNA was digested to low-molecular weight fragments in DNA from the cancer (labelled C in Fig. 1) than in DNA from the normal colonic mucosa (labelled N). In the hybridization conditions used, the HGH probe detected the human growth hormone genes as well as the related chorionic somatotropin

Table 1 Quantitation of methylation of specific genes in human cancers and adjacent analogous normal tissues

Patient	Carcinoma	Probe	Enzyme	% Hypomethylated fragments		
				N	C	M
1	Colon	HGH	*Hpa*II	<10	35	—
			*Hha*I	<10	39	—
		γ-Globin	*Hpa*II	<10	52	—
			*Hha*I	<10	39	—
		α-Globin	*Hpa*II	<10	<10	—
			*Hha*I	<10	<10	—
2	Colon	HGH	*Hpa*II	<10	76	—
			*Hha*I	<10	85	—
		γ-Globin	*Hpa*II	<10	58	—
			*Hha*I	<10	23	—
		α-Globin	*Hpa*II	<10	<10	—
			*Hha*I	<10	<10	—
3	Colon	HGH	*Hpa*II	<10	41	—
			*Hha*I	<10	38	—
		γ-Globin	*Hpa*II	<10	50	—
			*Hha*I	<10	[illegible]	—

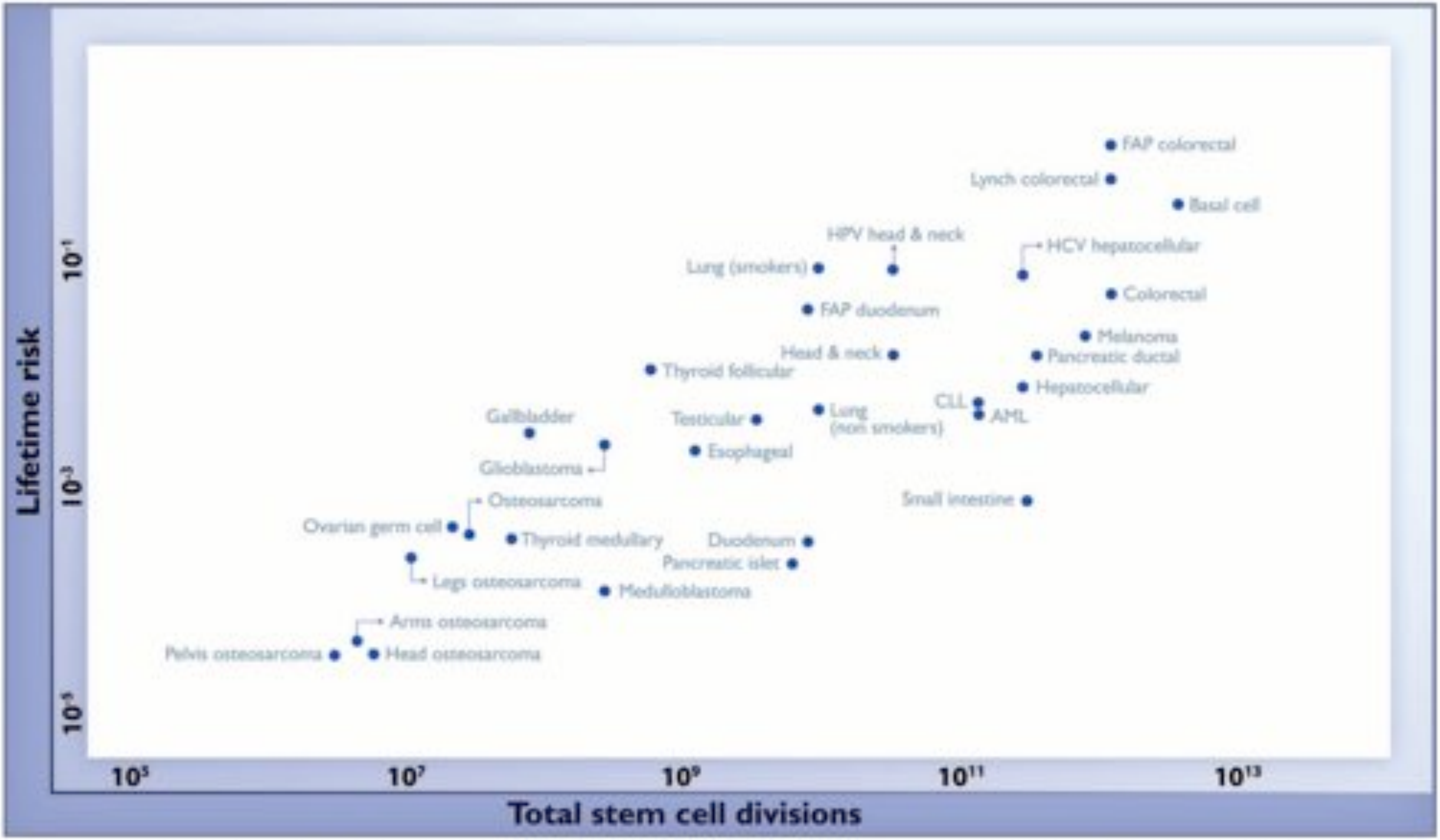
Nature 301, 89 - 92 (06 January 1983); doi:10.1038/301089a0

Causes of Cancer



Cancer is a Preventable Disease that Requires Major Lifestyle Changes

Anand et al (2008) Pharmaceutical Research. doi: 10.1007/s11095-008-9661-9

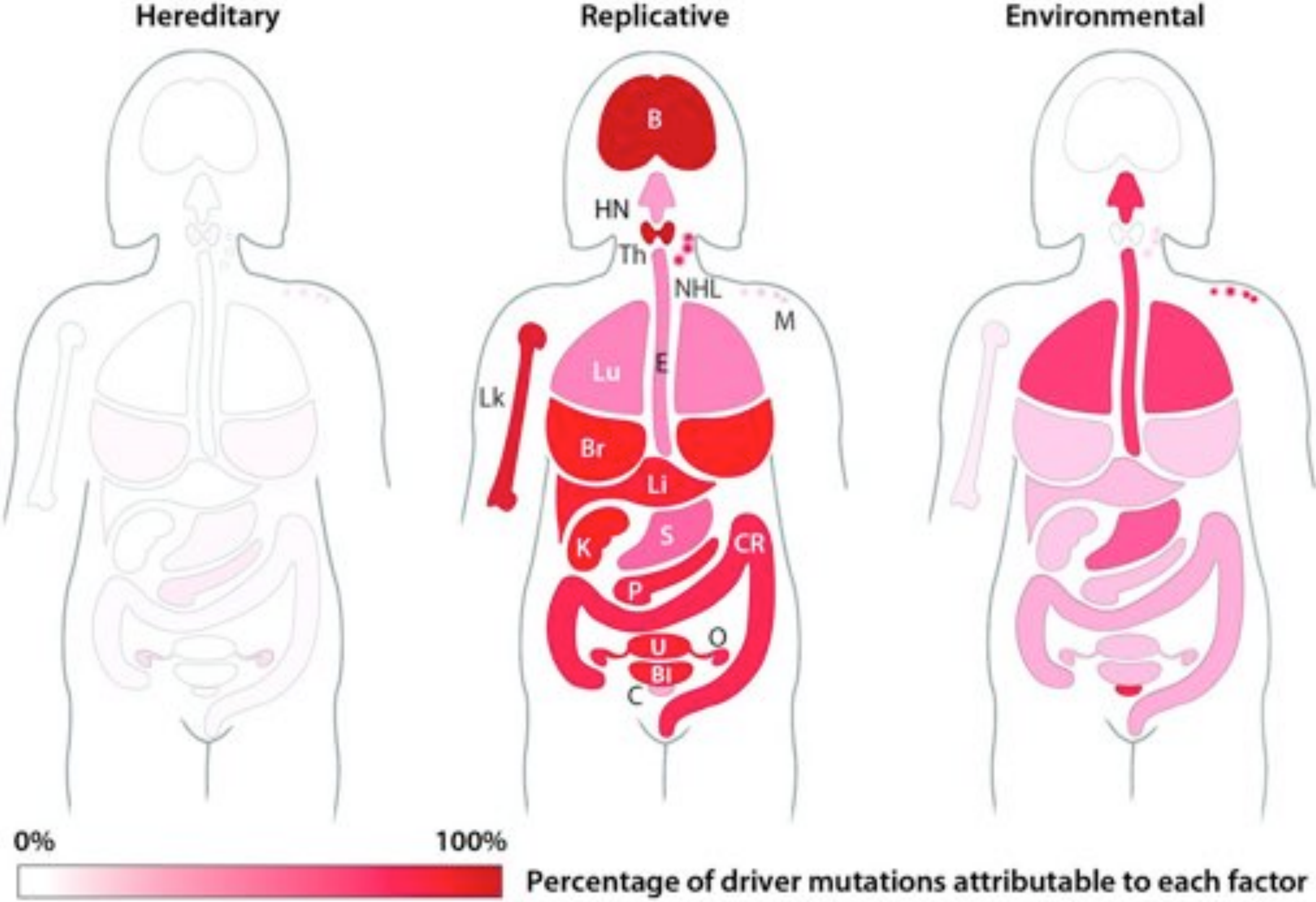


FAP = Familial Adenomatous Polyposis ◆ HCV = Hepatitis C virus ◆ HPV = Human papillomavirus ◆ CLL = Chronic lymphocytic leukemia ◆ AML = Acute myeloid leukemia

Fig. 1. The relationship between the number of stem cell divisions in the lifetime of a given tissue and the lifetime risk of cancer in that tissue. Values are from table S1, the derivation of which is discussed in the supplementary materials.

Variation in cancer risk among tissues can be explained by the number of stem cell divisions

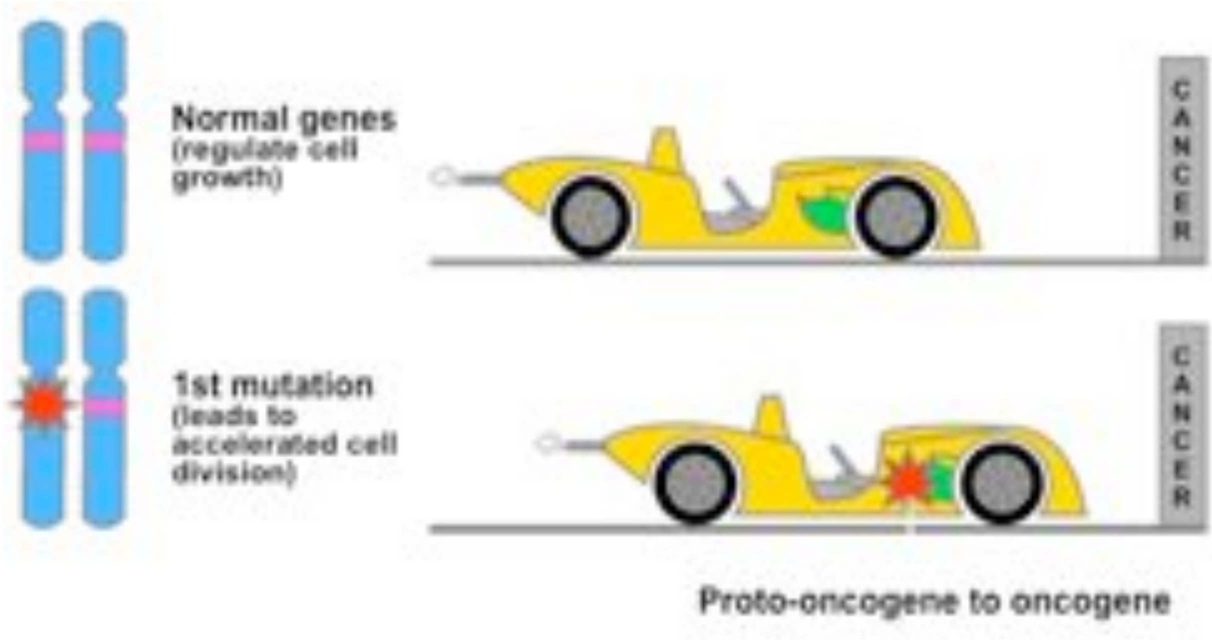
Tomasetti and Vogelstein (2015) Science. DOI: 10.1126/science.1260825



Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention

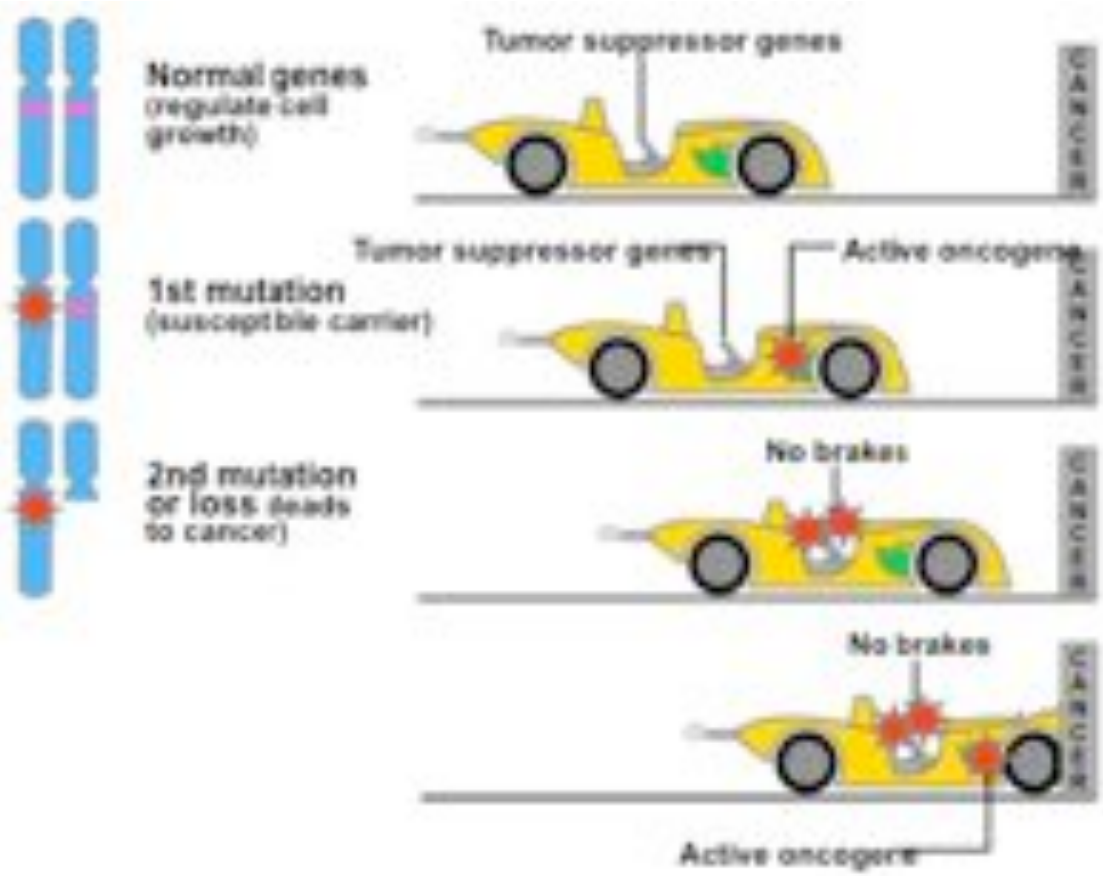
Tomasetti, Li, and Vogelstein (2017) Science. DOI: 10.1126/science.aaf9011

Oncogenes


Normal genes
(regulate cell growth)
CANCER
1st mutation
(leads to accelerated cell division)
CANCER
Proto-oncogene to oncogene

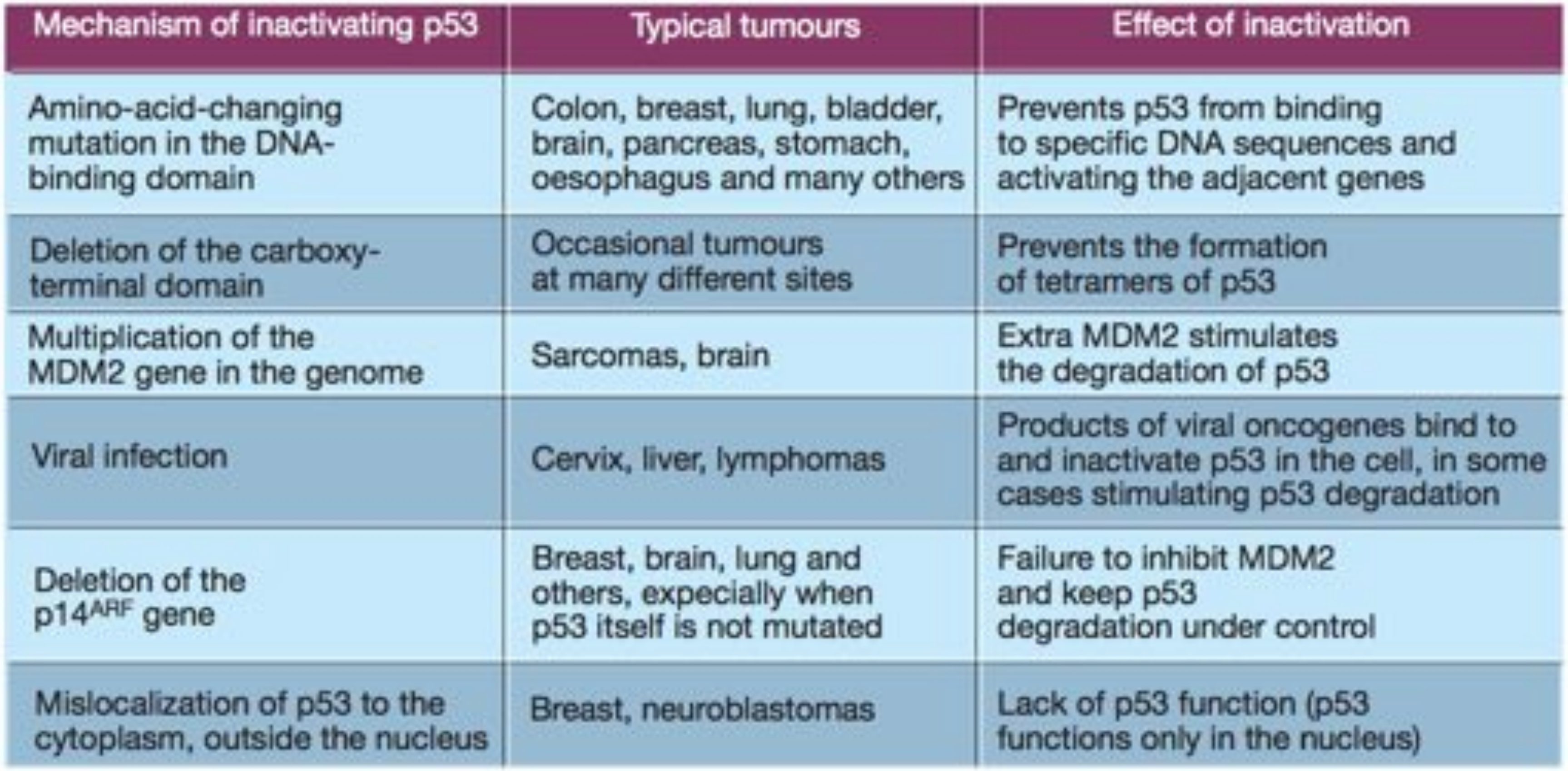
- ***HER-2/neuHER-2/neu:*** encodes for a cell surface receptor that can stimulate cell division. The HER-2/neu gene is amplified in up to 30% of human breast cancers.
- ***RAS:*** The Ras gene products are involved in kinase signaling pathways that ultimately control transcription of genes, regulating cell growth and differentiation.
- ***MYC:*** The Myc protein is a transcription factor and controls expression of several genes.
- ***SRC:*** First oncogene ever discovered. The Src protein is a tyrosine kinase, which regulates cell activity.
- ***hTER:*** Codes for an enzyme (telomerase) that maintains chromosome ends.

Tumor Suppressors



- ***TP53:*** a transcription factor that regulates cell division and cell death.
- ***Rb:*** alters the activity of transcription factors and therefore controls cell division.
- ***APC:*** controls the availability of a transcription factor.
- **PTEN:** acts by opposing the action of PI3K, which is essential for anti-apoptotic, pro-tumorogenic Akt activation.

TP53: The first and most important tumor suppressor



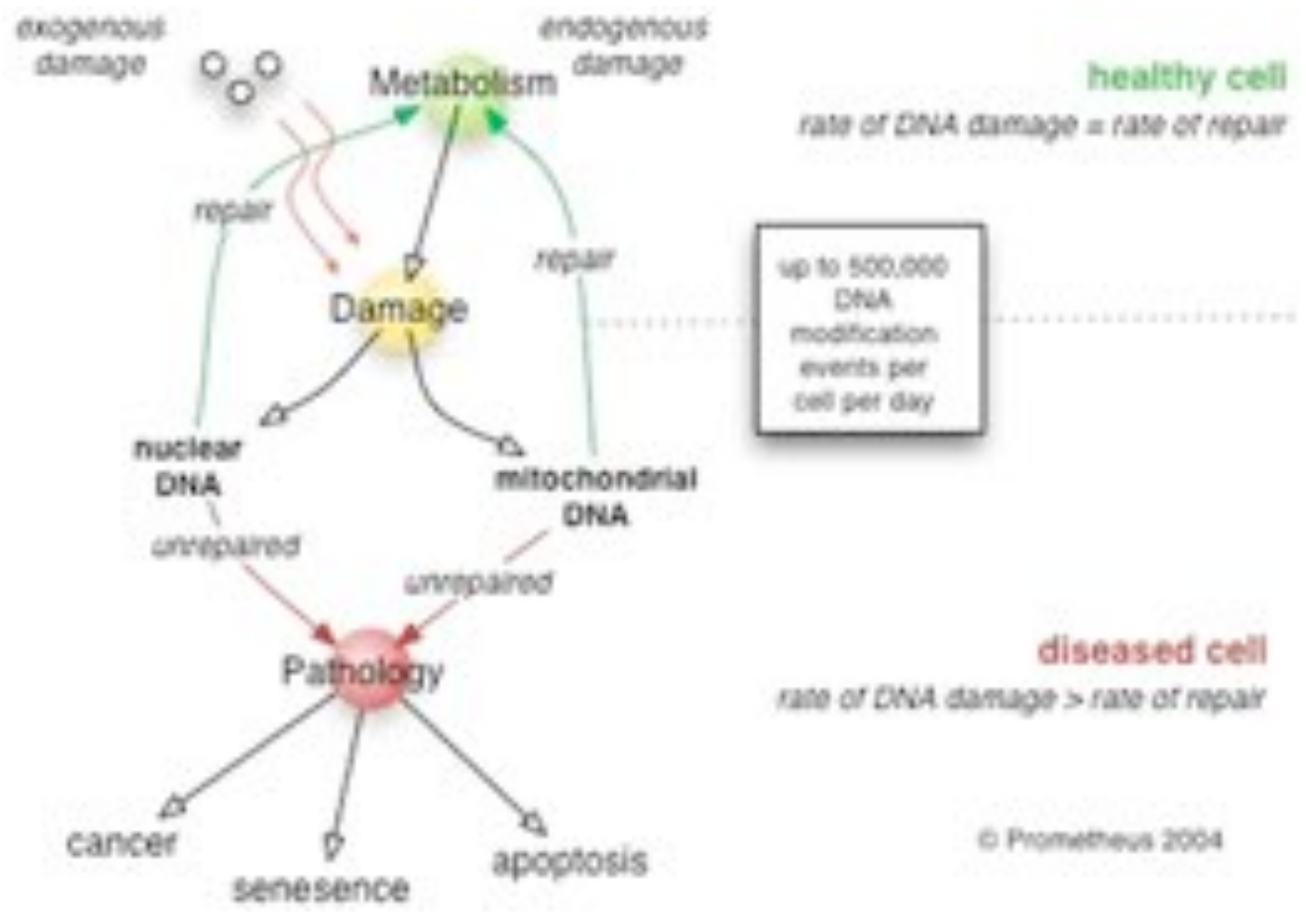
Mechanism of inactivating p53	Typical tumours	Effect of inactivation
Amino-acid-changing mutation in the DNA-binding domain	Colon, breast, lung, bladder, brain, pancreas, stomach, oesophagus and many others	Prevents p53 from binding to specific DNA sequences and activating the adjacent genes
Deletion of the carboxy-terminal domain	Occasional tumours at many different sites	Prevents the formation of tetramers of p53
Multiplication of the MDM2 gene in the genome	Sarcomas, brain	Extra MDM2 stimulates the degradation of p53
Viral infection	Cervix, liver, lymphomas	Products of viral oncogenes bind to and inactivate p53 in the cell, in some cases stimulating p53 degradation
Deletion of the $p14^{ARF}$ gene	Breast, brain, lung and others, expecially when p53 itself is not mutated	Failure to inhibit MDM2 and keep p53 degradation under control
Mislocalization of p53 to the cytoplasm, outside the nucleus	Breast, neuroblastomas	Lack of p53 function (p53 functions only in the nucleus)

Figure 1 **The many ways in which p53 may malfunction in human cancers.**

>10,000 known mutations
>17,000 publications

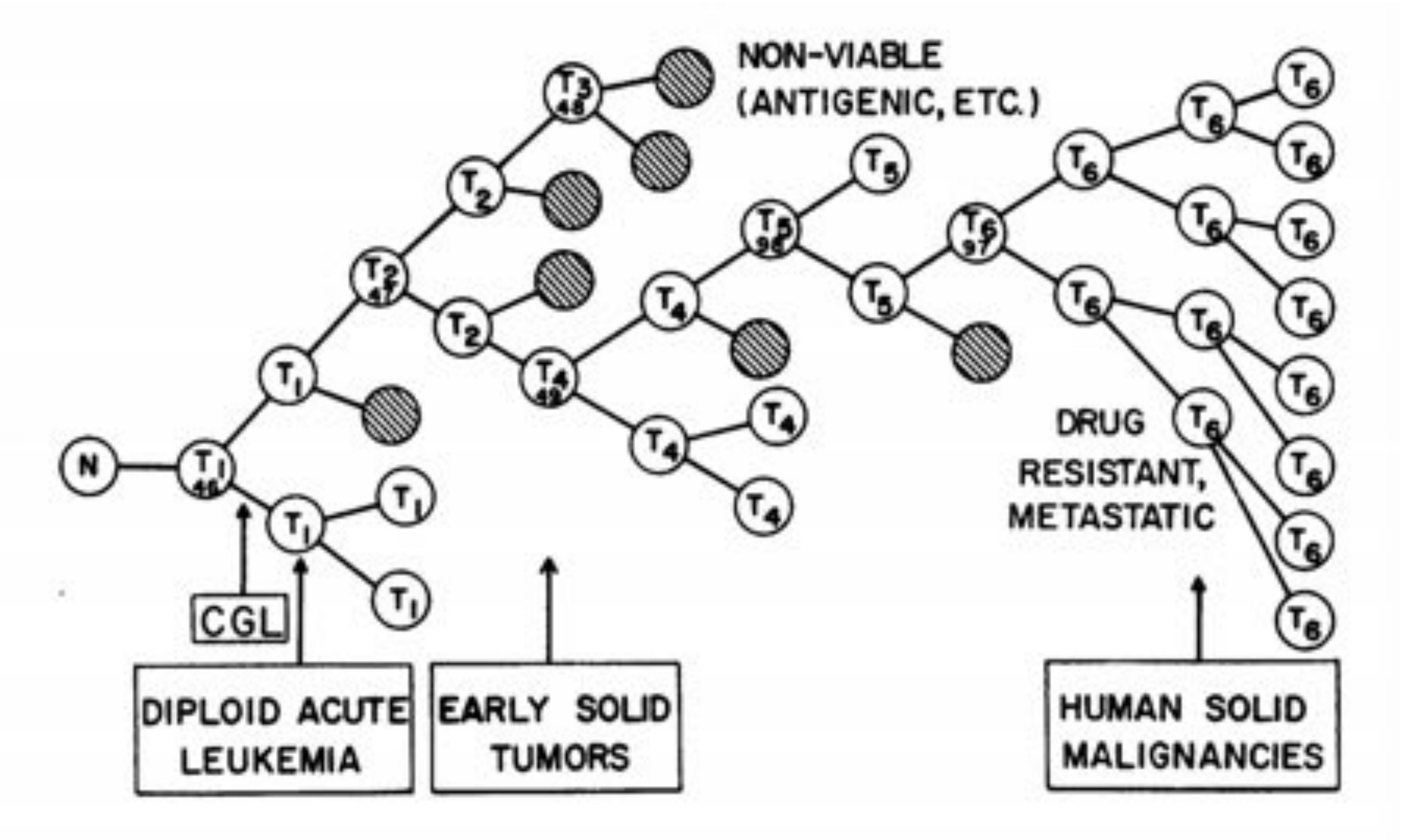
Surfing the p53 network
Volgelstein et al (2000) Nature. DOI: 10.1038/35042675

DNA Repair Genes



BRCA1 and BRCA2 (breast cancer type 1/2 susceptibility genes)
Normally expressed in the cells of breast and other tissue, where they help repair damaged DNA, or destroy cells if DNA cannot be repaired. They are involved in the repair of chromosomal damage with an important role in the error-free repair of DNA double-strand breaks

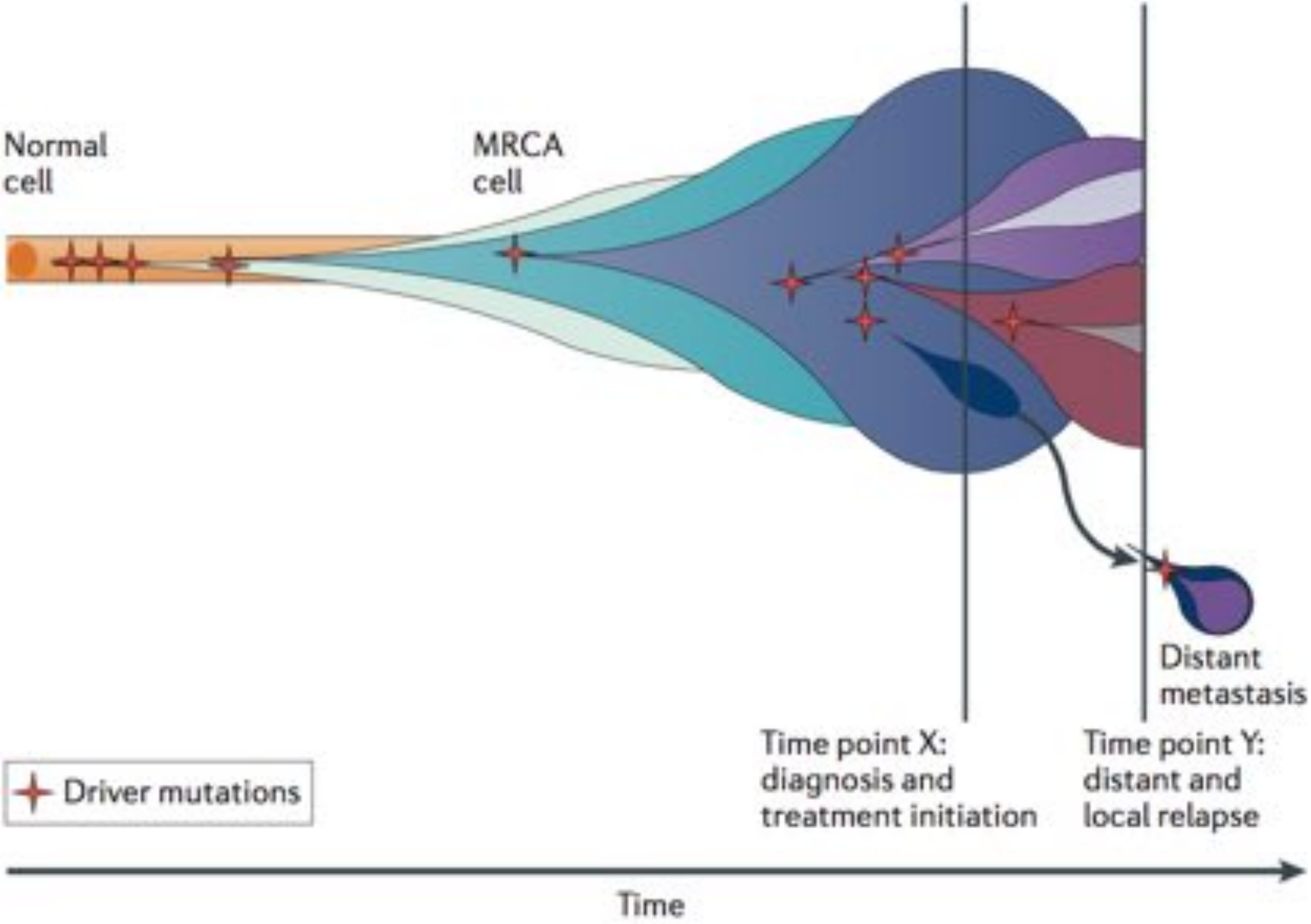
Tumor Evolution



The Clonal Evolution of Tumor Cell Populations

Peter C. Nowell (1976) *Science*. 194(4260):23-28 DOI: 10.1126/science.959840

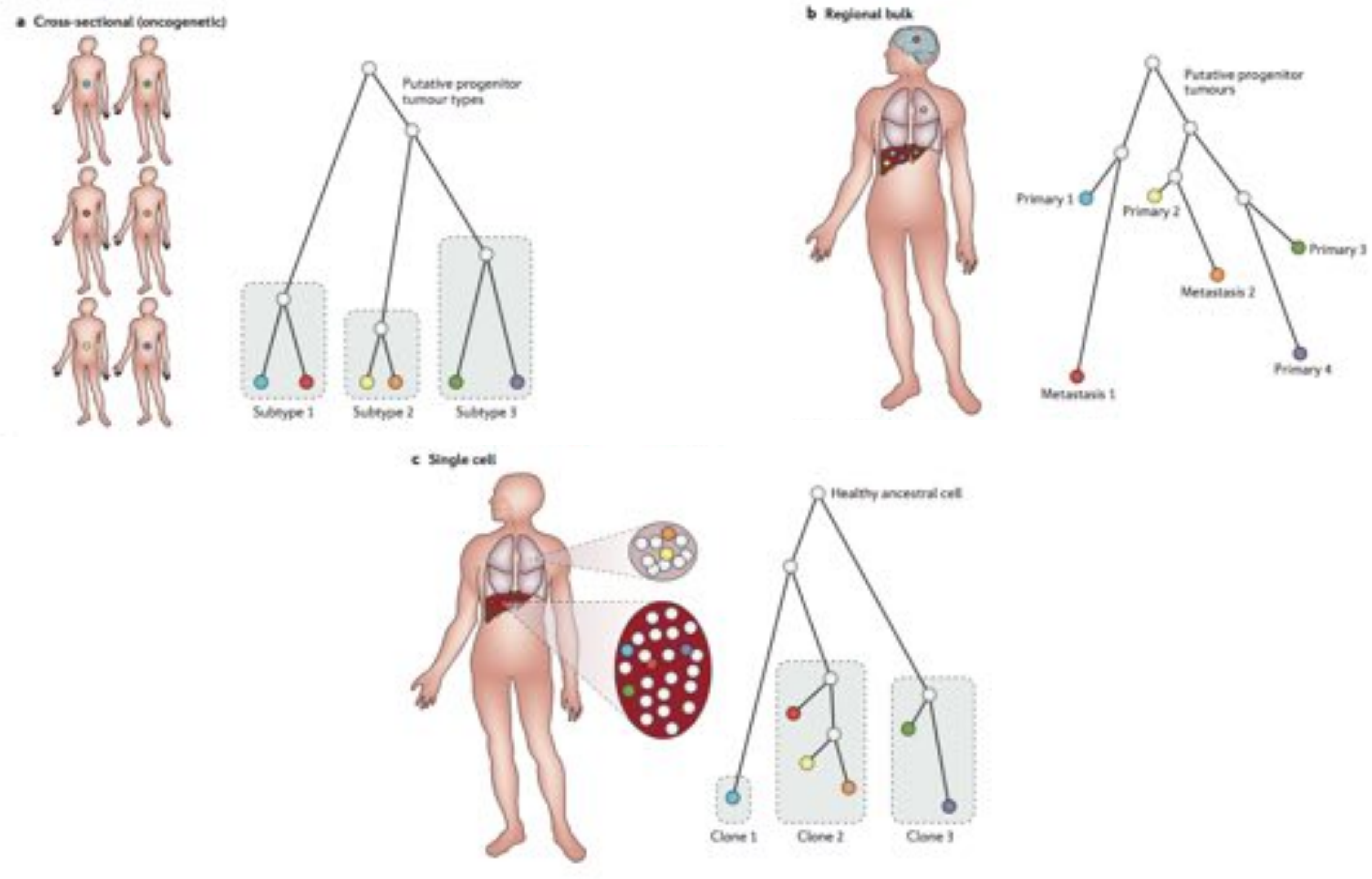
Tumor Evolution


Normal cell
MRCA cell
Distant metastasis
Time point X: diagnosis and treatment initiation
Time point Y: distant and local relapse
Driver mutations
Time

Evolution of the cancer genome

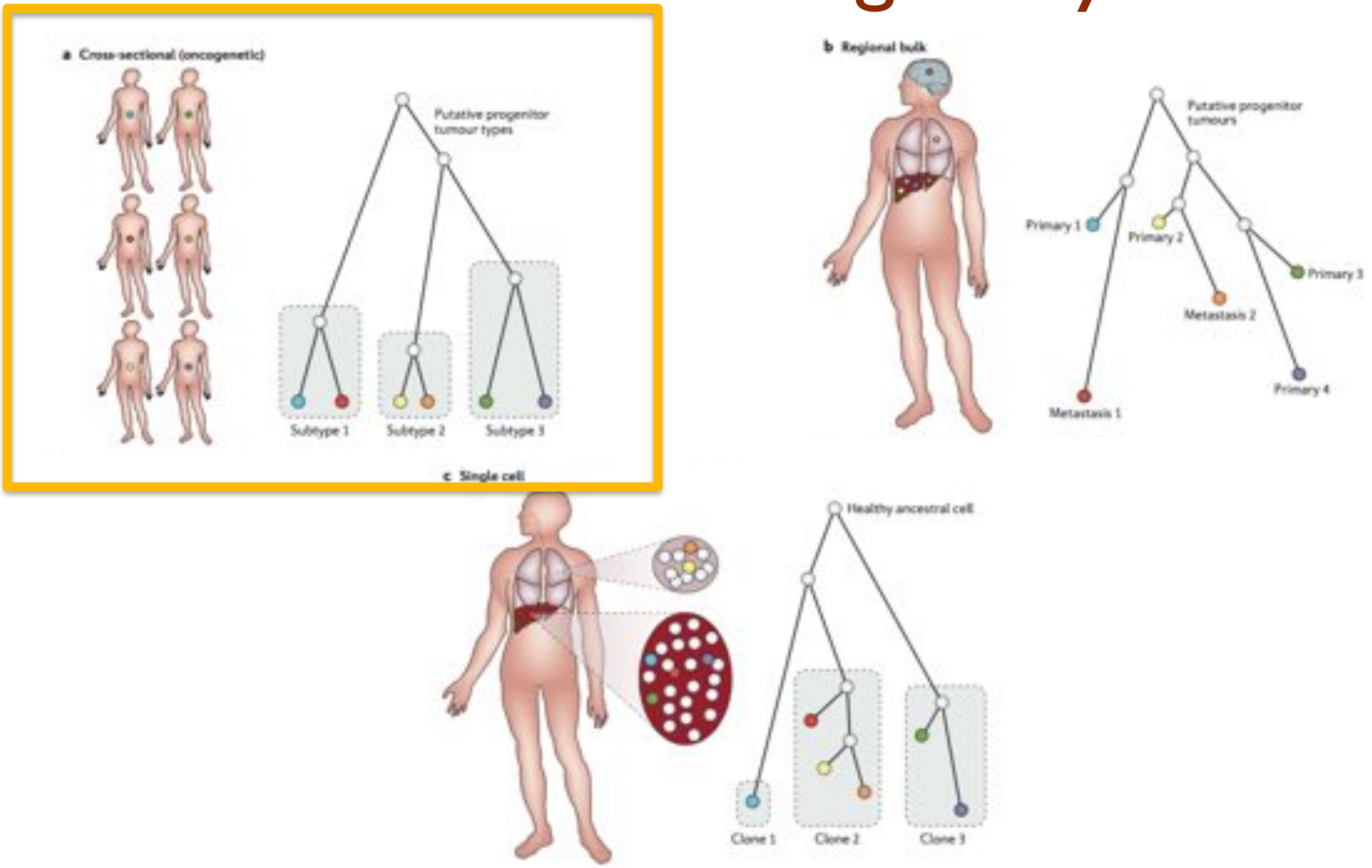
Yates & Campbell (2012) Nature Review Genetics. doi:10.1038/nrg3317

Tumor Heterogeneity



The evolution of tumour phylogenetics: principles and practice
Schwarz and Schaffer (2017) *Nature Reviews Genetics. doi:10.1038/nrg.2016.170*

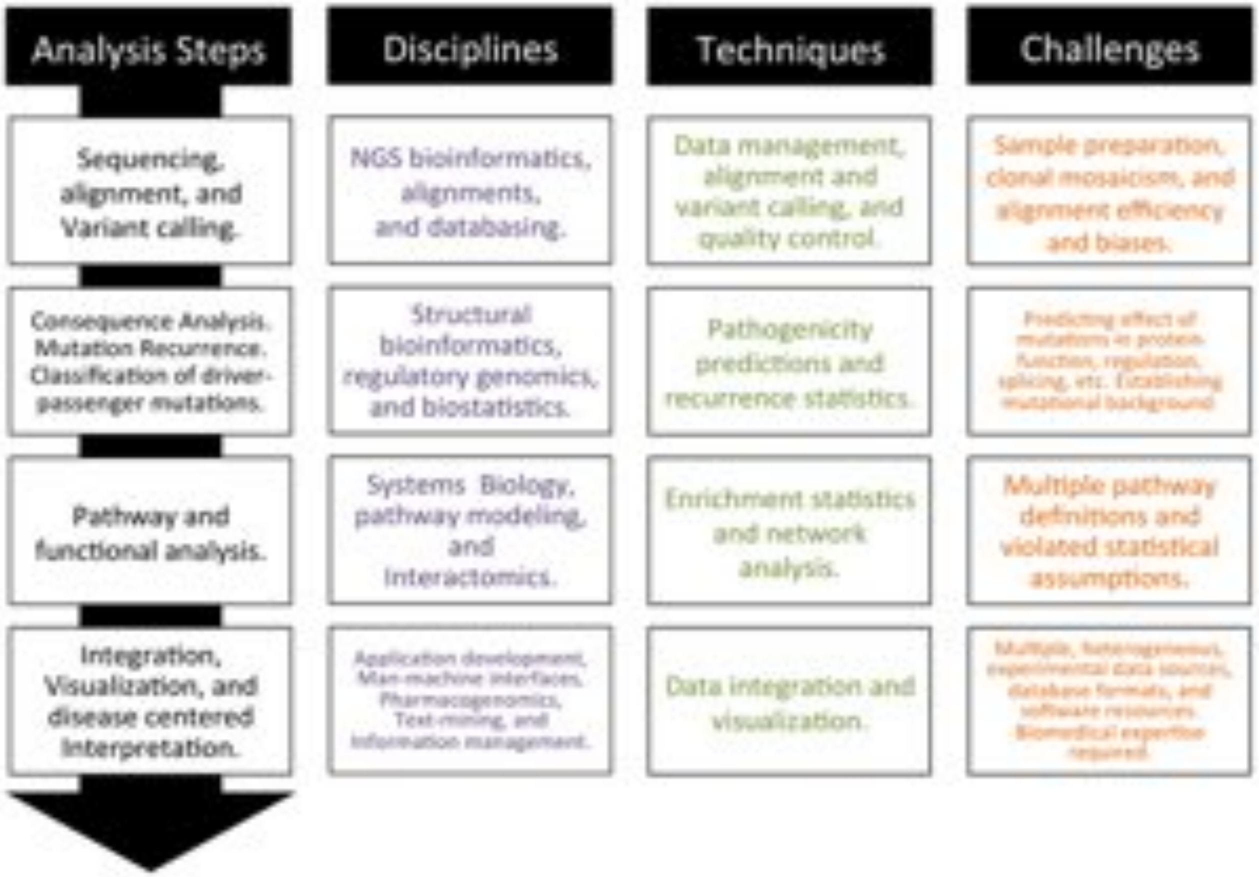
Tumor Heterogeneity



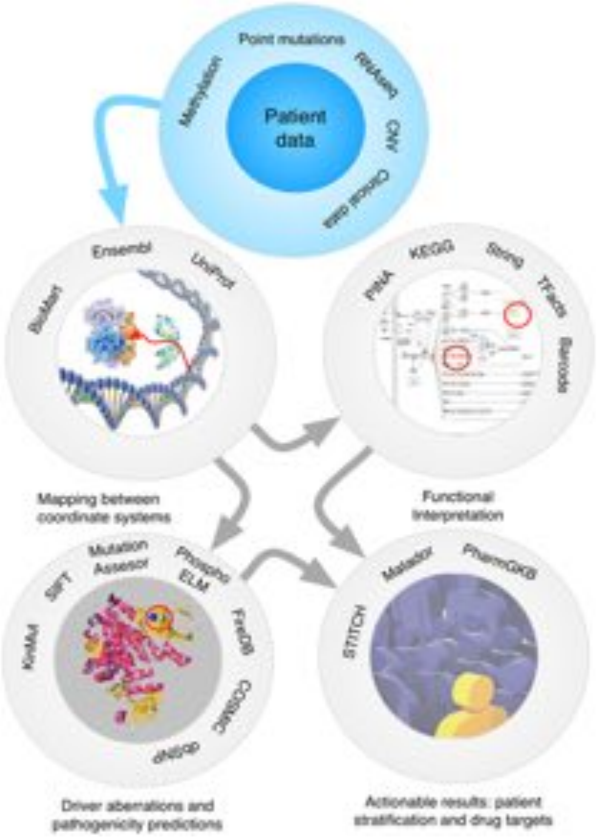
The evolution of tumour phylogenetics: principles and practice

Schwarz and Schaffer (2017) *Nature Reviews Genetics. doi:10.1038/nrg.2016.170*

Cancer Mutation Analysis



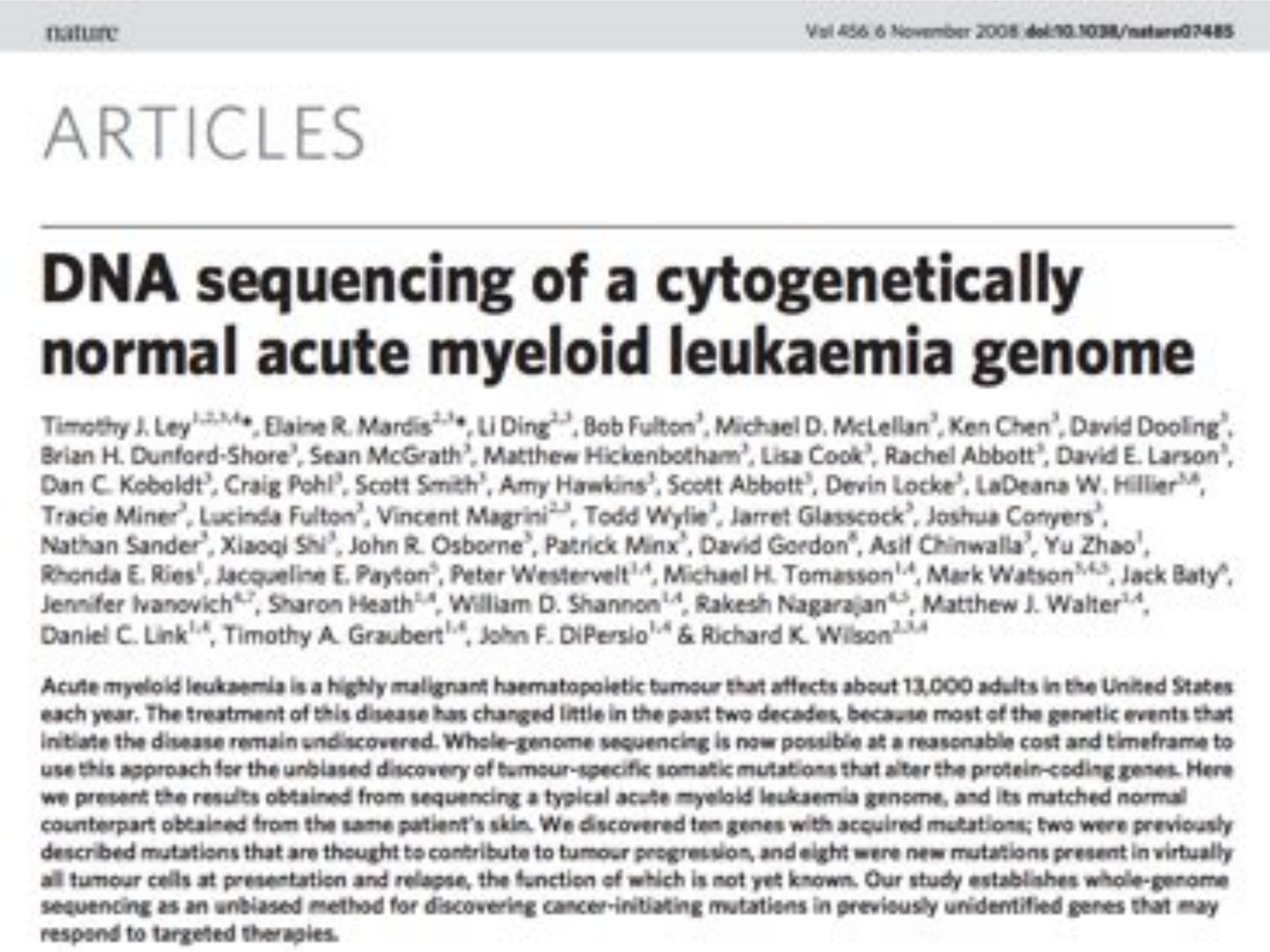
Analysis Steps	Disciplines	Techniques	Challenges
Sequencing, alignment, and Variant calling.	NGS bioinformatics, alignments, and databasing.	Data management, alignment and variant calling, and quality control.	Sample preparation, clonal mosaicism, and alignment efficiency and biases.
Consequence Analysis. Mutation Recurrence. Classification of driver-passenger mutations.	Structural bioinformatics, regulatory genomics, and biostatistics.	Pathogenicity predictions and recurrence statistics.	Predicting effect of mutations in protein function, regulation, splicing, etc. Establishing mutational background
Pathway and functional analysis.	Systems Biology, pathway modeling, and Interactomics.	Enrichment statistics and network analysis.	Multiple pathway definitions and violated statistical assumptions.
Integration, Visualization, and disease centered Interpretation.	Application development, Man-machine interfaces, Pharmacogenomics, Text-mining, and Information management.	Data integration and visualization.	Multiple, heterogeneous, experimental data sources, database formats, and software resources. Biomedical expertise required



Vazquez M, de la Torre V, Valencia A (2012) Chapter 14: Cancer Genome Analysis. PLOS Computational Biology 8(12): e1002824. https://doi.org/10.1371/journal.pcbi.1002824
http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002824

PLOS | COMPUTATIONAL BIOLOGY

First Cancer Genome



nature Vol 456 6 November 2008 doi:10.1038/nature07485

ARTICLES

DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome

Timothy J. Ley[1,2,3,4]*, Elaine R. Mardis[2,3]*, Li Ding[2,3], Bob Fulton[3], Michael D. McLellan[3], Ken Chen[3], David Dooling[3], Brian H. Dunford-Shore[3], Sean McGrath[3], Matthew Hickenbotham[3], Lisa Cook[3], Rachel Abbott[3], David E. Larson[3], Dan C. Koboldt[3], Craig Pohl[3], Scott Smith[3], Amy Hawkins[3], Scott Abbott[3], Devin Locke[3], LaDeana W. Hillier[3,8], Tracie Miner[3], Lucinda Fulton[3], Vincent Magrini[2,3], Todd Wylie[3], Jarret Glasscock[3], Joshua Conyers[3], Nathan Sander[3], Xiaoqi Shi[3], John R. Osborne[3], Patrick Minx[3], David Gordon[8], Asif Chinwalla[3], Yu Zhao[1], Rhonda E. Ries[1], Jacqueline E. Payton[5], Peter Westervelt[1,4], Michael H. Tomasson[1,4], Mark Watson[3,4,5], Jack Baty[6], Jennifer Ivanovich[6,7], Sharon Heath[1,4], William D. Shannon[1,4], Rakesh Nagarajan[4,5], Matthew J. Walter[1,4], Daniel C. Link[1,4], Timothy A. Graubert[1,4], John F. DiPersio[1,4] & Richard K. Wilson[2,3,4]

Acute myeloid leukaemia is a highly malignant haematopoietic tumour that affects about 13,000 adults in the United States each year. The treatment of this disease has changed little in the past two decades, because most of the genetic events that initiate the disease remain undiscovered. Whole-genome sequencing is now possible at a reasonable cost and timeframe to use this approach for the unbiased discovery of tumour-specific somatic mutations that alter the protein-coding genes. Here we present the results obtained from sequencing a typical acute myeloid leukaemia genome, and its matched normal counterpart obtained from the same patient's skin. We discovered ten genes with acquired mutations; two were previously described mutations that are thought to contribute to tumour progression, and eight were new mutations present in virtually all tumour cells at presentation and relapse, the function of which is not yet known. Our study establishes whole-genome sequencing as an unbiased method for discovering cancer-initiating mutations in previously unidentified genes that may respond to targeted therapies.

First Cancer Genome

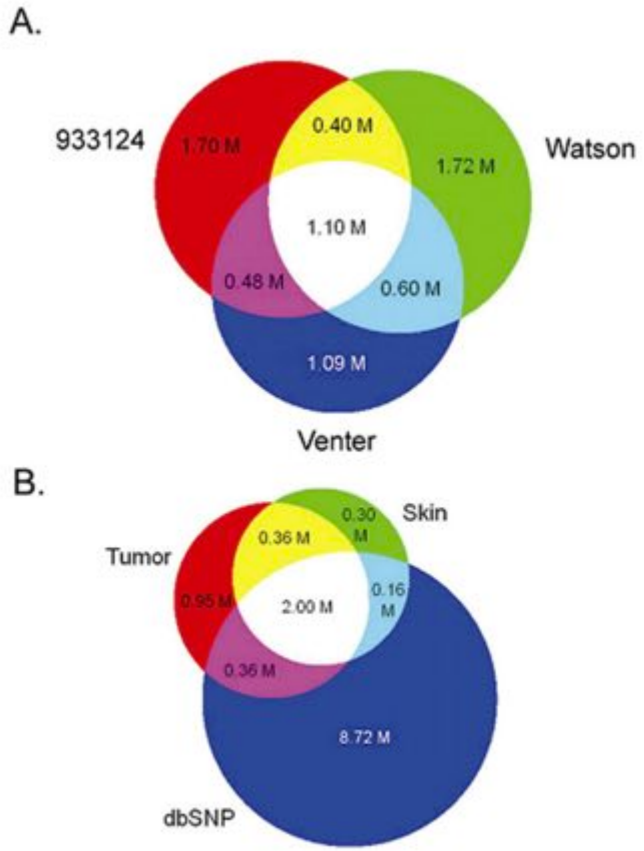
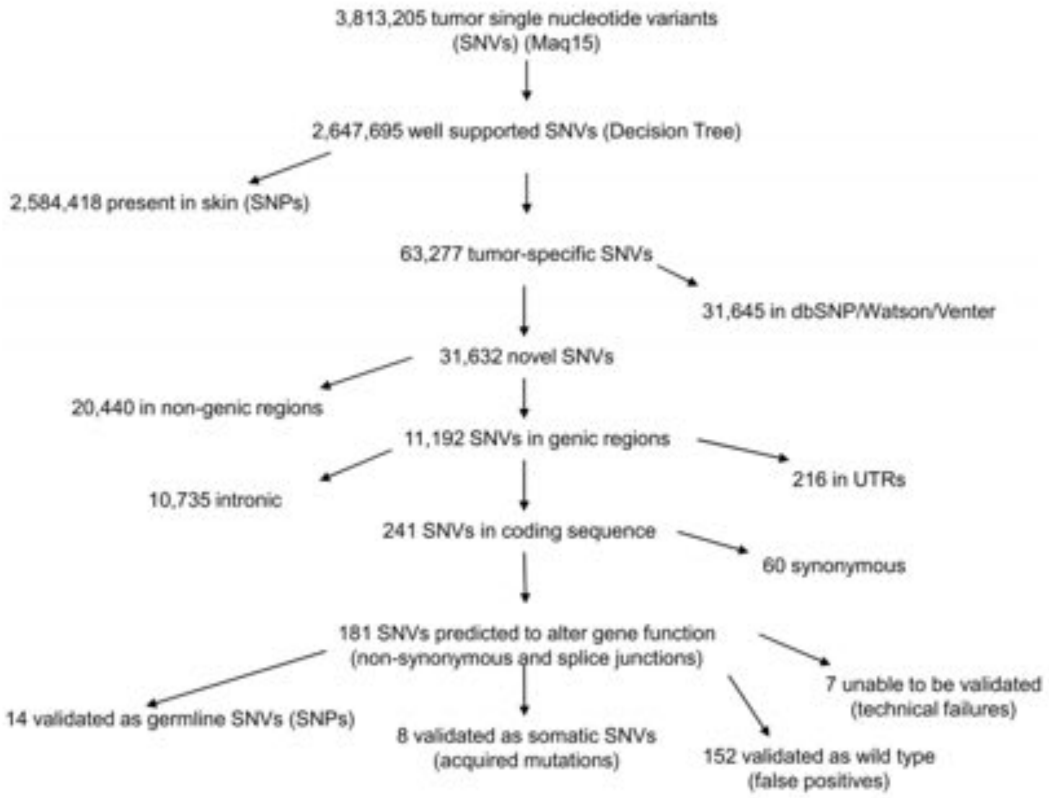


Figure 1. Overlap of SNPs detected in 933124 and other genomes
(A) Venn diagram of overlap between SNPs detected in the 933124 tumor genome and the genomes of Watson and Venter. (B) Venn Diagram of overlap among 933124 tumor genome, skin genome, and dbSNP (ver. 127). Single nucleotide variants were defined with a MAQ SNP quality ≥ 15.



3,813,205 tumor single nucleotide variants (SNVs) (Maq15)

2,647,695 well supported SNVs (Decision Tree)

2,584,418 present in skin (SNPs)

63,277 tumor-specific SNVs

31,645 in dbSNP/Watson/Venter

31,632 novel SNVs

20,440 in non-genic regions

11,192 SNVs in genic regions

216 in UTRs

10,735 intronic

241 SNVs in coding sequence

60 synonymous

181 SNVs predicted to alter gene function (non-synonymous and splice junctions)

7 unable to be validated (technical failures)

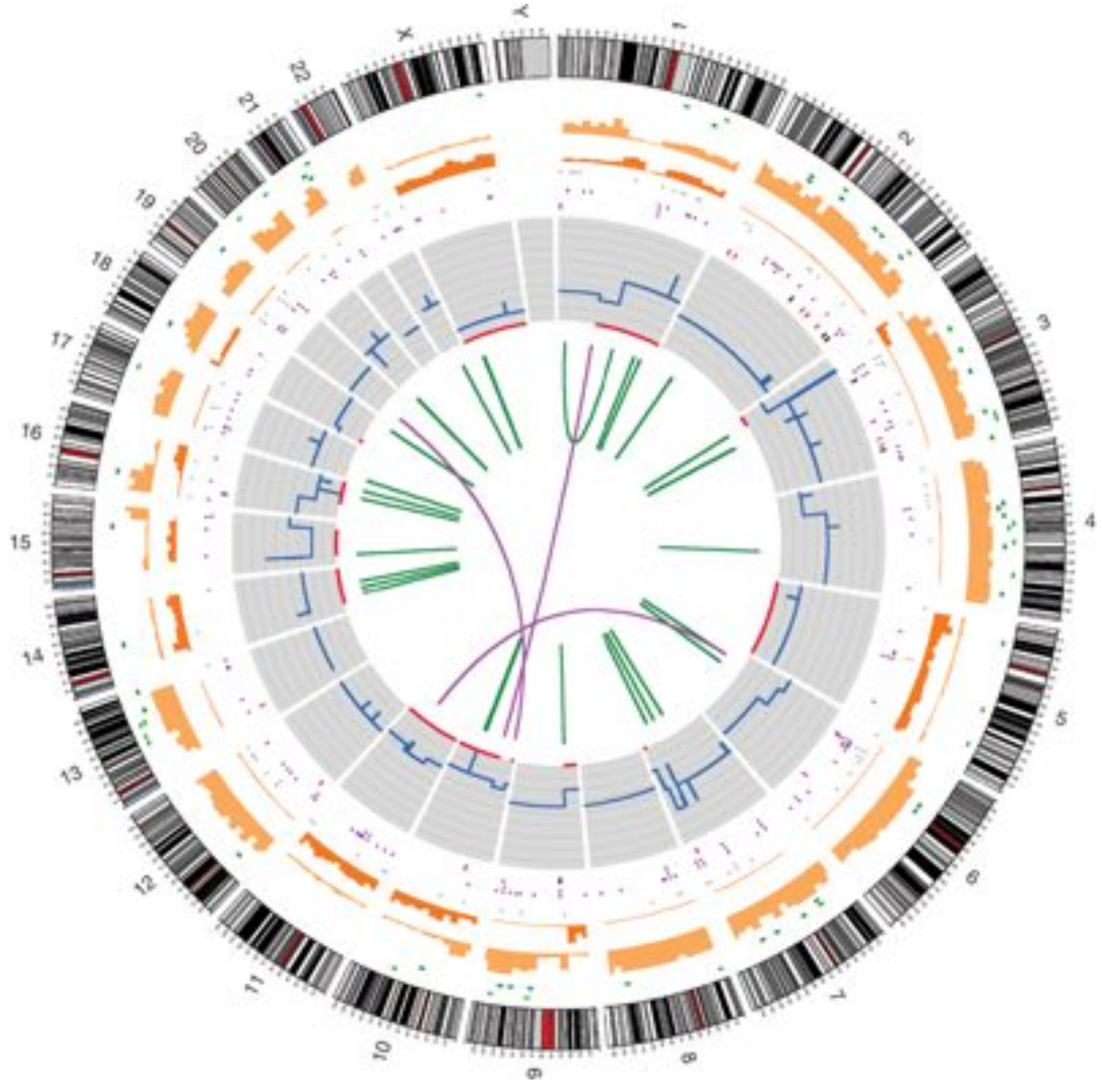
14 validated as germline SNVs (SNPs)

8 validated as somatic SNVs (acquired mutations)

152 validated as wild type (false positives)

Figure 2. Filters used to identify somatic point mutations in the tumor genome
See text for details.

First Melanoma Genome

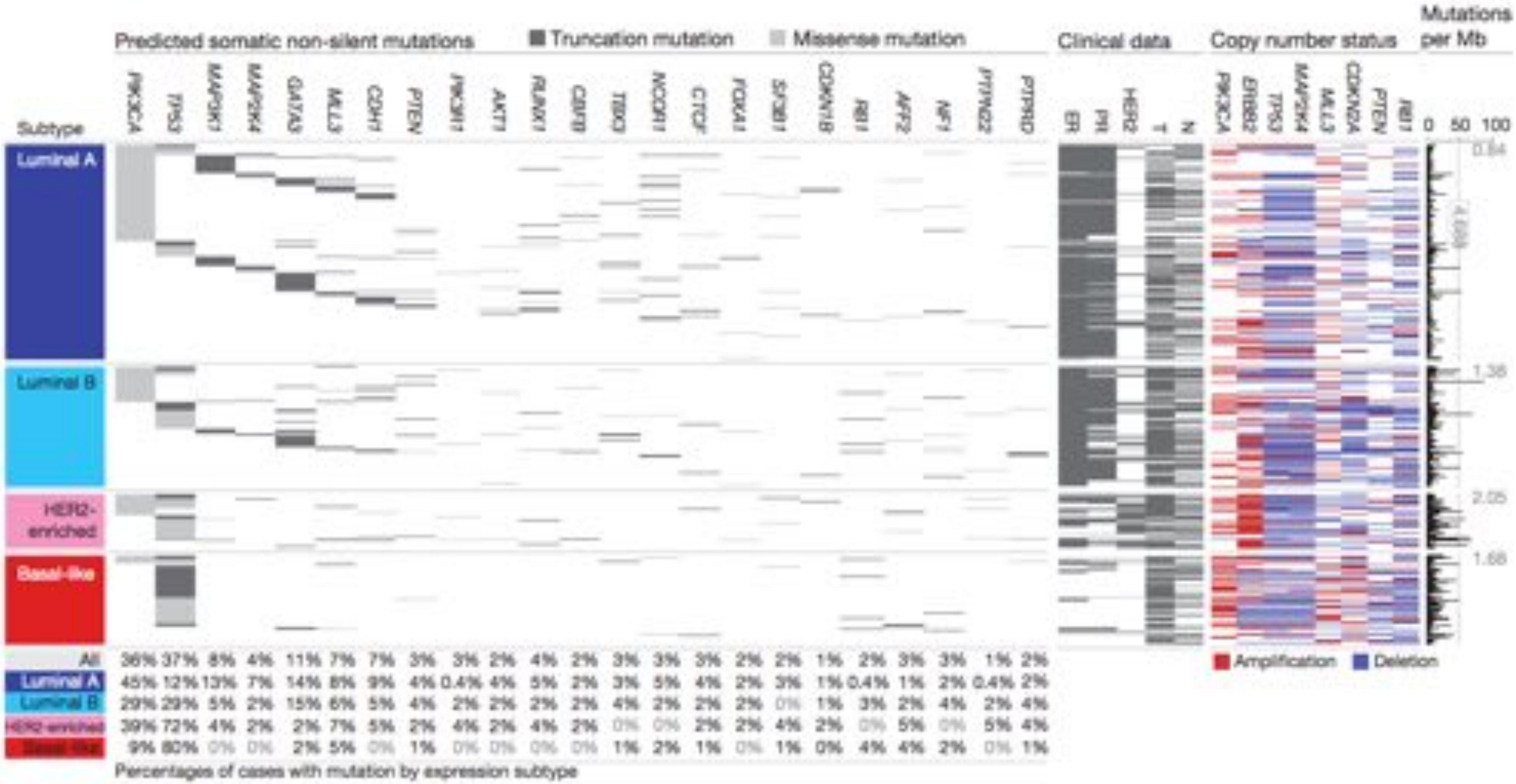


- Insertions (light-green rectangles);
- Deletions (dark-green rectangles);
- Heterozygous (light-orange bars) and Homozygous (dark-orange bars) Substitutions
- Coding substitutions (coloured squares: silent in grey, missense in purple, nonsense in red and splice site in black);
- Copy number (blue lines); regions of LOH (red lines);
- Intrachromosomal rearrangements (green lines);
- Interchromosomal rearrangements (purple lines).

A comprehensive catalogue of somatic mutations from a human cancer genome

Pleasance et al (2010) Nature. doi:10.1038/nature08658

Mutations in Breast Cancer



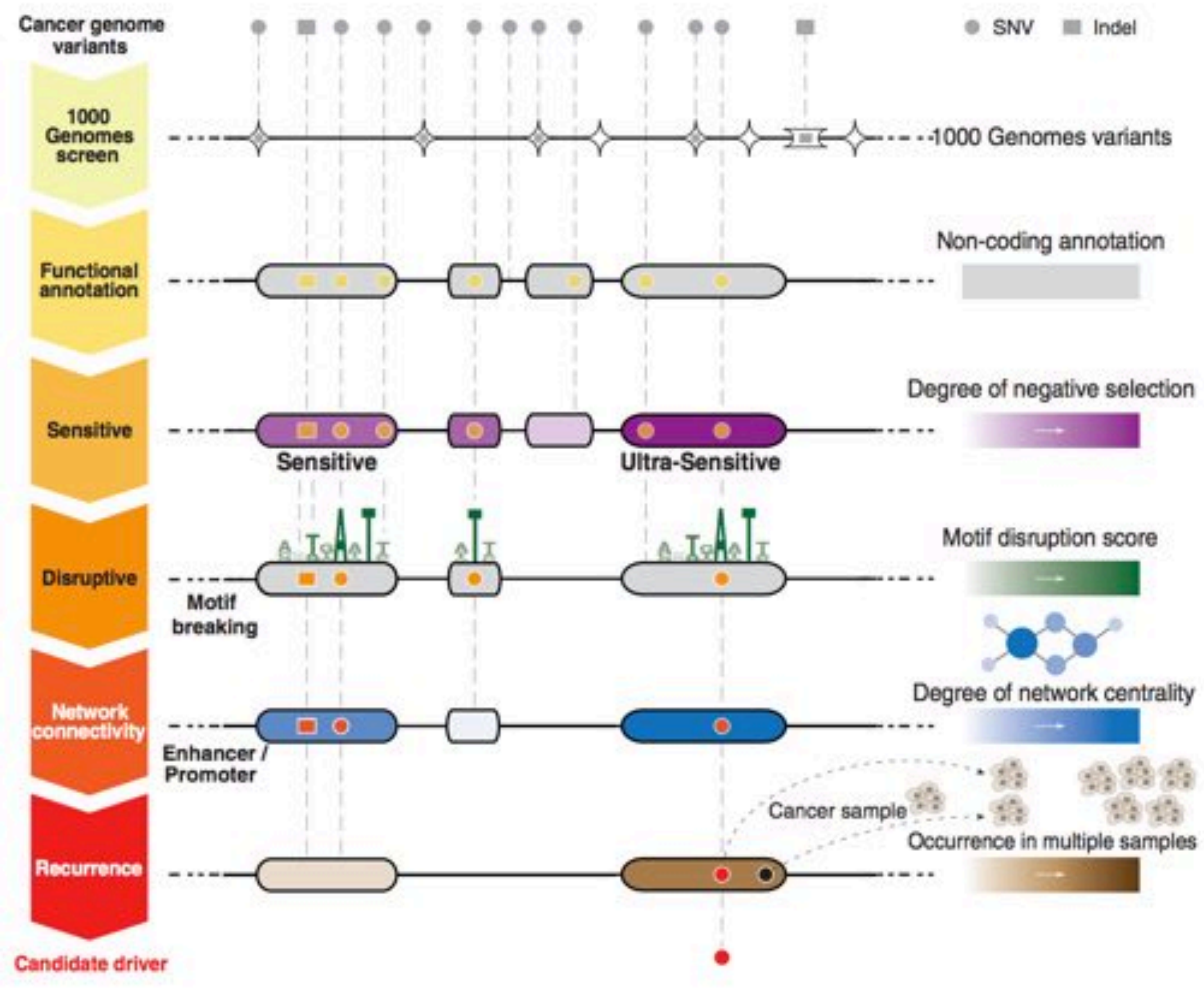
	PIK3CA	TP53	MAP3K1	MAP2K4	GATA3	MLL3	CDH1	PTEN	PIK3R1	AKT1	RUNX1	CBFB	TBX3	NCOR1	CTCF	FOXA1	SF3B1	CDKN1B	RB1	AFF2	NF1	PTPN22	PTPRD
All	36%	37%	8%	4%	11%	7%	7%	3%	3%	2%	4%	2%	3%	3%	3%	2%	2%	1%	2%	3%	3%	1%	2%
Luminal A	45%	12%	13%	7%	14%	8%	9%	4%	0.4%	4%	5%	2%	3%	5%	4%	2%	3%	1%	0.4%	1%	2%	0.4%	2%
Luminal B	29%	29%	5%	2%	15%	6%	5%	4%	2%	2%	2%	2%	4%	2%	2%	2%	0%	1%	3%	2%	4%	2%	4%
HER2-enriched	39%	72%	4%	2%	2%	7%	5%	2%	4%	2%	4%	2%	0%	0%	2%	2%	4%	2%	0%	5%	0%	5%	4%
Basal-like	9%	80%	0%	0%	2%	5%	0%	1%	0%	0%	0%	0%	1%	2%	1%	0%	1%	0%	4%	4%	2%	0%	1%

Percentages of cases with mutation by expression subtype

Comprehensive molecular portraits of human breast tumours

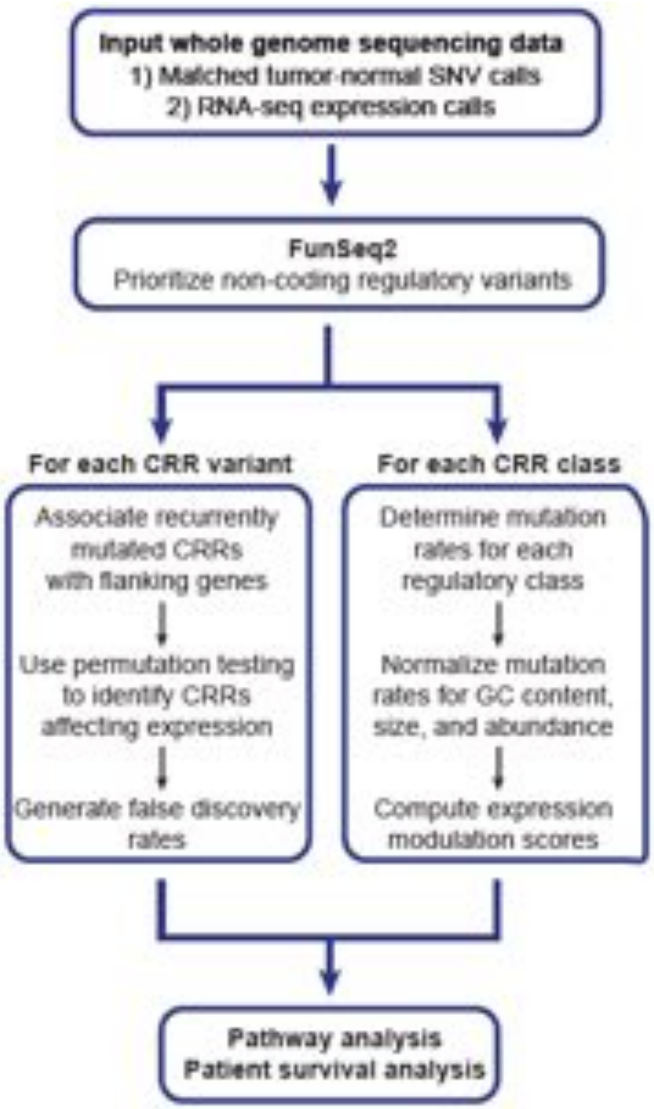
Cancer Genome Atlas Network (2012) Nature. doi:10.1038/nature11412

Finding Driving Mutations



Integrative Annotation of Variants from 1092 Humans: Application to Cancer Genomics
Khurana et al (2013) Science. DOI: 10.1126/science.1235587

Regulatory mutations in PDAC



Input whole genome sequencing data
1) Matched tumor-normal SNV calls
2) RNA-seq expression calls

FunSeq2
Prioritize non-coding regulatory variants

For each CRR variant

Associate recurrently mutated CRRs with flanking genes

Use permutation testing to identify CRRs affecting expression

Generate false discovery rates

For each CRR class

Determine mutation rates for each regulatory class

Normalize mutation rates for GC content, size, and abundance

Compute expression modulation scores

Pathway analysis
Patient survival analysis

Coding alterations of PDAC are now fairly well established but non-coding mutations (NCMs) largely unexplored

•Developed GECCO to analyze the thousands of somatic mutations observed from hundreds of tumors to find potential drivers of gene expression and pathogenesis

•NCMs are enriched in known and novel pathways

•NCMs correlate with changes in gene expression

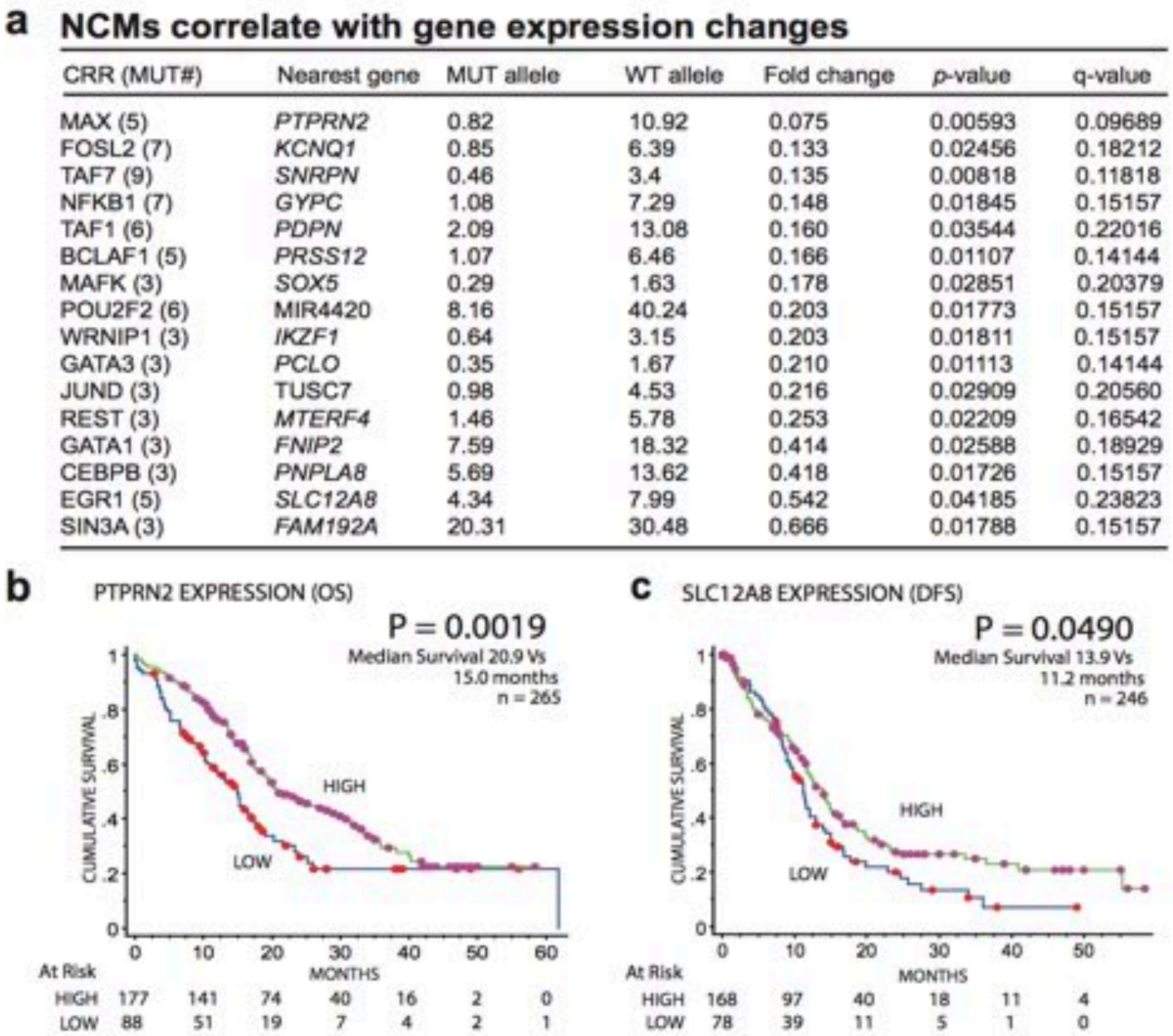
•NCMs can demonstrably modulate gene expression

•NCMs correlate with novel clinical outcomes

NCMs are an important mechanism for tumor genome evolution

Recurrent noncoding regulatory mutations in pancreatic ductal adenocarcinoma
Feigin, M, Garvin, T et al. (2017) Nature Genetics. doi:10.1038/ng.3861

Driving Non-Coding Mutations

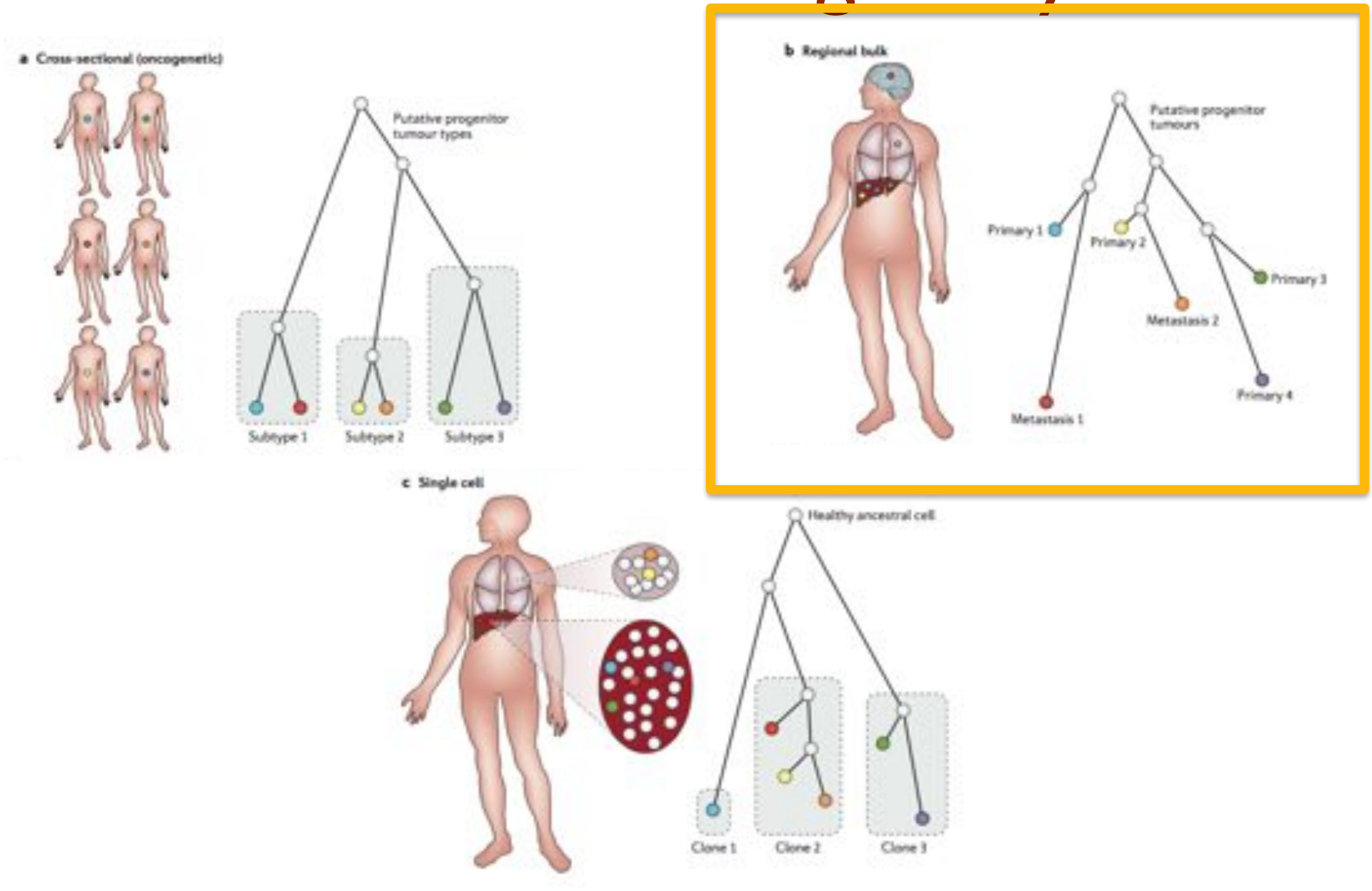


a NCMs correlate with gene expression changes

CRR (MUT#)	Nearest gene	MUT allele	WT allele	Fold change	*p*-value	q-value
MAX (5)	*PTPRN2*	0.82	10.92	0.075	0.00593	0.09689
FOSL2 (7)	*KCNQ1*	0.85	6.39	0.133	0.02456	0.18212
TAF7 (9)	*SNRPN*	0.46	3.4	0.135	0.00818	0.11818
NFKB1 (7)	*GYPC*	1.08	7.29	0.148	0.01845	0.15157
TAF1 (6)	*PDPN*	2.09	13.08	0.160	0.03544	0.22016
BCLAF1 (5)	*PRSS12*	1.07	6.46	0.166	0.01107	0.14144
MAFK (3)	*SOX5*	0.29	1.63	0.178	0.02851	0.20379
POU2F2 (6)	MIR4420	8.16	40.24	0.203	0.01773	0.15157
WRNIP1 (3)	*IKZF1*	0.64	3.15	0.203	0.01811	0.15157
GATA3 (3)	*PCLO*	0.35	1.67	0.210	0.01113	0.14144
JUND (3)	TUSC7	0.98	4.53	0.216	0.02909	0.20560
REST (3)	*MTERF4*	1.46	5.78	0.253	0.02209	0.16542
GATA1 (3)	*FNIP2*	7.59	18.32	0.414	0.02588	0.18929
CEBPB (3)	*PNPLA8*	5.69	13.62	0.418	0.01726	0.15157
EGR1 (5)	*SLC12A8*	4.34	7.99	0.542	0.04185	0.23823
SIN3A (3)	*FAM192A*	20.31	30.48	0.666	0.01788	0.15157

Recurrent noncoding regulatory mutations in pancreatic ductal adenocarcinoma
Feigin, M, Garvin, T et al. (2017) Nature Genetics. doi:10.1038/ng.3861

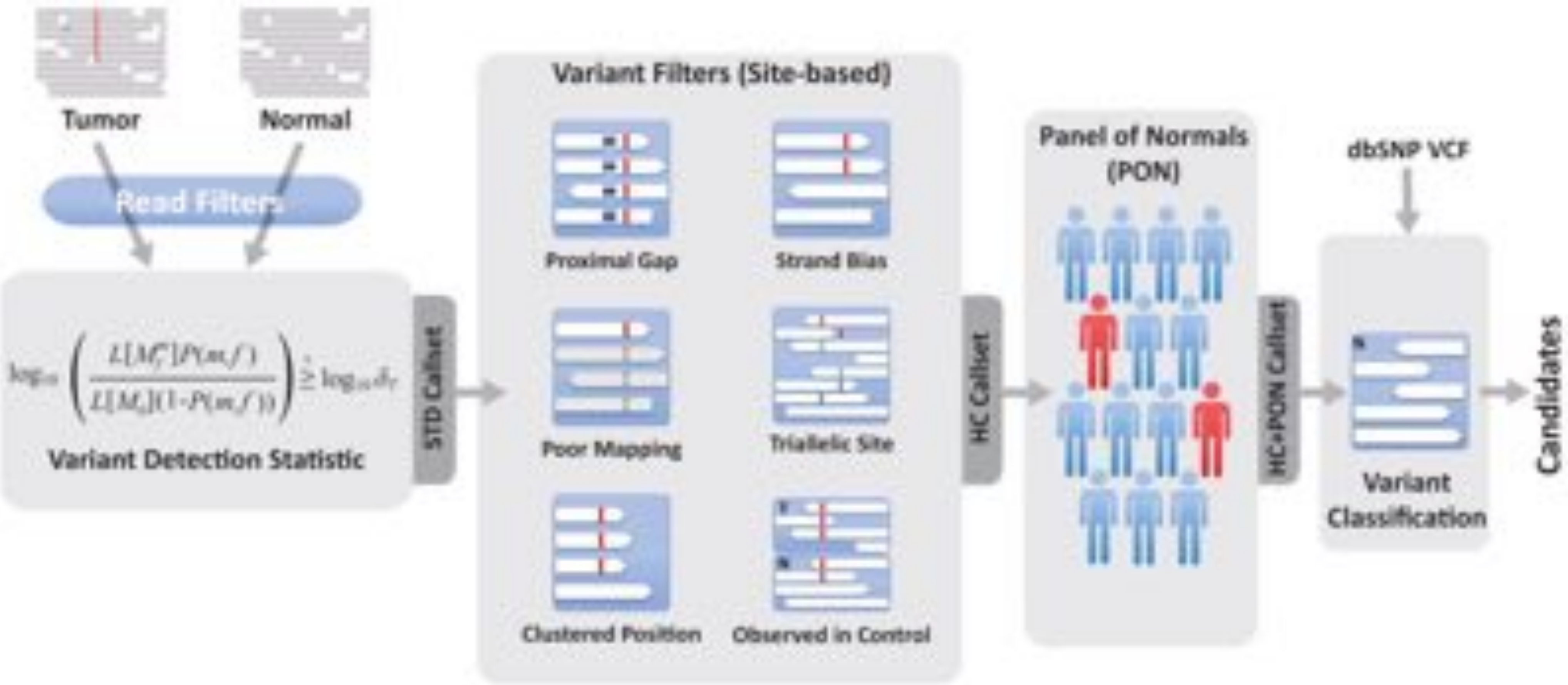
Tumor Heterogeneity



The evolution of tumour phylogenetics: principles and practice

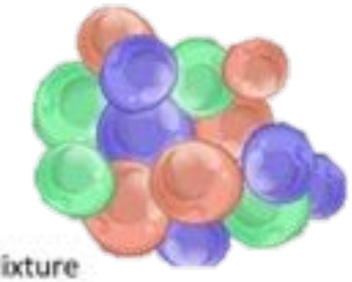
Schwarz and Schaffer (2017) *Nature Reviews Genetics. doi:10.1038/nrg.2016.170*

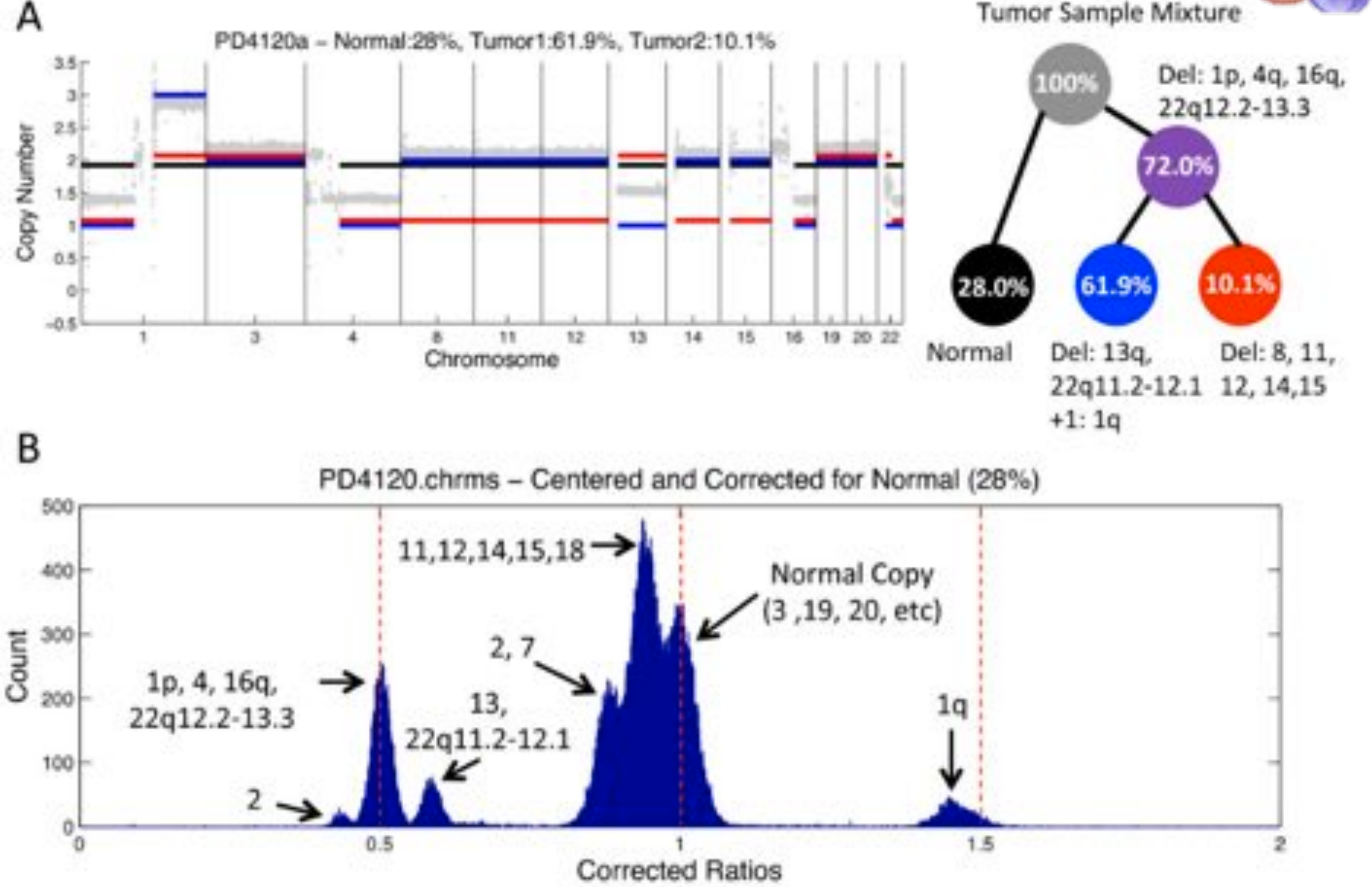
Tumor-Normal Pairs



Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples
Cibulskis et al (2013) Nature Biotech. doi:10.1038/nbt.2514

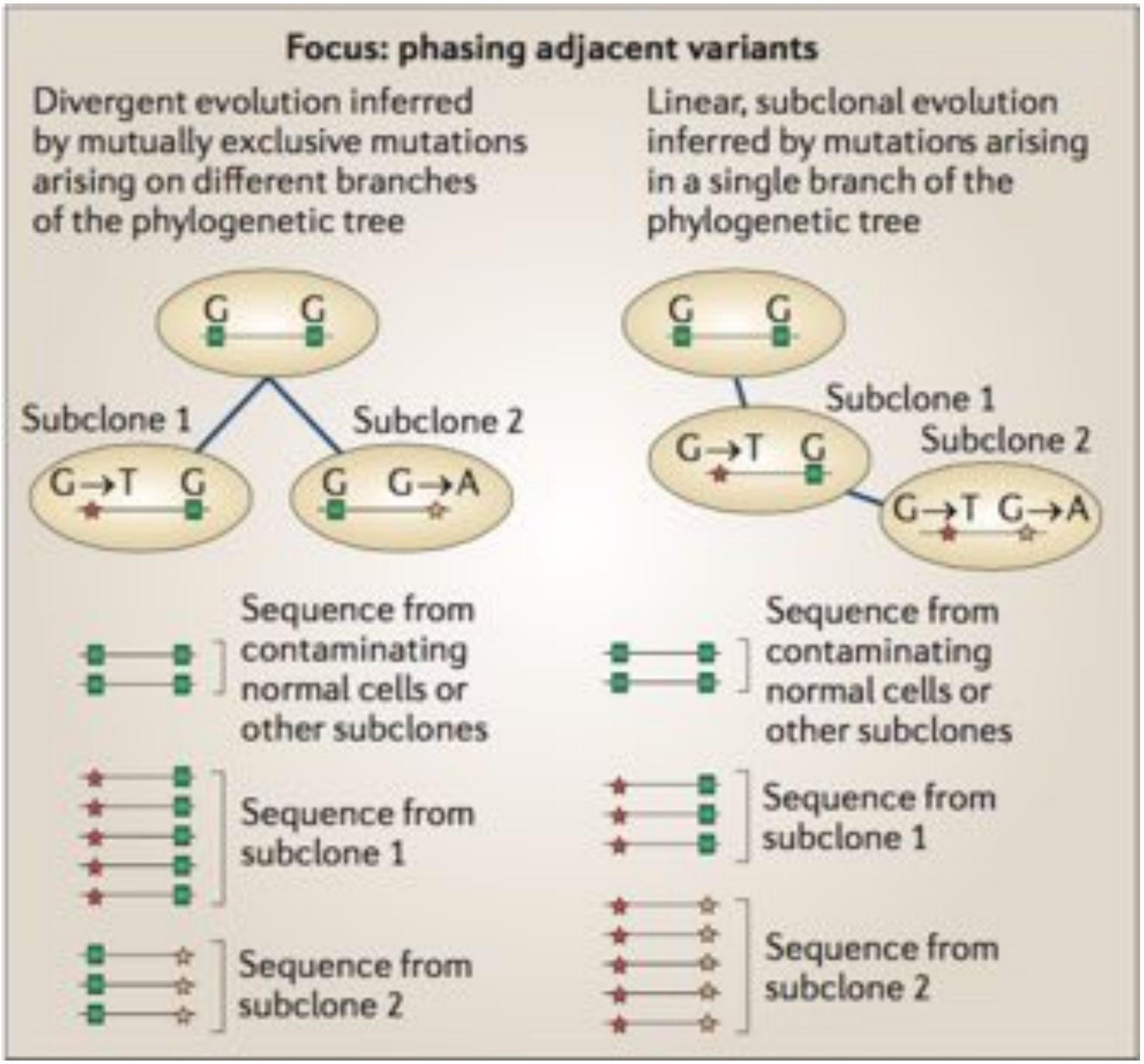
Bulk Heterogeneity





THetA: inferring intra-tumor heterogeneity from high-throughput DNA sequencing data
Oesperet al (2013) Genome Biology. DOI: 10.1186/gb-2013-14-7-r80

Somatic Variant Detection



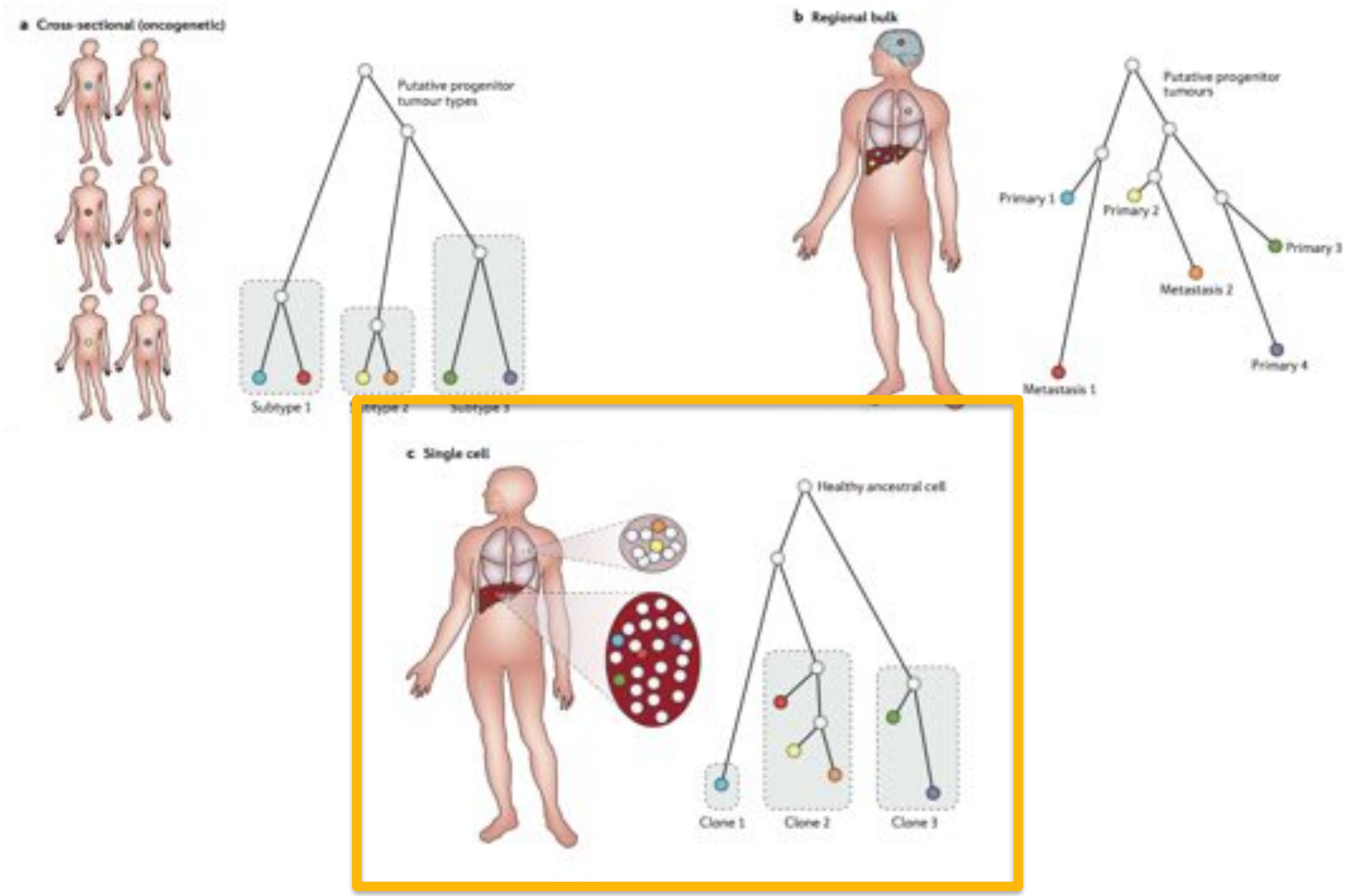
Focus: phasing adjacent variants

Divergent evolution inferred by mutually exclusive mutations arising on different branches of the phylogenetic tree

Linear, subclonal evolution inferred by mutations arising in a single branch of the phylogenetic tree

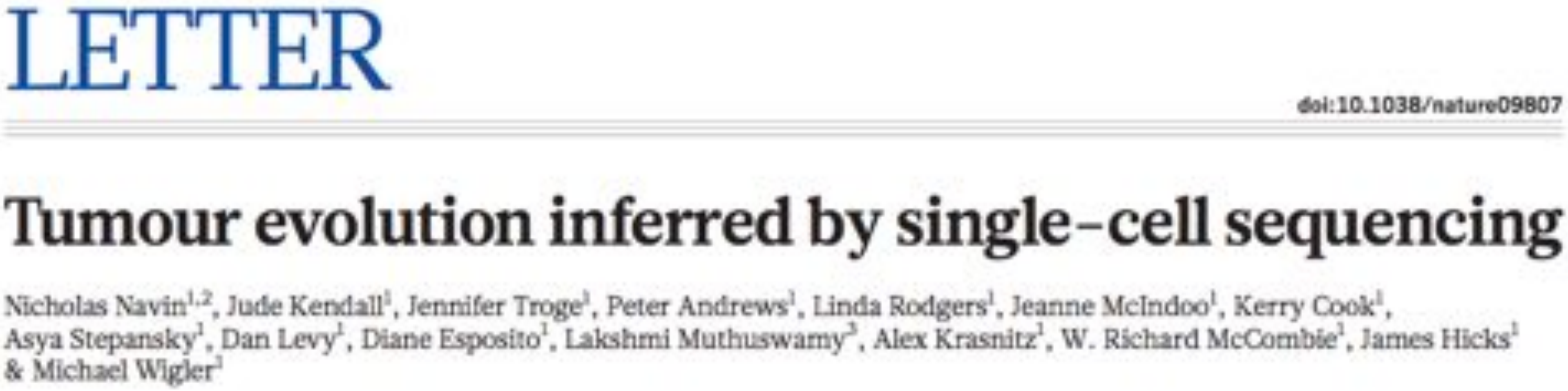
Evolution of the cancer genome
Yates & Campbell (2012) Nature Review Genetics. doi:10.1038/nrg3317

Tumor Heterogeneity



The evolution of tumour phylogenetics: principles and practice

Schwarz and Schaffer (2017) *Nature Reviews Genetics. doi:10.1038/nrg.2016.170*

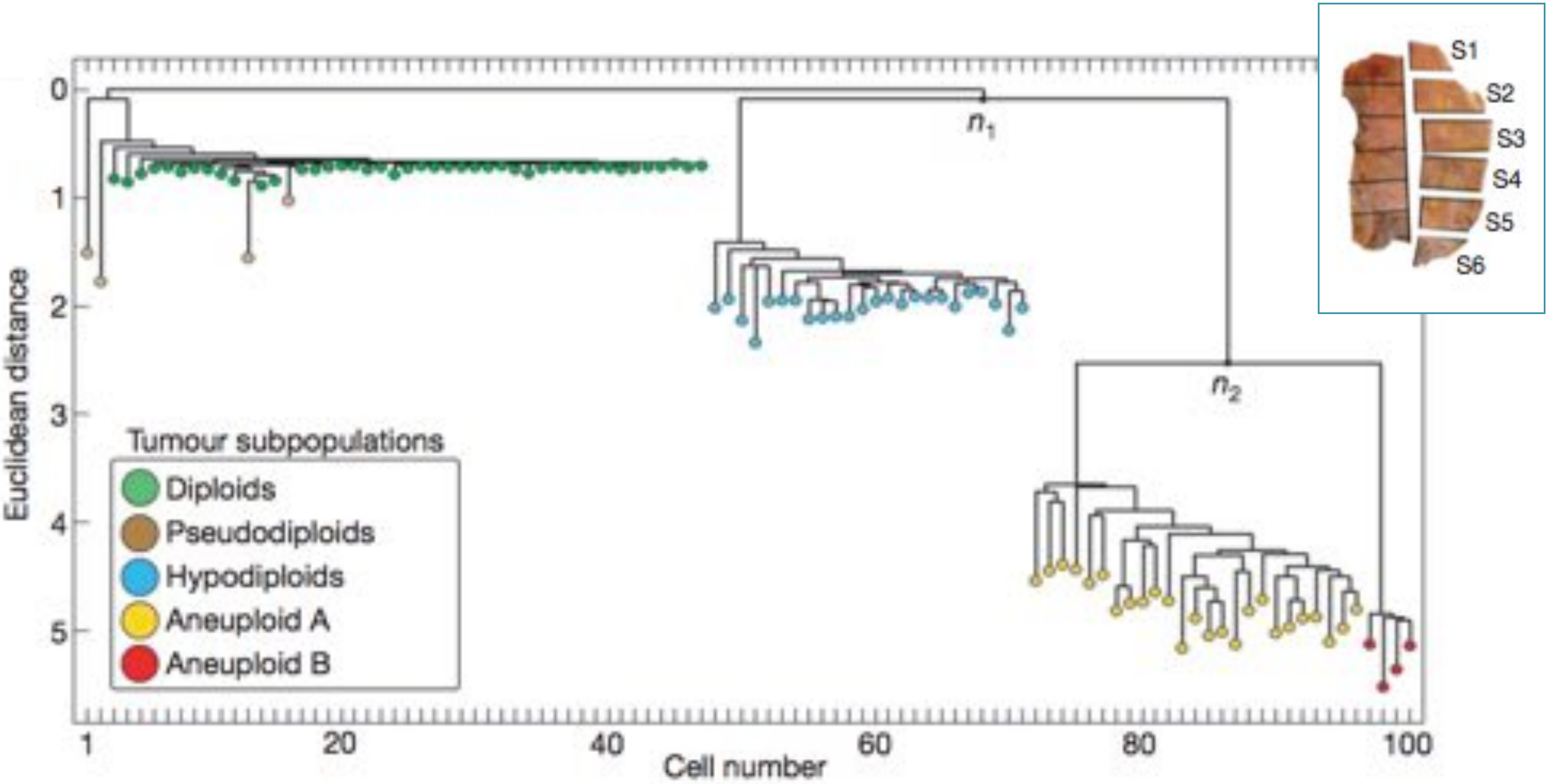


LETTER

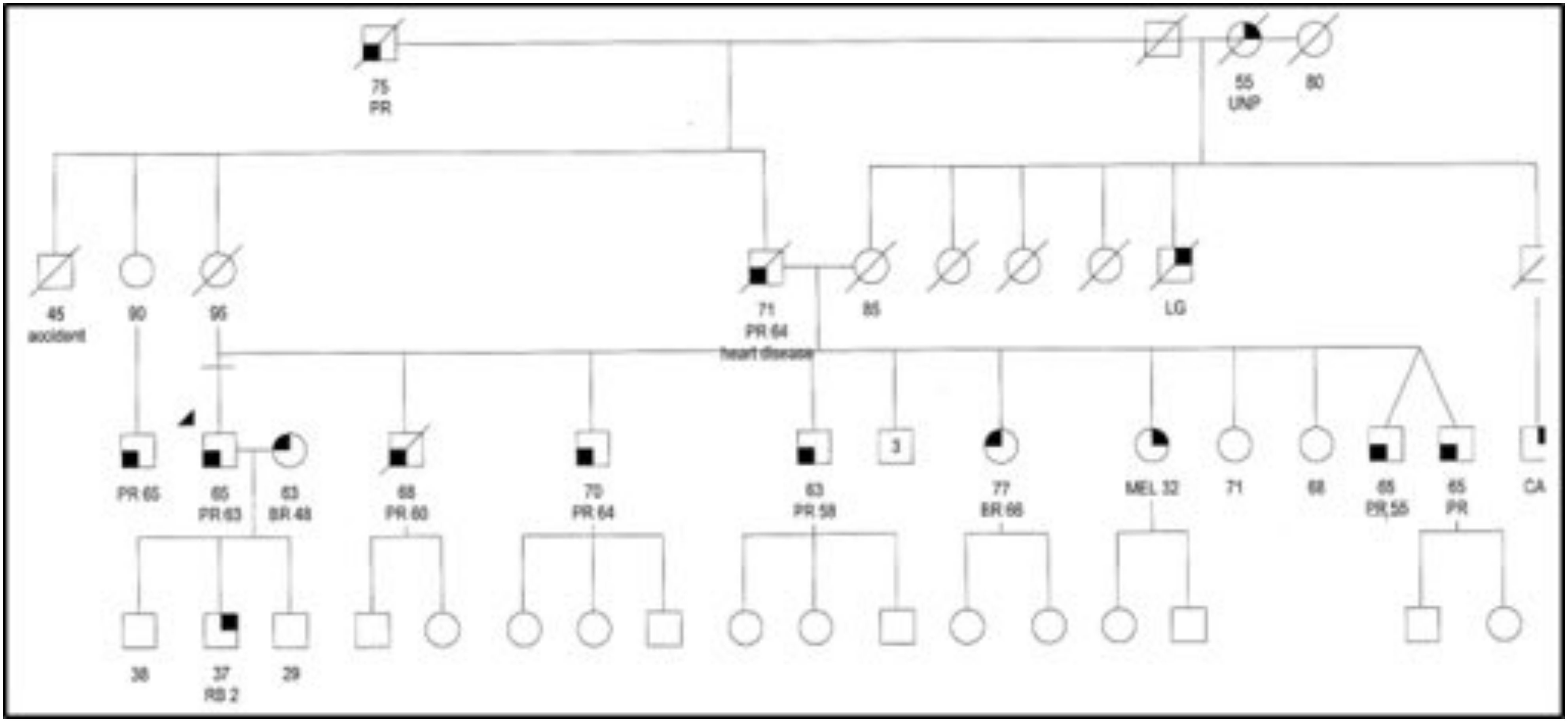
doi:10.1038/nature09807

Tumour evolution inferred by single-cell sequencing

Nicholas Navin[1,2], Jude Kendall[1], Jennifer Troge[1], Peter Andrews[1], Linda Rodgers[1], Jeanne McIndoo[1], Kerry Cook[1], Asya Stepansky[1], Dan Levy[1], Diane Esposito[1], Lakshmi Muthuswamy[3], Alex Krasnitz[1], W. Richard McCombie[1], James Hicks[1] & Michael Wigler[1]



What causes "outlier" families?



The Mark Foundation® for Cancer Research

Eli Van Allen,
Dana-Farber Cancer Institute