SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF BIO AND CHEMICAL ENGINEERING

DEPARTMENT OF CENTRE FOR MOLECULAR AND NANOMEDICAL SCIENCES

UNIT – I - HUMAN GENETICS – (SMB5202)

Introduction to human genetics

Human genetics is the study of inheritance as it occurs in human beings. Human genetics encompasses a variety of overlapping fields including: classical genetics, cytogenetics, moleculargenetics, biochemical genetics, genomics, population genetics, developmental genetics, clinical genetics, and genetic counseling. Genes are the common factor of the qualities of most human-inherited traits. Study of human genetics can answer questions about human nature, can help understand diseases and the development of effective disease treatment, and help us to understand the genetics of human life. This article describes only basic features of human genetics; for the genetics of disorders please see: medical genetics.

History of Human Genetics

For almost all human diseases, individual susceptibility is, to some degree, influenced by genetic variation. Consequently, characterizing the relationship between sequence variation and disease predisposition provides a powerful tool for identifying processes fundamental to disease pathogenesis and highlighting novel strategies for prevention and treatment. Over the past 25 years, advances in technology and analytical approaches, often building on major community projects—such as those that generated the human genome sequence1 and elaborated on that reference to capture sites of genetic variation2,3,4,5,6—have enabled many of the genes and variants that are causal for rare diseases to be identified and enabled a systematic dissection of the genetic basis of common multifactorial traits. There is growing momentum behind the application of this knowledge to drive innovation in clinical care, most obviously through developments in precision medicine. Genomic medicine, which was previously restricted to a few specific clinical indications, is poised to go mainstream.

This Review charts recent milestones in the history of human disease genetics and provides an opportunity to reflect on lessons learned by the human genetics community. We focus first on the long-standing division between genetic discovery efforts targeting rare variants with large effects and those seeking alleles that influence predisposition to common diseases. We describe how this division, with its echoes of the century-old debate between Mendelian and

biometric views of human genetics, has obscured the continuous spectrum of disease risk alleles—across the range of frequencies and effect sizes—observed in the population, and outline how genome-wide analyses in large biobanks are transforming genetic research by enabling a comprehensive perspective on genotype–phenotype relationships. We describe how the expansion in the scale and scope of strategies for enumerating the functional consequences of genetic variation is transforming the torrent of genetic discoveries from the past decade into mechanistic insights, and the ways in which this knowledge increasingly underpins advances in clinical care. Finally, we reflect on some of the challenges and opportunities that confront the field, and the principles that will, over the coming decade, drive the application of human genetics to enhance understanding of health and disease and maximize clinical benefit.

Rare diseases, rare variants

During the 1980s and 1990s, efforts to map disease genes were focused on rare, monogenic and syndromic diseases and were mostly driven by linkage analysis and fine mapping within large multiplex pedigrees. Localization of genetic signals was typically followed by Sanger sequencing of the genes found to map within the linked locus to identify disease-causing alleles. Assessments of pathogenicity, based on segregation of a putatively causal variant with disease across multiple families and evidence that the risk genotype was absent in healthy individuals, were typically followed by confirmatory functional studies in cellular and animal models. This path to gene identification was laborious; nevertheless, by 2000, around 1,000 of the estimated 7,000 single-gene inherited diseases had been characterized, including many with substantial biomedical impact, such as Huntington's disease and cystic fibrosis[7,8,9].

Completion of the draft human genome sequence[1] reduced many of the obstacles to disease- gene mapping and propelled a fourfold increase in the genes implicated as causal for rare, single-gene disorders (Fig. 1). Microarray-based detection of structural variation[10] and exome- and genome-wide sequencing[11,12] have been pivotal, bolstered by in silico analysis and prioritization of the discovered genetic variants. Increasing availability of reference datasets cataloguing population genetic variation across diverse ethnic backgrounds has supported robust causal inference[2,3,5,6]. More recently, the adoption of high-throughput sequencing technologies has enabled the full range of causal genetic variation, from single mutations to large structural

rearrangements, to be identified in a single assay. These technologies have extended from research into clinical usage, driving earlier and faster diagnosis for genetic disorders.

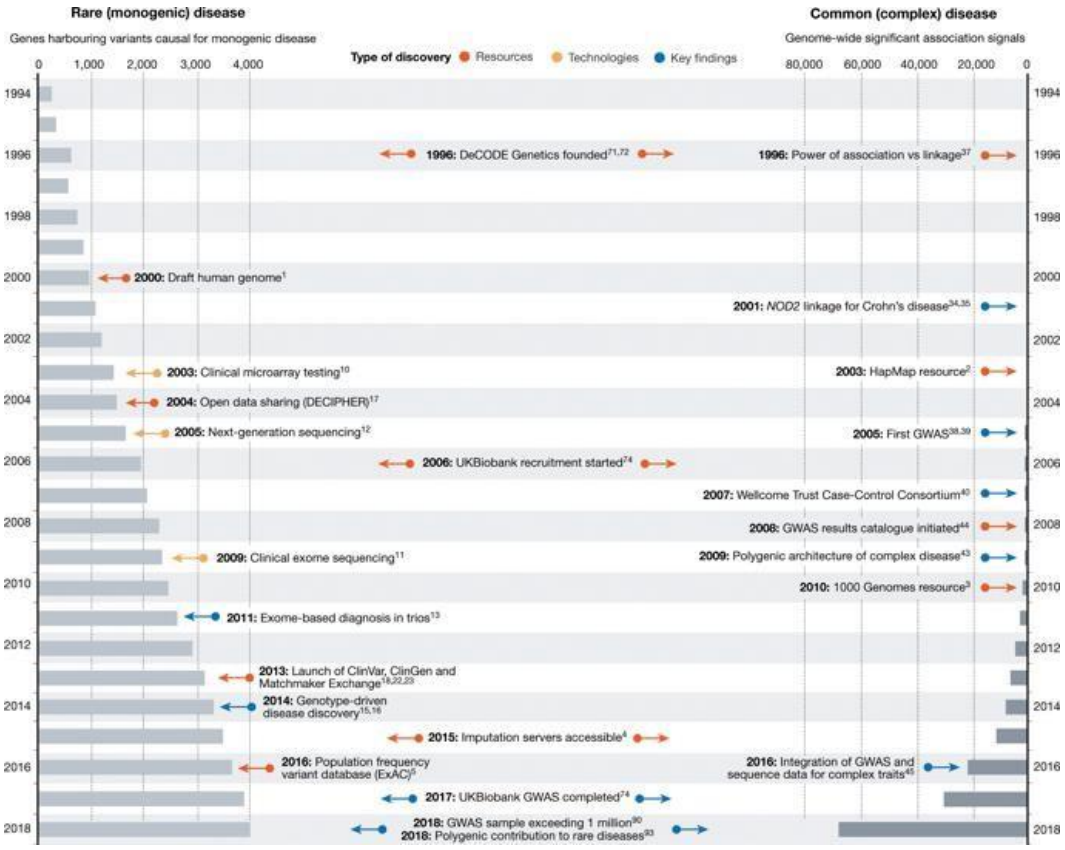


Fig. 1: Growth in the discovery of disease-associated genetic variation.

The cumulative numbers of genes harbouring variants causal for rare, monogenic diseases and traits and of significant GWAS associations implicated in common, complex diseases and traits are shown. Left, the advent of high-throughput sequencing technologies and availability of reference genomes from diverse populations has supported a fourfold increase in the discovery of rare disease-causing genes between 1999 and 2019. Right, international efforts such as the Human Genome Project and the HapMap Project, combined with GWAS and sequencing studies, have supported identification of more than 60,000 genetic associations across thousands of human diseases and traits. Centre, more recent developments have brought a synthesis of the rare- and common-variant approaches based around the combination of sequence-informed analyses in large cohorts. Key events contributing to these themes are depicted in the timeline. GA4GH, Global Alliance for Genomics and Health160; ExAC, Exome Aggregation Consortium5.

Reduced reliance on multiplex pedigrees in favour of collections of affected

cases, often with parents13, has proven decisive in identifying new dominant disorders, many of which were previously considered recessive14. Increasingly, discovery of rare disease genes has transitioned from genetic characterization of small numbers of individuals with similar clinical presentations to genome-wide sequencing of larger cohorts of phenotypically diverse patients. This genotype-driven approach has revealed new disorders associated with more variable clinical presentation15,16.

A more systematic approach to data sharing has been critical, both for the characterization of new disorders and diagnostic interpretation of potential causal alleles. The value of sharing genetic and phenotypic data from those thought to harbour rare undiagnosed genetic diseases has fostered global collaborative networks (for example, Matchmaker Exchange, DECIPHER and GeneMatcher) designed to match patients with similar genetic variants and/or phenotypic manifestations, even across continents17,18,19. Interactions between researchers and families with rare disease have enabled natural history studies to be driven by family support groups positioned to initiate data collection from patient cohorts once a causal gene is discovered20.

Clinical translation of these technologies has benefited from a series of information resources, including open databases of genes associated with rare disorders (for example, OMIM and ORPHANET)21, clinically interpreted variants (for example, ClinVar and ClinGen)22,23 and patient records (for example, DECIPHER and MyGene2 (https://mygene2.org/MyGene2))17. Access to resources that catalogue genetic variation across populations (such as ExAC and its successor gnomAD)5,6 has enabled the confident exclusion of genetic variants too common in population-level data to be plausible causes of rare, penetrant early-onset genetic diseases24. These analyses have reduced the contamination of databases with variants erroneously interpreted as causal for disease, and are addressing the overestimation of disease penetrance arising from the historical focus on multiplex pedigrees25. Improved recognition of the variable penetrance of many ‘monogenic’ disease alleles has invigorated efforts to identify the genetic and environmental modifiers responsible26,27.

Although huge strides have been made in associating specific genes with particular disorders, establishing the causal role of individual variants within those genes remains problematic, and many patients with suspected rare genetic

diseases are left without a definitive diagnosis28. Even for variants with established causality, the penetrance is often unclear. Resolving these uncertainties represents the central challenge for the field. Aggregation of sequencing data from large numbers of affected cases and population reference samples will provide the evidence base required for robust interpretation of variants. Highly parallelized in-vitro cellular assays that allow assessment of the functional effects of all variants in a disease- associated gene can transform interpretation of novel variants29, although developing well-calibrated functional assays predictive of pathogenicity for all disease genes represents a daunting prospect. Direct functional genomic exploration of accessible and disease-relevant tissues from patients using RNA sequencing and DNA methylation assays30,31 can identify previously cryptic causal genetic variants, particularly in under-explored regions outside protein-coding genes32,33. Developments in each of these areas will extend the range of variants and genes for which diagnostic and prognostic clinical information can be provided to patients and their families.

Common diseases, common variants

Efforts to apply the approach—linkage analysis in multiplex pedigrees—that had been so successful for the high-penetrance variants responsible for Mendelian disease were, with notable exceptions34,35,36, largely unsuccessful for common, later-onset traits with more complex multifactorial aetiologies, such as asthma, diabetes and depression. Recognition that association-based methods, focused on detecting phenotype-related differences in variant allele frequencies might have greater traction for identifying less penetrant common alleles redirected attention to analysis of case–control samples37. However, initial efforts targeting variants within ‘candidate’ genes were plagued by inadequate power, unduly liberal thresholds for declaring significance and scant attention to sources of bias and confounding, resulting in overblown claims and failed replication.

Systematic efforts to characterize genome-wide patterns of genomic variation, initially through the HapMap Consortium2, proved catalytic, demonstrating that the allelic structure of the genome was segmented into haplotype blocks, each containing sets of correlated variants. Recognition that this configuration could support genome-wide surveys of association energized the technological innovation—in the form of massively parallel genotyping arrays—to make such

studies possible (Fig. 1). Early wins in acute macular degeneration38 and inflammatory bowel disease39 were encouraging, and progress on several fronts—expansion of study size, denser genotyping arrays, novel strategies for imputation, attention to biases and appropriate significance thresholds—delivered robust associations across a range of diseases40. Most variants uncovered by these early genome- wide association studies (GWAS) were common, with more subtle effects than many had anticipated. A host of trait-specific consortia formed, covering diverse dichotomous and quantitative phenotypes, to accelerate genetic discovery through the aggregation and meta-analysis of data from multiple GWAS studies41,42,43. Many tens of thousands of robust associations were identified44. Recently, increased access to exome and whole-genome sequence data has, through both direct association analysis45,46 and imputation3,4, extended discovery to low-frequency and rare alleles previously inaccessible to GWAS.

In the decade since the first GWAS, understanding of the genetic basis of common human disease has been transformed. The disparity between the observed effects of the variants first identified by GWAS and estimates of overall trait heritability (the 'missing heritability' conundrum) is now largely resolved47. Common diseases are not simply aggregations of related Mendelian conditions: for most complex traits, genetic predisposition is shared across thousands of mostly common variants with individually modest effects on population risk41,43.

Although the collective contribution of low-frequency and rare risk alleles to overall trait variability appears modest compared with that attributable to common variants45,48, the rare risk alleles detected in current sample sizes necessarily have large phenotypic effects and are proportionately more likely to be coding, enhancing their value for biological inference. Founder populations (such as those from Finland and Iceland) have provided multiple examples of otherwise rare risk alleles driven to higher frequency locally through drift and/or selection49,50,51,52. In addition, studies in populations with high rates of consanguinity make it possible to identify individuals homozygous for otherwise rare loss-of-function
alleles, the basis for a 'human knockout' project to systematically investigate the phenotypic consequences of gene disruption in humans53,54.

For most diseases, large-scale GWAS-aggregation efforts have been disproportionately powered by information from individuals of European descent55. Whereas patterns of genetic predisposition appear broadly similar across major population groups and many common risk alleles discovered in one population group are detectable in others, allele frequencies can vary substantially; extending GWAS and sequencing studies to diverse populations will surely generate a rich harvest of novel risk alleles.

The relative contributions of common and rare variants indicate that, for many traits, particularly those with post-reproductive onset, purifying selection has had only limited effect45,56. For a few risk alleles, hallmarks of balancing selection reflect increased carrier survival, usually through protection from infectious diseases. This includes well-known examples of alleles maintained at high frequency in populations of African descent57,58.

While the extensive linkage disequilibrium within human populations has been essential to discovery in GWAS, high correlation between adjacent variants frustrates mapping of the specific variants responsible for these associations. Increasing sample size, improved access to trans-ethnic data, and more representative imputation reference panels3 provide a path to improved resolution of the causal variants59 and clues to the molecular mechanisms through which they operate. Functional interpretation is easiest for causal variants within coding sequences; however, most common disease-risk variants map to noncoding sequences, and are presumed to influence predisposition through effects on transcriptional regulation. In these cases, mechanistic inference depends on connecting association signals to their downstream targets (see below). For many traits, there is clear convergence between common-variant association signals and genes implicated in monogenic forms of the same disease, as well as enrichment of GWAS signals in regulatory elements specifically active in cell types consistent with known disease biology60,61. This provides reassurance that, even as the number of association signals for a given disease proliferates, the genetic associations uncovered will coalesce around molecular and cellular processes with a core role in pathogenesis62,63.

Importantly, the signals discovered by GWAS have revealed many unexpected insights into the biological basis of complex disease. Examples include the role of complement in the pathogenesis of acute macular degeneration38, synaptic

pruning in schizophrenia64 and autophagy in inflammatory bowel disease65. In addition, as inherited sequence variation is a prominent cause of phenotypic variation (but the reverse is not true), risk variants identified by GWAS have value as genetic instruments, mapping causal relationships between traits and inferring contributions made by circulating biomarkers and environmental exposures to disease development66.

As described below, findings from GWAS have increasing translational impact through identification of novel therapeutic targets67, prioritization (and deprioritization) of existing ones68 and development of polygenic scores that quantify individual genetic risk69.

Comprehensive genotype–phenotype maps

The historical division of disease-gene discovery into monogenic and polygenic strands arose from development and implementation of analytical approaches—family-based linkage and case–control association37—that are best-suited for detecting particular subsets of causal alleles. This obscured the true state of nature, with disease-risk alleles being distributed across a continuous spectrum of frequencies and effect sizes. In addition, the trait- and disease-specific perspective of early GWAS discovery (mostly reliant on case–control studies) was poorly equipped to investigate the contribution of genetic variants to phenotypic effects that are nested within or spread across classical disease definitions. Recent developments have enabled a more holistic perspective on genotype–phenotype relationships (Fig. 1).

One major advance has been the increasing availability of large prospective population-based cohorts. These biobank efforts, pioneered in studies such as the Framingham Cohort70 and the efforts of DeCODE in Iceland71,72, now encompass a growing inventory of national cohorts in North America, Europe, Asia and beyond73,74,75,76. The UK Biobank study, including 500,000 largely healthy, middle-aged participants has been particularly influential, transforming human genetic research in part through permissive data-sharing policies that have allowed multiple research groups to analyse the data74. Efforts to make clinical data embedded in electronic health records and registries available for research77,78 mean that biobanks increasingly provide access to a wide range of demographic, clinical and lifestyle data, captured in harmonized, systematic fashion from large, often multi-ethnic collections of individuals. For millions of

biobank participants, this rich phenotypic information has been combined with genome-wide genetic data. There are nascent efforts to capture transcriptomic, proteomic and metabolomic phenotypes, although these are not yet at equivalent scale to the genetic data[79,80]. Biobank analyses have provided more generalizable estimates of the relevance of genetic risk factors in the context of the separate and joint effects of non-genetic factors[81]. Increasingly, integration with healthcare data brings a longitudinal dimension to phenotypic characterization, which facilitates analyses of disease progression and lifelong disease risk[82].

The rich phenotypic scope of these cohorts has enabled variants of interest to be interrogated for associations across the gamut of available phenotypes. These phenome-wide association studies (PheWAS) have revealed the extent to which many variants have pleiotropic effects across multiple traits[83]. Some of these relationships are expected, such as the impact of obesity variants on risk of hepatic steatosis and type 2 diabetes[84] or variants that influence multiple autoimmune conditions[85]. Others connect diseases and traits in surprising ways, highlighting shared polygenic, pleiotropic effects and cell-type specificity, and delivering insights into shared biology and overlapping mechanisms[86,87]. These findings inform the prioritization of therapeutic targets, providing clues to potential on-target side effects and opportunities for drug repurposing[87,88,89].

The second enabler of inclusive, systematic analysis of genotype–phenotype relationships has been access to whole-genome sequence data. The scale of genetic analysis based on sequence data still lags behind that of genome-wide genotyping data (the largest sequence-based datasets are one tenth the size of the largest GWAS[90,91,92]), although reductions in sequencing costs are decreasing the differential. Most direct analysis of high-throughput sequence data has focused on the coding regions. Strategies for assigning variant function and jointly analysing sets of variants of similar functional effect have enabled aggregate, gene- level tests of rare functional-variant association that are often better powered than single- variant tests[91,92]. However, the principal benefit to date of whole-genome sequence data to genetic discovery has been to bolster array-based access to lower-frequency alleles, either directly, through their inclusion on genotyping platforms, or indirectly, through imputation from sequence-based reference samples[3,4].

These developments have enabled researchers to bridge the gap between the monogenic and polygenic realms, identifying common variant modifiers of monogenic phenotypes contributing to the variable expression of rare, large-effect alleles[26,93], and low-frequency and rare variants that influence common multifactorial traits[94,95]. This enables more rigorous evaluation of the contribution of rare and common variants to trait susceptibility[48] and supports the enumeration of 'allelic series' (sets of alleles of varying frequency, effect size and direction that disrupt the same gene) critical for studies of disease mechanism and therapeutic target optimization[89,96]. These developments are rapidly converging towards the ultimate destination: a comprehensive matrix of the effect of all observable genetic variants across the widest possible range of cross-sectional and longitudinal biomedical phenotypes. Success in this endeavour depends on ever greater harmonization between, and integration of results from, individual studies through sustained investments in data sharing.

Adding function

From the first linkage maps to whole-genome sequencing of large cohorts, human genetics has deployed increasingly sophisticated and inherently systematic approaches for mapping the genetic factors that underlie traits and diseases. However, progress in determining how these variants influence disease, through systematic interrogation of their functional effects on molecular, cellular and physiological processes, has been far slower.

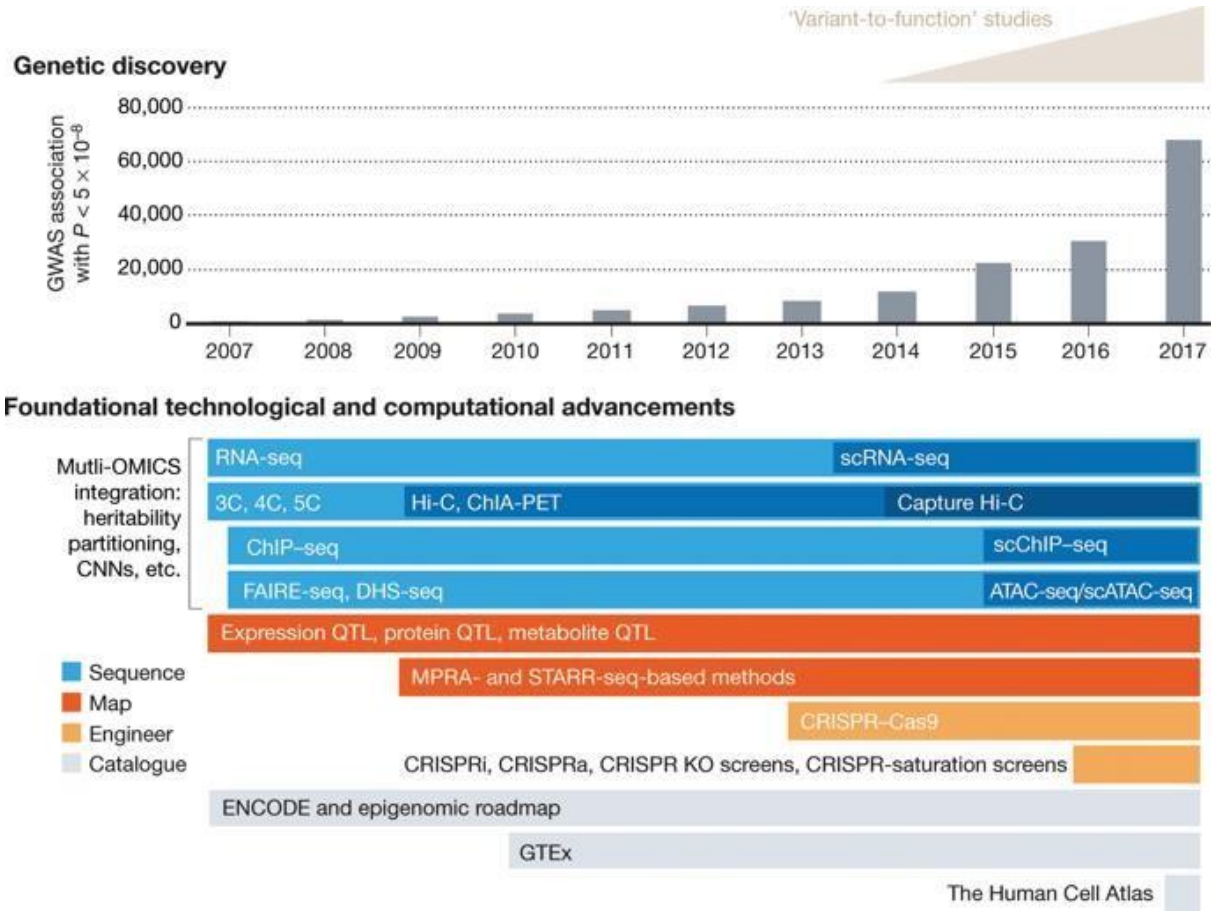
For monogenic diseases, for which the alleles responsible are typically rare, penetrant and coding, genetic approaches have generally been both necessary and sufficient to implicate a gene as causal[28]. However, as efforts to elucidate the genetic basis of Mendelian disorders progress towards completion[97], functional studies remain important to understand the mechanisms by which disruptive variation within a causal gene leads to disease phenotypes. Unlike common diseases, the clarity of causation for Mendelian disorders usually simplifies the task of generating models (including human cells and organoids or rodents) to connect genotype to organismal phenotype; these have led to many critical insights into the biology of health and disease in humans[98,99]. In addition, for genes harbouring variants with medically actionable consequences (as with the *BRCA1* and *BRCA2* mutations that are causal for early- onset breast and ovarian cancer), functional studies can support the translational

interpretation of novel alleles identified by medical sequencing29. For common diseases, functional studies have a more fundamental role. Although tens of thousands of associations have been discovered across thousands of common human diseases and traits44, multiple factors have frustrated efforts to convert these genetic signals to knowledge about causal variants, genes and mechanisms. For the common variants that underlie the bulk of complex-disease risk, the resolution of association mapping is often limited by the haplotype structure of the human genome2,3,4. Furthermore, most GWAS associations map to the noncoding genome and thus lack a direct address to the gene that mediates their effects. Growing appreciation of the pervasive role of pleiotropy complicates matters: many variants identified by GWAS are associated with multiple traits and exert diverse effects across multiple cell types100.

To date, relatively few studies have achieved the goal of connecting variants causal for complex traits to the molecular and cellular functions that mediate that predisposition. One early success described how regulatory variants that modulate *SORT1* expression influence low-density lipoprotein cholesterol and myocardial infarction risk101. More recent examples have focused on the relationship between obesity-associated variants intronic to *FTO*, altered expression of *IRX3* and *IRX5*, and adipocyte102 and hypothalamic103 function. Similar functional descriptions have been reported for individual loci implicated in schizophrenia64, cardiovascular disease104, type 2 diabetes105 and Alzheimer's disease106, among others.

Over the past decade, the challenge for the functional genomics community has been to convert this 'one-locus-at-a-time' workflow to a systematic, multidimensional, integrative approach able to deliver genome-scale functional analyses to match genome-wide variant discovery (Fig. 2). At the molecular level, one cornerstone has been generation of genome- wide catalogues of functional activity. For example, the ENCODE and Roadmap Epigenomics projects have generated maps of histone modifications, transcription-factor binding, chromatin accessibility, three-dimensional genome structure and other regulatory annotations across hundreds of cell types and tissues 107,108. The patterns of genomic overlap between these data and GWAS results enable the functional inference of risk variants, deliver clues to the specific cell types driving disease pathogenesis60,109 and accelerate locus-specific mechanistic insights.

Fig. 2: Genetic discovery is paralleled by advances in functional genomics technologies.



Top, the growth in the number of genetic loci associated by GWAS with human traits and diseases (bars) and of variant-to-function studies (area under line, not to scale). Bottom, foundational technological and computational advances over the last decade that enabled (1) development of systematic, genome-wide catalogues of functional elements across multiple cell types and tissues (blue); (2) mapping of QTLs in the context of gene expression, metabolites, proteins and regulatory elements (red); (3) engineering of genes, genetic elements and genetic variation at increasing scale (orange); and (4) systematic tissue-specific surveys of regulatory elements and transcription (grey). scRNA-seq, single-cell RNA- sequencing analysis; ChIA-PET, chromatin interaction analysis by paired-end tag sequencing; ChIP–seq, chromatin immunoprecipitation followed by sequencing; FAIRE- seq, formaldehyde-assisted isolation of regulatory elements with sequencing; DHS-seq, DNase I-hypersensitive sites sequencing; ATAC-seq, assay for transposase-accessible chromatin using sequencing; MPRA, massively parallel reporter assay; STARR-seq, self-

transcribing active regulatory region sequencing; CNN: convolutional neural networks. For further details and primary literature on many of these assays, see ref. 173.

In parallel, there has been a scaling of efforts to connect trait-associated regulatory variants to the genes and processes that they regulate in cell types relevant to the disease of interest[110,111]. For example, the GTEx (Genotype-Tissue Expression) consortium has mapped thousands of expression quantitative trait loci (QTLs) across hundreds of individuals and dozens of tissues[112]. Further clues to the relationships between regulatory variants and their effector genes can be gathered from DNA proximity assays (such as Hi-C) and single- cell data[113] (Fig. 2). Programs such as HubMAP[114] and the Human Cell Atlas[115] are set to deliver comprehensive, high-resolution reference maps of individual human cell types across diverse developmental stages, providing new opportunities to understand how regulatory genetic variation results in cellular and organismal phenotypes.

Efforts to probe the clinical consequences of coding alleles with large phenotypic effects (particularly null alleles) in humans[53,54] and across diverse animal models[116] represent powerful strategies for extending functional analyses to the whole-body level. Connections between genetic variation and circulating proteomic and metabolomic data provide additional mechanistic links between cellular events and whole-body physiology[79,80]. These efforts are paralleled by PheWAS approaches[83], which, by mapping variant effects across the range of traits available in biobanks and EMRs, can inform priors for cell types and pathways at individual loci. Importantly, whereas early studies typically linked GWAS risk alleles to data from a single functional assay, the focus is increasingly on maximizing biological insight through the multi-dimensional integration of multiple genome-wide data types using approaches such as heritability partitioning[117], functional enrichment analyses[60,109], integration of the three-dimensional genome structure[118] and deep convolutional neural networks[119,120].

Although QTL analyses can implicate a haplotype in a molecular, cellular or organismal phenotype, they are, in isolation, insufficient to define the specific causal variants responsible. To address this, there has been rapid maturation of technologies, such as massively parallel reporter assays[121,122,123] and CRISPR genome editing, to support functional characterization of targeted

sequence perturbations at scale. Variations on these methods enable the functional evaluation of genes (via knockout screens124), regulatory elements (using CRISPR interference and CRISPR activation screens125,126), and genetic variants (base editors127) at increasing scale and resolution29. Combined with complex readouts—including high-content imaging128 and single-cell transcriptomics and epigenomics129,130—these methods can generate empirical 'truth' data, supporting the development of in silico models to predict causal variants, effector transcripts126 and cellular effects. In due course, such models should reduce the need for exhaustive experimental characterization of function for all variants across all cell types.

The goal of such efforts is to enumerate the cascade of molecular events that underlie observed genotype–phenotype associations using physiologically relevant cellular systems (from primary cells to organoids and 'organ-on-chip' designs) and whole-body assays appropriate to the disease of interest. Collectively, strategies that offer large-scale functional evaluation of variants and genes of interest will reduce (but probably not eliminate) the intensive effort required for 'final mile' validation of disease mechanisms in dedicated systems, thereby accelerating downstream translational application.

Clinical implementation

Medical genetics, as applied to rare diseases, has been characterized by the rapid application in the clinic of the transformative genomic technologies that drove initial research discoveries. There are now targeted genetic tests for nearly all clinical presentations attributable to large-impact alleles, alongside more extensive genome-sequencing assays that, when necessary, enable interrogation of a longer list of relevant genes. Genetic testing for symptomatic individuals and at-risk relatives occurs routinely in many medical specialties. In parallel, the use of somatic cancer testing has increased as therapies targeted to specific mutational events have entered clinical practice (these developments are reviewed elsewhere131,132).

For patients with symptoms that indicate a probable monogenic aetiology (such as retinal degeneration, hearing loss or cardiomyopathy), targeted panels are typically the platform of choice133, although they are increasingly performed on a more extensive sequence backbone. For more complex phenotypes—those without a clear match to a specific syndrome, such as neurodevelopmental

disorders and multiple congenital anomalies—testing has gravitated towards early deployment of exome and genome-sequencing platforms that offer speedy resolution of what has historically often been a traumatic diagnostic odyssey15,134. The power of genomic diagnosis is especially clear for those presenting with monogenic neurodevelopmental disorders and critically ill infants135,136. Sequencing of the parent–offspring trio can detect de novo variation in dominant disorders and phase biallelic rare variants in recessive disease13. The transition from targeted gene tests to genomic sequencing enables recursive reanalysis, including reinterpretation of individual sequences on the basis of subsequent discoveries regarding causal disease alleles and their phenotypic consequences137. However, improved molecular diagnostics are required to ensure reliable detection of a subset of genetic disorders, including those arising from triplet repeats and complex rearrangements138. Deep sequencing of affected tissues for mosaic variants and the use of RNA sequencing to detect noncoding variants that drive early-onset disease (for example, through effects on splicing) represent new fronts for clinical diagnostics30.

Other examples of the rapid adoption of new genomic technologies include noninvasive prenatal testing (more than ten million tests by 2018 across multiple countries139,140,141) and the use of recessive carrier panels for couples planning pregnancies. Newborn screening is now universal in many countries, although it is limited to disorders combining high- throughput low-cost detection with effective early interventions (such as diet restrictions or enzyme replacement)142. Genetic diagnostics are also increasingly applied to newborn screening as a reflex test following an abnormal (for example, metabolic) screening test143. Over the next decade, the repertoire of disorders captured by neonatal screening and prenatal testing is likely to expand markedly. Whereas prenatal testing may be more effective at avoiding disease, the associated ethical issues are more complex144.

Although genetic testing for rare disease and cancer has exploded, there has been more limited uptake of genetic information in other aspects of healthcare. For example, despite multiple examples of clinically important genetic markers related to drug efficacy and side- effect profile145, the roll-out of pharmacogenetics has been hampered by a range of factors, including lack of clinical decision support in electronic medical systems to guide the drug choice

or dosing by the physician. This has been compounded by challenges in diagnostic testing: complex haplotype structures and structural variants at some key drug metabolism loci necessitate genome sequencing or specific targeted panels to detect all clinically relevant variants.

For common diseases, translational attention is currently focused on the clinical potential of polygenic risk scores. The development of robust polygenic scores for several common diseases has been catalysed by more precise per-variant effect estimates from larger GWAS datasets, improved algorithms for combining information across millions of single-nucleotide polymorphisms, and large-scale biobanks that support score validation[69,146,147]. For example, a genome-wide polygenic score for heart attack, incorporating 6.6 million variants, indicates that 5% of European-descent individuals have a risk of future cardiac events equivalent to that seen in those with less frequent monogenic forms of hypercholesterolaemia[69]. Increasingly, the shift from array-based genotyping to sequence- based analysis is facilitating risk prediction, which integrates information from rare, large- effect alleles with that from polygenic scores[93]. By improving the capture of genetic risk, particularly in non-European populations, and integrating environmental and biomarker data to quantify aspects of non-genetic risk, it should be possible to achieve increasingly accurate prediction of individual disease risk, and to use this information to tailor screening, prevention and treatment. Success will depend on developing models of risk that robustly integrate these diverse data types and on optimizing the strategies deployed to ensure effective implementation.

The absence of evidence-based guidelines to support healthcare recommendations continues to hinder the clinical applications of genetic data. In some countries, this is compounded by confusion over reimbursement and disparities in testing across society[148]. Many healthcare professionals lack experience in genomic medicine and need education and guidance to practice in the rapidly evolving space of genetic and genomic testing[149]. One consequence of these difficulties has been an expanding direct-to-consumer testing market, variably controlled by country-specific regulations[150], which is moving beyond a focus on ancestry and personal traits, towards models in which individuals have direct access to ordering physicians and genetic counselors[151]. The risk of commercial influence in this model remains high. There are concerns about the consequences of unfettered release of genetic data of dubious or

inflated clinical relevance, and limited infrastructure to pull these results into mainstream medical systems.

These advances have fostered debate about the value of genetics for population screening, for both monogenic and complex disorders. Population screening for monogenic disorders is most likely to be initiated for conditions for which risk estimates are well-understood and there are actionable interventions (for example, Lynch syndrome and familial hypercholesterolaemia). Expansion to other disorders requires better understanding of the penetrance of pathogenic alleles in unselected populations152 and caution before extending screening to longer lists of genes that are less securely implicated in disease causation153. As certain countries consider universal capture of genome-wide genetic data at birth or later in life, key questions concern the strategies for releasing this information to citizens and their medical teams to support individual healthcare.

Ultimately, barriers to genomic medicine are most directly overcome by demonstrating clinical utility in disease management and therapeutic decision-making, with evidence for improved patient outcomes. Hereditary cancers provide multiple examples, such as the use of *BRCA1/BRCA2* testing to inform PARP inhibitor treatment in patients with cancer154. (some examples are listed in Table 1), and there are currently more than 50 FDA-approved drugs for genetic disorders155. Although gene therapy has been slow to evolve since its early introduction, recent advances in gene editing are reinvigorating approaches to treat disorders by manipulation of the underlying genetic defects156.

Looking forward

Over the coming decade, the challenge will be to optimize and to implement at scale, strategies that use human genetics to further the understanding of health and disease, and to maximize the clinical benefit of those discoveries. Realizing these goals will require the concerted effort of researchers in academia and industry to bring about transformational change across a range of highly interconnected domains, for example, through the auspices of the recently established International Common Disease Alliance (https://www.icda.bio). Such efforts will be directed towards establishing: (a) comprehensive inventories of genotype– phenotype relationships across populations and environments; (b) systematic assays of variant- and gene-level function across

cell types, states and exposures; (c) improved scalable strategies for turning this basic knowledge into fully developed molecular, cellular and physiological models of disease pathogenesis; and (d) application of those biological insights to drive novel preventative and therapeutic options.

The first of these will involve documenting the full spectrum of natural genetic variation across all human populations, including capture of structural variants, and somatic mutations that accumulate with aging157,158, and associating these variations with the ever-richer disease-related intermediate and clinical traits available through biobanks and electronic health records. It will be particularly important to include populations historically under- represented in genomic research, following the pioneering work of the H3Africa consortium159. As over time, clinically sequenced genomes will outnumber those collected in academia, research and healthcare communities will need to develop a harmonized approach to genomics to transcend historical boundaries. Progress will be critically dependent
on platforms and governance that lower barriers to the integration of genetic and phenotypic data across studies and countries, along with technical standards that are reliable, secure and compatible with the international regulatory landscape160.

Mechanistic interpretation of genetic associations, particularly those in regulatory regions, will be driven by the systematic annotation of sequence variants and genes for functional impact across disease-relevant cell types, enabling mapping of processes contributing to disease development with respect to place (tissue and cell type), time (developmental stage) and context (external influences)161. Accelerating efforts to characterize the cellular composition of tissues through single-cell assays115 will increase the granularity of these observations. Large-scale perturbation studies across diverse cellular and animal models will, together with analyses of coding variants in humans53,54, provide confidence in causal inference. Large-scale proteomic and metabolomic analyses (in tissues and biological fluids) will provide a bridge to downstream pathways79,80. Research access to such functional data, generated at scale, should lower the barriers to mechanistic inference, provide system-wide context and enable researchers to focus wet-laboratory validation on the most critical experiments. Collectively, these efforts will support compilation of a systematic catalogue of key networks and processes that influence normal physiology and

disease development and inform a revised molecular taxonomy of disease.

This knowledge will reinforce the essential contribution of human genetics to the identification and prioritization of targets for therapeutic development[89,162]. Insights into the efficacy of target perturbation and potential for adverse events, allied to characterization of translatable biomarkers, provide ways to boost the efficiency of drug-development pipelines[162]. Given the clinical importance of slowing disease progression[163], target- discovery efforts will increasingly need to embrace the genetics of disease progression and treatment response, as these may involve processes distinct from those captured by studies of disease onset.

In parallel, the clinical use of human genetics will benefit from progress towards universal determination of individual genome sequences built through a combination of biobank expansion and direct access within healthcare systems. This will power clinical applications that extend beyond the current focus on neonatal sequencing, Mendelian diagnostics and somatic tumour sequencing[164]. In particular, improvements in polygenic score derivation will boost risk prediction for multifactorial traits, provide a molecular basis for disease classification, support biomarker discovery and therapeutic optimization and contribute to understanding of the variable penetrance of monogenic conditions[69]. Implementing genomic medicine as a routine component of clinical care across diverse healthcare environments will inevitably require investment in the training of healthcare professionals and attention to optimal strategies for returning genetic findings to patients.

The limited heritability of many multifactorial traits constrains the clinical precision available from genetic data alone. This will drive efforts to integrate information on personal environment, lifestyle and behaviour, and to combine prognostic, predictive information on disease risk with longitudinal measures of molecular and clinical state that track an individual's journey from health to disease. Human genetics will also, given its unique potential for causal inference, support identification of the non-genetic risk factors (often modifiable) that directly contribute to disease predisposition and development[165]. As polygenic score performance improves, analysis of individuals who show marked divergence between genetic predisposition and real-world clinical outcomes should define exposures (such as lifestyle choices

or gut microbiome) the contribution of which to disease causation remains unclear166.

Collectively, these developments can be expected to accelerate personalization of healthcare delivery. Provided costs are sustainable, a more preventative perspective on health could emerge, managed through proactive genomic, clinical and lifestyle surveillance using risk scores, complex biomarkers, liquid biopsies and wearables. Improved understanding of aetiological heterogeneity, patterns of sharing of genetic risk across diseases, variation in therapeutic response and risk of adverse events will enhance targeting of preventative and therapeutic interventions167. At the population level, intervention strategies will seek to combine population-wide and targeted strategies to best effect168. It will be critical to ensure that these benefits are available to as many as possible, so that genomics reduces, rather than exacerbates, national and global health disparities55,169 (Box 1).

The developments described above, represent variations on the theme of 'reading' the genome. The emerging capacity to block this reading (for example, through siRNA therapies170) or even to 'write' the genome (through CRISPR editing) promises to be equally transformative, providing new opportunities to correct, and even cure, Mendelian disease. Spectacular advances in developing novel therapeutic strategies are likely for many diseases, based, for example, on ex vivo cellular manipulation171 or in vivo somatic cell editing172.

Importantly, developments in genomic medicine need to proceed in a bioethical framework for research and clinical use that recognizes the personal relevance of human genetics and the critical importance of autonomous consent and the protection of privacy, while minimizing the adverse consequences of genetic exceptionalism. Governance needs to reaffirm the rights of citizens to make individual contributions to scientific progress through research participation and encourage the responsible exchange of data for clinical and research purposes.

Box 1 Global genomics

Present and future advances in genetics and genomics have the potential to provide benefits to individuals and societies across the world, but equitable and fair access to those benefits will require proactive measures to address

entrenched disparities in scientific capacity and clinical opportunities. This includes:

Future prospects

Over the past two decades, understanding of the genetic basis of human disease has been transformed by a combination of spectacular technological and analytical advances, collaborative commitment to the development of foundational resources and the collection and analysis of vast amounts of genetic, molecular and clinical data. The biological insights derived from these data are, increasingly, drivers of translational innovation, and widening personal access to large-scale genetic and molecular data promises to reshape medical care.

However, for the full potential of genomic medicine to be realized, there will need to be sustained collaborative endeavour on several fronts to ensure that the capacity to generate ever more detailed maps of the relationships between sequence variation and biomedical phenotypes delivers a comprehensive understanding of disease mechanisms that can be translated into the medicines of tomorrow.
Genome organization and function,

Genome sizes and corresponding composition of six major model organisms as pie charts. The increase in genome size correlates with the vast expansion of noncoding (i.e., intronic, intergenic, and interspersed repeat sequences) and repeat DNA (e.g., satellite, LINEs, Short interspersed nuclear element (SINEs), DNA (Alu sequence), in red) sequences in more complex muiticellular organisms. This expansion is accompanied by an increase in the number of epigenetic mechanisms (particularly repressive) that regulate the genome. Expansion of the genome also correlates with an increase in size and complexity of transcription units, with the exception of plants. P = Promoter DNA element.

The hereditary material i.e. DNA(deoxyribonucleic acid) of an organism is composed of a sequence of four nucleotides in a specific pattern, which encode information as a function of their order. Genomic organization refers to the linear order of DNA elements and their division into chromosomes. "Genome organization" can also refer to the 3D structure of chromosomes and the positioning of DNA sequences within the nucleus.

Description

Organisms have a vast array of ways in which their respective genomes are organized. A comparison of the genomic organization of six major model organisms shows size expansion with the increase of complexity of the organism. There is a more than 300-fold difference between the genome sizes of yeast and mammals, but only a modest 4- to 5-fold increase in overall gene number (see the figure on the right). However, the ratio of coding to noncoding and repetitive sequences is indicative of the complexity of the genome: The largely "open" genomes of unicellular fungi have relatively little noncoding DNA compared with the highly heterochromatic genomes of multicellular organisms.[citation needed]

In particular, mammals have accumulated considerable repetitive elements and noncoding regions, which account for the majority of their DNA sequences (52% non-coding and 44% repetitive DNA).[1][2] Only 1.2% of the mammalian genome thus encodes for protein function. This massive expansion of repetitive and noncoding sequences in multicellular organisms is most likely due to the incorporation of invasive elements, such as DNA transposons, retrotransposons, and other repetitive elements.[3] The expansion of repetitive elements (such as Alu sequences) has even infiltrated the transcriptional units of the mammalian genome. This results in transcription units that are frequently much larger (30-200 kb), commonly containing multiple promoters and DNA repeats within untranslated introns.[citation needed]
The vast expansion of the genome with noncoding and repetitive DNA in higher eukaryotes implies more extensive epigenetic silencing mechanisms. Studies of the genomic organization is thought to be the future of genomic medicine, which will provide the opportunity for personalized prognoses in clinics.[4]

Genome Function

The term genome was created in 1920 by Hans Winkler,[3] professor of botany at the University of Hamburg, Germany. The Oxford Dictionary suggests the name is a blend of the words gene and chromosome. [4] However, see omics for a more thorough discussion. A few related -ome words already existed, such as biome and rhizome, forming a vocabulary into which genome fits systematically.[5]

Sequencing and mapping

Further information: Genome project

A genome sequence is the complete list of the nucleotides (A, C, G, and T for DNA genomes) that make up all the chromosomes of an individual or a species. Within a species, the vast majority of nucleotides are identical between individuals, but sequencing multiple individuals is necessary to understand the genetic diversity. Part of DNA sequence - prototypification of complete genome of virus

In 1976, Walter Fiers at the University of Ghent (Belgium) was the first to establish the complete nucleotide sequence of a viral RNA-genome (Bacteriophage MS2). The next year, Fred Sanger completed the first DNA-genome sequence: Phage Φ-X174, of 5386 base pairs. [6] The first complete genome sequences among all three domains of life were released within a short period during the mid-1990s: The first bacterial genome to be sequenced was that of Haemophilus influenzae, completed by a team at The Institute for Genomic Research in 1995. A few months later, the first eukaryotic genome was completed, with sequences of the 16 chromosomes of budding yeast Saccharomyces cerevisiae published as the result of a European-led effort begun in the mid-1980s. The first genome sequence for an archaeon, Methanococcusjannaschii, was completed in 1996, again by The Institute for Genomic Research.

The development of new technologies has made genome sequencing dramatically cheaper and easier, and the number of complete genome sequences is growing rapidly. The US National Institutes of Health maintains one of several comprehensive databases of genomic information.[7] Among the thousands of completed genome sequencing projects include those for rice, a mouse, the plant Arabidopsis thaliana, the puffer fish, and the bacteria E. coli. In December 2013, scientists first sequenced the entire genome of a Neanderthal, an extinct species of humans. The genome was extracted from the toe bone of a 130,000-year- old Neanderthal found in a Siberian cave.[8][9]. New sequencing technologies, such as massive parallel sequencing have also opened up the prospect of personal genome sequencing as a diagnostic tool, as

pioneered by Manteia Predictive Medicine. A major step toward that goal was the completion in 2007 of the full genome of James D. Watson, one of the co-discoverers of the structure of DNA.[10]

Whereas a genome sequence lists the order of every DNA base in a genome, a genome map identifies the landmarks. A genome map is less detailed than a genome sequence and aids in navigating around the genome. The Human Genome Project was organized to map and to sequence the human genome. A fundamental step in the project was the release of a detailed genomic map by Jean Weissenbach and his team at the Genoscope in Paris.[11][12]

Reference genome sequences and maps continue to be updated, removing errors and clarifying regions of high allelic complexity.[13] The decreasing cost of genomic mapping has permitted genealogical sites to offer it as a service,[14] to the extent that one may submit one's genome to crowdsourced scientific endeavours such as DNA.LAND at the New York Genome Center,[15] an example both of the economies of scale and of citizen science.[16]

Viral genomes

Viral genomes can be composed of either RNA or DNA. The genomes of RNA viruses can be either single-stranded or double-stranded RNA, and may contain one or more separate RNA molecules (segments: monopartit or multipartit genome). DNA viruses can have either single-stranded or double-stranded genomes. Most DNA virus genomes are composed of a single, linear molecule of DNA, but some are made up of a circular DNA molecule.[17]

Prokaryotic genomes

Prokaryotes and eukaryotes have DNA genomes. Archaea have a single circular chromosome.[18] Most bacteria also have a single circular chromosome; however, some bacterial species have linear chromosomes[19] or multiple chromosomes.[20] If the DNA is replicated faster than the bacterial cells divide, multiple copies of the chromosome can be present in a single cell, and if the cells divide faster than the DNA can be replicated, multiple replication of the chromosome is initiated before the division occurs, allowing daughter cells to inherit complete genomes and already partially replicated chromosomes. Most prokaryotes have very little repetitive DNA in their genomes .[21] However,

some symbiotic bacteria (e.g. Serratia symbiotica) have reduced genomes and a high fraction of pseudogenes: only ~40% of their DNA encodes proteins.[22][23]
Some bacteria have auxiliary genetic material, also part of their genome, which is carried in plasmids. For this, the word genome should not be used as a synonym of chromosome.

Eukaryotic genomes

Eukaryotic genomes are composed of one or more linear DNA chromosomes. The number of chromosomes varies widely from Jack jumper ants and an asexual nemotode,[24] which each have only one pair, to a fern species that has 720 pairs.[25] A typical human cell has two copies of each of 22 autosomes, one inherited from each parent, plus two sex chromosomes, making it diploid. Gametes, such as ova, sperm, spores, and pollen, are haploid, meaning they carry only one copy of each chromosome.
In addition to the chromosomes in the nucleus, organelles such as the chloroplasts and mitochondria have their own DNA. Mitochondria are sometimes said to have their own genome often referred to as the "mitochondrial genome". The DNA found within the chloroplast may be referred to as the "plastome". Like the bacteria they originated from, mitochondria and chloroplasts have a circular chromosome.
Unlike prokaryotes, eukaryotes have exon-intron organization of protein coding genes and variable amounts of repetitive DNA. In mammals and plants, the majority of the genome is composed of repetitive DNA.[26]

Coding sequences

DNA sequences that carry the instructions to make proteins are coding sequences. The proportion of the genome occupied by coding sequences varies widely. A larger genome does not necessarily contain more genes, and the proportion of non-repetitive DNA decreases along with increasing genome size in complex eukaryotes.[26]
Simple eukaryotes such as C. elegans and fruit fly, have more non-repetitive DNA than repetitive DNA,[26][27] while the genomes of more complex eukaryotes tend to be composed largely of repetitive DNA. In some plants and amphibians, the proportion of repetitive DNA is more than 80%.[26] Similarly, only 2% of the human genome codes for proteins.

Composition of the human genome Noncoding sequences

Noncoding sequences include introns, sequences for non-coding RNAs, regulatory regions, and repetitive DNA. Noncoding sequences make up 98% of the human genome. There are two categories of repetitive DNA in the genome: tandem repeats and interspersed repeats.[28]

Chromosomal Disorders

Almost every cell in our body contains 23 pairs of chromosomes, for a total of 46 chromosomes. Half of the chromosomes come from our mother, and the other half come from our father. The first 22 pairs are called autosomes. The 23rd pair consists of the sex chromosomes, X and Y. Females usually have two X chromosomes, and males usually have one X and one Y chromosome in each cell. All of the information that the body needs to grow and develop comes from the chromosomes. Each chromosome contains thousands of genes, which make proteins that direct the body's development, growth, and chemical reactions.

Many types of chromosomal abnormalities exist, but they can be categorized as either numerical or structural. Numerical abnormalities are whole chromosomes either missing from or extra to the normal pair. Structural abnormalities are when part of an individual chromosome is missing, extra, switched to another chromosome, or turned upside down. Chromosomal abnormalities can occur as an accident when the egg or the sperm is formed or during the early developmental stages of the fetus.

The age of the mother and certain Prenatal screening and testing can be performed to examine the chromosomes of the fetus and detect some, but not all, types of chromosomal abnormalities.

Chromosomal abnormalities can have many different effects, depending on the specific abnormality. For example, an extra copy of chromosome 21 causes Down syndrome (trisomy 21). Chromosomal abnormalities can also cause miscarriage, disease, or problems in growth or development.

The most common type of chromosomal abnormality is known as aneuploidy, an abnormal chromosome number due to an extra or missing chromosome.Most people with aneuploidy have trisomy (three copies of a chromosome) instead of monosomy (single copy of a chromosome). Down syndrome is probably the

most well-known example of a chromosomal aneuploidy. Besides trisomy 21, the major chromosomal aneuploidies seen in live-born babies are: trisomy 18; trisomy 13; 45, X (Turner syndrome); 47, XXY (Klinefelter syndrome); 47, XYY; and 47, XXX.

Structural chromosomal abnormalities result from breakage and incorrect rejoining of chromosomal segments. A range of structural chromosomal abnormalities result in disease. Structural rearrangements are defined as balanced if the complete chromosomal set is still present, though rearranged, and unbalanced if information is additional or missing. Unbalanced rearrangements include deletions, duplications, or insertions of a chromosomal segment. Ring chromosomes can result when a chromosome undergoes two breaks and the broken ends fuse into a circular chromosome. An isochromosome can form when an arm of the chromosome is missing and the remaining arm duplicates.

Balanced rearrangements include inverted or translocated chromosomal regions. Since the full complement of DNA material is still present, balanced chromosomal rearrangements may go undetected because they may not result in disease. A disease can arise as a result of a balanced rearrangement if the breaks in the chromosomes occur in a gene, resulting in an absent or nonfunctional protein, or if the fusion of chromosomal segments results in a hybrid of two genes, proucing a new protein product whose function is damaging to the cell.

Questions for Practice

6 marks

1. Brief note on reason for chromosomal disorders.
2. Describe the term ‘Heritabilty’
3. Explain the morphology of chromosomes.
4. Describe the term ‘Central Dogma of life’ with illustration.
5. Explain the Gene structure with schematic representations.
6. Brief note on ‘Genetic Variations’ with emphasis on ‘Allele, Polymorphism and mutation’
7. Describe the applications of Human Genetics.

10 marks

1. Elaborate the History of Human Genetics.
2. Describe the Genome organization and function.
3. Explain Chromosomal Disorders.
4. Illustrate the techniques involved in identification of Chromosomal Disorders.

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF BIO AND CHEMICAL ENGINEERING

DEPARTMENT OF CENTRE FOR MOLECULAR AND NANOMEDICAL SCIENCES

UNIT - II – PEDIGREE ANALYSIS - (SMB5202)

Unit-II

Pedigrees- gathering family history, pedigree symbols, construction of pedigrees, presentation of molecular genetic data in pedigrees Pedigree analysis describes the process of interpretation of information displayed as a family tree. The family tree or pedigree is constructed using a standardized set of symbols and will include information about the disease status of each individual. If only a single individual is affected within the family, then the pedigree cannot in itself provide proof for a particular mode of inheritance and cannot distinguish inherited from non-inherited conditions. When more than one individual is affected then the pattern may provide important clues or even proof of the mode of inheritance. There are four main patterns of inheritance that may be seen in a pedigree.

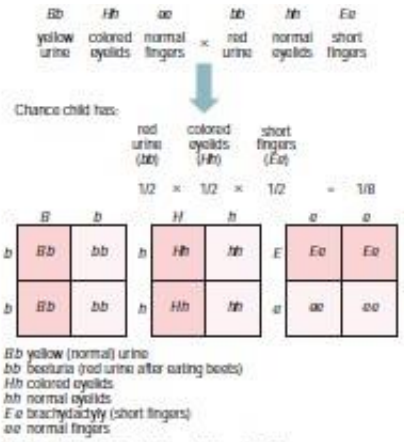


Figure 4.12 Using probability to track three traits. A man with normal urine, colored eyelids, and normal fingers wants to have children with a woman who has red urine after she eats beets, normal eyelids, and short fingers. The chance that a child of theirs will have red urine after eating beets, colored eyelids, and short fingers is 1/8.

genes at once. The increasingly computational nature of genetics in this century has produced an entirely new field called bioinformatics. So, in this sense, genetics is continuing the theme of mathematical analysis that Gregor Mendel began more than a century ago.

Key Concepts

1. Mendel's law of independent assortment considers genes transmitted on different chromosomes.
2. In a dihybrid cross of heterozygotes for seed color and shape, Mendel saw a phenotypic ratio of 9:3:3:1. He concluded that transmission of one gene does not influence that of another.
3. Meiotic events explain independent assortment.
4. Punnett squares and probability can be used to follow independent assortment.
5. Knowing the human genome sequence has made it possible to analyze more than one gene at a time.

4.4 Pedigree Analysis

For researchers, families are tools, and the bigger the family the better—the more children in a generation, the easier it is to discern modes of inheritance. Geneticists use charts called **pedigrees** to display family relationships and to depict which relatives have specific phenotypes and, sometimes, genotypes. A human pedigree serves the same purpose as one for purebred dogs or cats or thoroughbred horses—it represents relationships and traits. A pedigree in genetics differs from a family tree in genealogy, and from a genogram in social work, in that it indicates disorders or traits as well as relationships. Pedigrees may also include molecular data, test results, and haplotypes (genes linked in segments on a chromosome).

A pedigree consists of lines that connect shapes. Vertical lines represent generations; horizontal lines that connect two shapes at their centers depict partners; shapes connected by vertical lines that are joined horizontally represent siblings. Squares indicate males; circles, females; and diamonds, individuals of unspecified sex. Roman numerals designate generations. Arabic numerals or names indicate individuals. **Figure 4.13** shows these and other commonly used pedigree symbols. Colored or shaded shapes indicate individuals who express a trait, and half-filled shapes are known

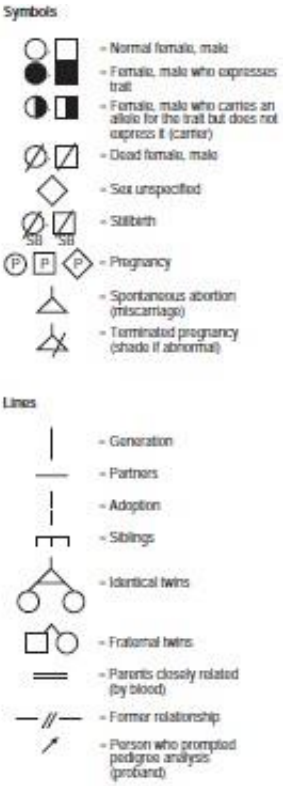


Figure 4.13 Pedigree components. Symbols representing individuals are connected to form pedigree charts, which display the inheritance patterns of particular traits.

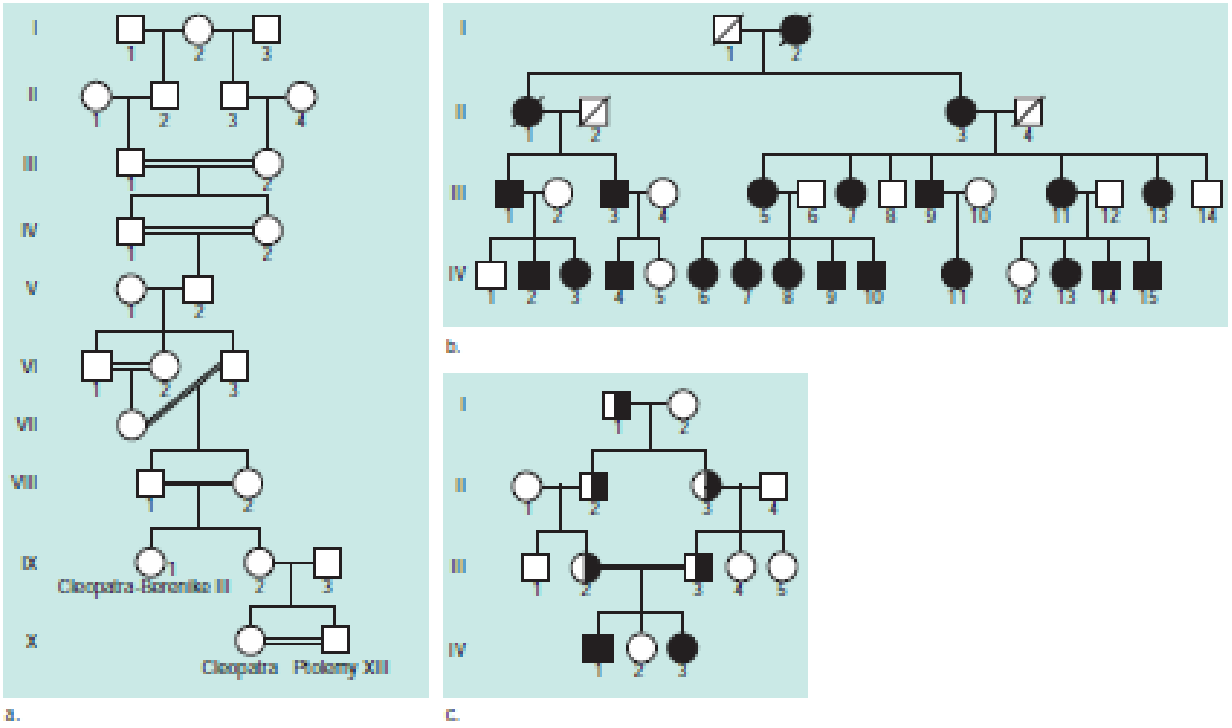


Figure 4.14 Some unusual pedigrees. (a) A partial pedigree of Egypt's Ptolemy dynasty shows only genealogy, not traits. It appears almost laddarlike because of the extensive inbreeding. From 323 B.C. to Cleopatra's death in 30 B.C., the family experienced one pairing between cousins related through half-brothers (generation III), four brother-sister pairings (generations IV, VI, VIII, and X), and an uncle-niece relationship (generations VI and VII). Cleopatra married her brother, Ptolemy XIII, when he was 10 years old. These marriage patterns were an attempt to preserve the royal blood. (b) In contrast to the Egyptian pedigree, a family with polydactyly (extra fingers and toes) extends laterally, with many children. (c) The most common form of consanguinity is marriage of first cousins. They share one set of grandparents, and therefore risk passing on the same recessive alleles to offspring.

carriers. A genetic counselor will often sketch out a pedigree while interviewing a client, then use a computer program and add test results that indicate genotypes.

Pedigrees Then and Now

The earliest pedigrees were strictly genealogical, not indicating traits. **Figure 4.14** shows such a pedigree for a highly inbred part of the ancient Egyptian royal family. The term *pedigree* arose in the fifteenth century, from the French *pie de grue*, which means "crane's foot." Pedigrees at that time, typically depicting large families, showed parents linked by curved lines to their many offspring. The overall diagram often resembled a bird's foot.

One of the first pedigrees to trace an inherited illness was an extensive family tree of several European royal families, indicating which members had the clotting disorder hemophilia (see figure 6.8). The mutant gene probably originated in Queen Victoria of England in the nineteenth century. In 1845, a genealogist named Pliny Earle constructed a pedigree of a family with colorblindness using musical notation—half notes for unaffected females, quarter notes for colorblind females, and filled-in and squared-off notes to represent the many colorblind males. In the early twentieth century, eugenicists tried to use pedigrees to show that traits such as criminality, feebleмindedness, and promiscuity were the consequence of faulty genes.

Today, pedigrees are important both for helping families identify the risk of transmitting an inherited illness and as starting points for identifying a gene from the human genome sequence. People who have kept meticulous family records are invaluable in helping researchers follow the inheritance of particular genes in groups such as the Mormons and the Amish. Very large pedigrees can provide information on many individuals with a particular rare disorder. The researchers can then search affected individuals' DNA to identify a particular sequence they have all inherited that is not found in healthy family members. This is where the causative mutation lies. Discovery of the gene that causes HD, for example, took researchers to a remote village in Venezuela to study an enormous family. The gene was eventually traced to a Portuguese sailor who introduced the mutation in the nineteenth century.

Pedigrees Display Mendel's Laws

Visual learners can easily "see" a mode of inheritance in a pedigree. Consider a pedigree for an autosomal recessive trait, albinism. Homozygous recessive individuals in the third (F_2) generation lack an enzyme necessary to manufacture the pigment melanin and, as a result, hair and skin are very pale (**figure 4.15**) Their parents are inferred to be heterozygotes (carriers). One partner from each pair of grandparents must also be a carrier. Carriers can sometimes be identified using a carrier test, inferred from family history, or deduced from the DNA sequence.

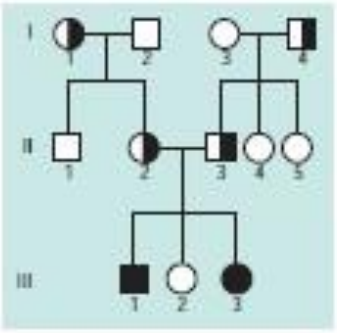


Figure 4.15 A pedigree for an autosomal recessive trait. Albinism affects males and females and can skip generations, as it does here in generations I and II. The homozygous recessive individual lacks an enzyme needed to produce melanin, which colors the eyes, skin, and hair.

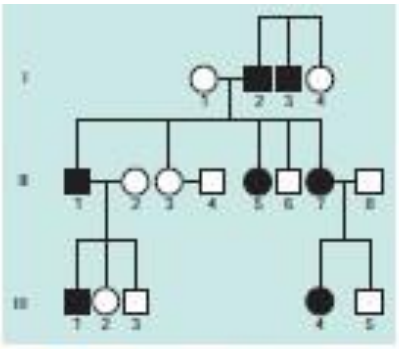


Figure 4.16 A pedigree for an autosomal dominant trait. Autosomal dominant traits do not skip generations.

An autosomal dominant trait does not skip generations and can affect both sexes. A typical pedigree for an autosomal dominant trait has some squares and circles filled in to indicate affected individuals in each generation (**figure 4.16**).

A pedigree may be inconclusive, which means that either autosomal recessive or autosomal dominant inheritance can explain the pattern of filled-in symbols. **Figure 4.17** shows one such pedigree, for a type of hair loss called alopecia areata (OMIM 104000). According to the pedigree, this trait can be passed in an autosomal dominant mode because it affects both males and females and is present in every generation. However, the pedigree can also depict autosomal recessive inheritance if the individuals represented by unfilled symbols are carriers. Inconclusive pedigrees tend to arise when families are small and the trait is not severe enough to impair fertility.

Solving a Problem: Conditional Probability

Often genetic counselors are asked to predict the probability that a condition will occur in a particular individual. Mendel's laws, pedigrees, and Punnett squares provide clues, as do logic and common sense. Consider the family depicted in **figure 4.18.**

Michael Stewart has sickle cell disease, which is autosomal recessive. His unaffected parents, Kate and Brad, must each be heterozygotes (carriers). Michael's sister,

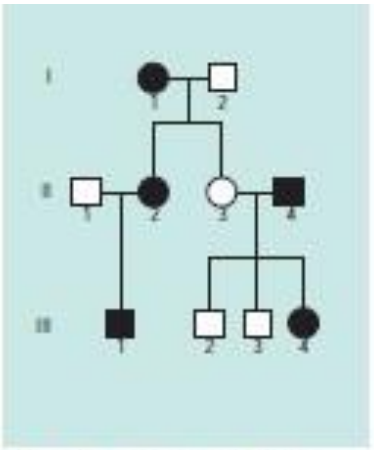
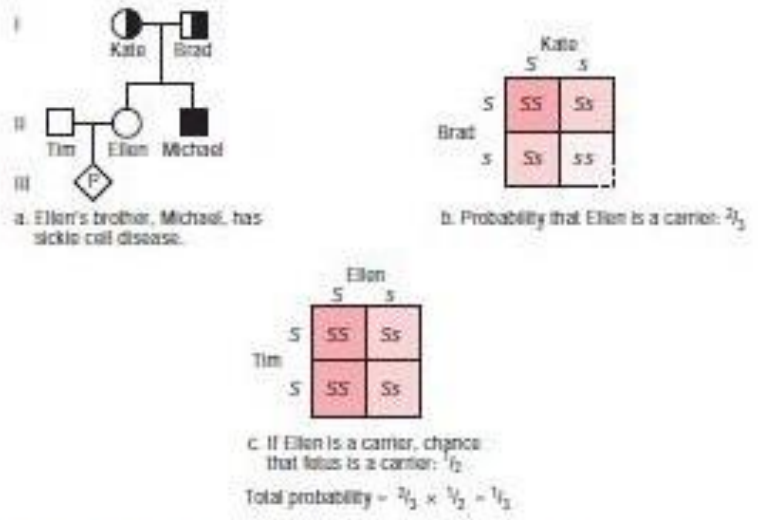


Figure 4.17 An inconclusive pedigree. This pedigree could account for an autosomal dominant trait or an autosomal recessive trait that does not prevent affected individuals from having children. (Unfilled symbols could represent carriers.)

Ellen, also healthy, is expecting her first child. Ellen's husband, Tim, has no family history of sickle cell disease. Ellen wants to know the risk that her child will inherit the mutant allele from her and be a carrier.

Ellen's request raises two questions. First, what is the risk that she herself is a carrier? Because Ellen is the product of a monohybrid cross, and we know that she is not homozygous recessive, she has a 2 in 3 chance of being a carrier, as the Punnett square indicates. If Ellen is a carrier, what is the chance that she will pass the mutant allele to an offspring? It is 1 in 2, because she has two copies of the gene, and according to Mendel's first law, only one allele goes into each gamete.

To calculate the overall risk to Ellen's child, we can apply the product rule and multiply the probability that Ellen is a carrier by the chance that, if she is, she will pass the mutant allele on. This result, following two events, is a conditional probability, because the likelihood of the second event—the child being a carrier—depends upon the first event—that Ellen is a carrier. If we assume Tim is not a carrier, Ellen's chance of giving birth to a child who carries the mutant allele is therefore 2/3 times 1/2, which equals 2/6, or 1/3. Ellen thus has a



a. Ellen's brother, Michael, has sickle cell disease.

b. Probability that Ellen is a carrier: $^2/_3$

c. If Ellen is a carrier, chance that fetus is a carrier: $^1/_2$

Total probability = $^2/_3 \times {}^1/_2 = {}^1/_3$

Figure 4.18 Making predictions. Ellen's brother, Michael, has sickle cell disease (a). Ellen wonders if her fetus has inherited the sickle cell allele from her. First, she must calculate the chance that she is a carrier. The Punnett square in (b) shows that this risk is 2 in 3. (She must be genotype *SS* or *Ss*, but cannot be *ss* because she does not have the disease.) The risk that the fetus is a carrier, assuming that the father is not a carrier, is half Ellen's risk of being a carrier, or 1 in 3 (c).

theoretical 1 in 3 chance of giving birth to a child who is a carrier for sickle cell disease.

Pedigrees work well in theory but may be difficult to construct or interpret in practice for several reasons. People sometimes hesitate to supply information because they are embarrassed by symptoms affecting behavior or mental stability. Family relationships can be complicated by adoption, children born out of wedlock, serial relationships, blended families, and assisted reproductive technologies such as surrogate mothers and intrauterine insemination by donor (see chapter 21). Moreover, many people cannot trace their families back more than three or four generations, so they lack sufficient evidence to reveal a mode of inheritance. Still, the pedigree remains a powerful way to see, at a glance, how a trait passes from generation to generation—just as Gregor Mendel did with peas.

Key Concepts

1. Pedigrees depict family relationships and the transmission of inherited traits. Squares represent males, and circles, females; horizontal lines link partners, vertical lines show generations, and elevated horizontal lines depict siblings. Heterozygote symbols are half-shaded, and symbols for individuals who express a trait are completely shaded.
2. Pedigrees can reveal modes of inheritance. Along with Punnett squares, they are tools that apply Mendel's laws to predict the recurrence risks of inherited disorders or traits.

Summary

4.1 Following the Inheritance of One Gene—Segregation

1. Gregor Mendel described the two basic laws of inheritance using pea plant crosses. The laws, which derive from the actions of chromosomes during meiosis, apply to all diploid organisms.
2. Mendel used a statistical approach to investigate why some traits seem to disappear in the hybrid generation. The **law of segregation** states that alleles of a gene are distributed into separate gametes during meiosis. Mendel demonstrated this using seven traits in pea plants.
3. A diploid individual with two identical alleles of a gene is **homozygous**. A **heterozygote** has two different alleles of a gene. A gene may have many alleles.
4. A **dominant** allele masks the expression of a **recessive** allele. An individual may be homozygous dominant, homozygous recessive, or heterozygous.
5. Mendel found that when he crossed two true-breeding types, then bred the resulting hybrids to each other, the two variants of the trait appeared in a 3:1 phenotypic ratio. Crossing these progeny further revealed a genotypic ratio of 1:2:1.
6. A **Punnett square** follows the transmission of alleles and is based on probability.

4.2 Single-Gene Inheritance in Humans

7. **Modes of inheritance** enable geneticists to predict phenotypes. In **autosomal dominant** inheritance, males and females may be affected, and the trait does not skip generations. Inheritance of an **autosomal recessive** trait may affect either males or females and may skip generations. Autosomal recessive conditions are more likely to occur in families with **consanguinity**. Recessive disorders tend to be more severe and cause symptoms earlier than dominant disorders.

Pedigree analysis and Importance of pedigree

Pedigree charts are diagrams that show the phenotypes and/or genotypes for a particular organism and its ancestors. While commonly used in human families to track genetic diseases, they can be used for any species and any inherited trait. Geneticists use a standardized set of symbols to represent an individual's sex, family relationships and phenotype. These diagrams are used to determine the **mode of inheritance** of a particular disease or trait, and to predict the probability of its appearance among offspring. Pedigree analysis is therefore an important tool in both basic research and **genetic counseling**.

Each pedigree chart represents all of the available information about the inheritance of a single trait (most often a disease) within a family. The pedigree chart is therefore drawn using factual information, but there is always some possibility of errors in this information, especially when relying on family members' recollections or even clinical diagnoses.

In real pedigrees, further complications can arise due to **incomplete penetrance** (including age of onset) and **variable expressivity** of disease alleles, but for the examples presented in this book, we will presume complete accuracy of the pedigrees. A pedigree may be drawn when trying to determine the nature of a newly discovered disease, or when an individual with a family history of a disease wants to know the probability of passing the disease on to their children. In either case, a tree is drawn, as shown in Figure 5.2, with circles to represent females, and squares to represent males. Matings are drawn as a line joining a male and female, while a consanguineous mating (closely related is two lines.

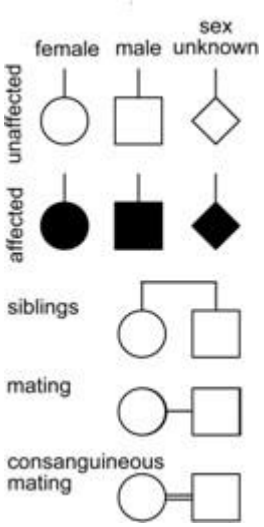


Figure 5.2: *Symbols used in drawing a pedigree. (Original-Deyholos-CC:AN)*

The affected individual that brings the family to the attention of a geneticist is called the **proband** (or propositus). If an individual is known to have symptoms of the disease (**affected**), the symbol is filled in. Sometimes a half-filled in symbol is used to indicate a known **carrier** of a disease; this is someone who does not have any symptoms of the disease, but who passed the disease on to subsequent generations because they are a heterozygote. Note that when a pedigree is constructed, it is often unknown whether a particular individual is a carrier or not, so not all carriers are always explicitly indicated in a pedigree. For simplicity, in this chapter we will assume that the pedigrees presented are accurate, and represent fully penetrant traits.

QUESTIONS FOR PRACTICE

UNIT – 2

6 marks

1. Describe the steps involved in the construction of Pedigree.
2. Define Pedigree with schematic representation.
3. Brief note on Pedigree symbols.
4. Explain the importance of Pedigree.
5. Briefly explain the steps involved in the analysis of Pedigree.
6. Elaborate the terms involved in the construction of Pedigree.

10 marks

1. Elaborate the construction of Pedigree and add notes on the application of Pedigree.
2. Explain the importance of Pedigree with two examples.
3. Neatly explain how the molecular genetic data are present in the pedigree?
4. Is it possible to identify the Autosomal, X-linked, Dominant and Recessive in Pedigree. If So, neatly sketch the pedigree for each condition mentioned above.

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF BIO AND CHEMICAL ENGINEERING

DEPARTMENT OF CENTRE FOR MOLECULAR AND NANOMEDICAL SCIENCES

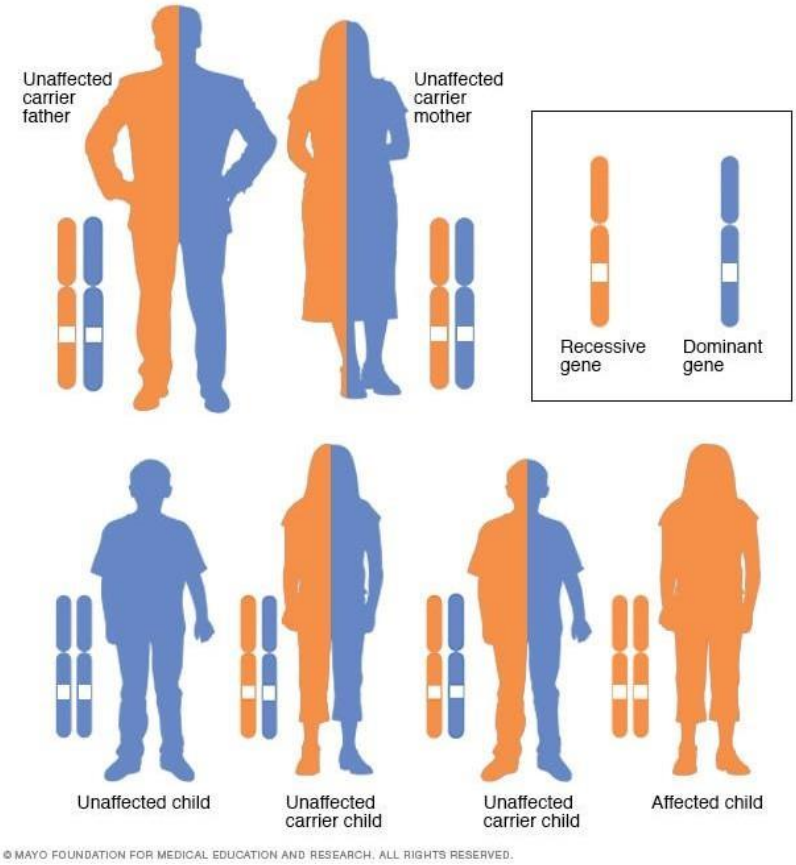
UNIT - III – MONOGENIC TRAITS - (SMB5202)

Unit-III

Monogenic traits, Autosomal inheritance-dominant, recessive. Sex-linked inheritance, Mitochondrial inheritance, genomic imprinting, spontaneous mutations, male lethality, X-inactivation, Consanguinity and its effects.

Monogenic Traits

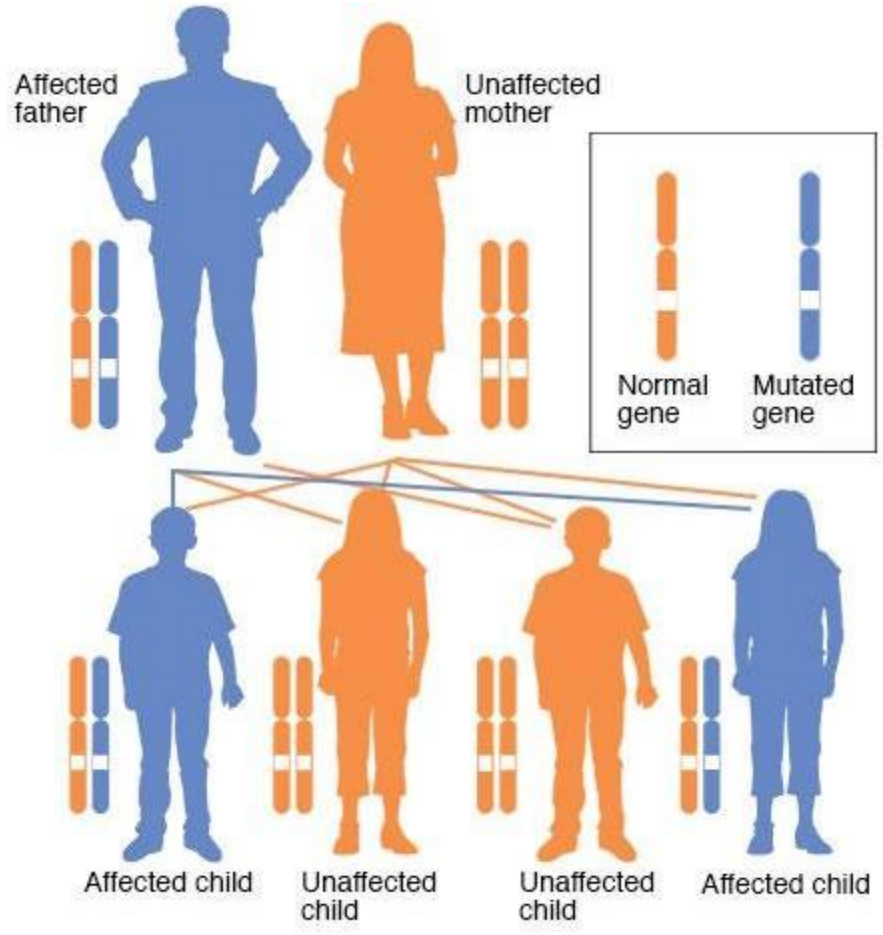
A monogenic trait is a characteristic that is produced by a single gene or a single allele. A gene is a DNA segment that is responsible for physical traits and the phenotype of an organism. An allele is one of a pair of genes on a chromosome - the pair controls the same trait.



An example of a monogenic trait is eye color or the ability to roll your tongue. A polygenic trait is a characteristic controlled by two or more genes that are located at different areas on different chromosomes.

To have an autosomal recessive disorder, you inherit two mutated genes, one

from each parent. These disorders are usually passed on by two carriers. Their health is rarely affected, but they have one mutated gene (recessive gene) and one normal gene (dominant gene) for the condition. Two carriers have a 25% chance of having an unaffected child with two normal genes (left), a 50% chance of having an unaffected child who also is a carrier (middle), and a 25% chance of having an affected child with two recessive genes (right).



In an autosomal dominant disorder, the mutated gene is a dominant gene located on one of the non-sex chromosomes (autosomes). You need only one mutated gene to be affected by this type of disorder. A person with an autosomal dominant disorder in this case, the father has a 50% chance of having an affected child with one mutated gene (dominant gene) and a 50% chance of having an unaffected child with two normal genes (recessive genes).

Sex-linked Inheritance

Inheritance because of a recessive gene found in sex chromosomes (the X chromosome of the X and Y chromosomes in humans and drosophila) is called **sex-linked inheritance**. This is primarily a trait observed in males. Sex-linked inheritance has been explained below by examining color blindness in humans (impaired ability to distinguish red and green). If an X chromosome that has a recessive gene is called an X′ chromosome, then an X′Y male and an X′X′ female will both be color blind, but an X′X female (a carrier) will be normal. Generally, color blindness occurs in a male child born to a carrier X′X female and a normal XY male. Color blindness is rare in a female child born to a color blind male and a carrier female.

Following the same pattern for sex determination, other sex chromosomes such as the XO type (mainly in insects such as locust; males have a single X chromosome and females have two X chromosomes); ZO type (loggerhead sea turtles and similar species; females have a single Z chromosome and males have two Z chromosomes); or ZW type (chickens and similar species; females have Z and W chromosomes and males have two Z chromosomes) are found in organisms.

Mitochondrial inheritance,

Mitochondrial inheritance: The inheritance of a trait encoded in the mitochondrial genome. Because of the oddities of mitochondria, mitochondrial inheritance does not obey the classic rules of genetics. Persons with a mitochondrial disease may be male or female but they are always related in the maternal line and no male with the disease can transmit it to his children.

The mitochondria are normal structures or organelles in cells. They are located in the cell's cytoplasm outside the nucleus.

The mitochondria are responsible for energy production. They consist of two sets of membranes, a smooth continuous outer coat and an inner membrane

arranged in tubules or in folds that form plate- like double membranes (cristae). The mitochondria are in fact the principal energy source of the cell (thanks to the cytochrome enzymes of terminal electron transport and the enzymes of the citric acid cycle, fatty acid oxidation, and oxidative phosphorylation). The mitochondria convert nutrients into energy as well as doing many other specialized tasks.

Each mitochondrion has a chromosome that is made of DNA (mitochondrial DNA or mtDNA) but is otherwise quite different from the better known chromosomes in the nucleus. The mitochondrial chromosome is much smaller. It is round (whereas the chromosomes in the nucleus are shaped like rods). There are many copies of the mitochondrial chromosome in every cell (whereas there is normally only one set of chromosomes in the nucleus). Mitochondrial DNA contains 37 genes which all are essential for normal function of the mitochondria. Many genetic conditions are related to changes in particular mitochondrial genes.

Genomic imprinting,

Even though both parents contribute equally to the genetic content of their offspring, a developmental process called genomic imprinting sometimes leads to the exclusive expression of specific genes from only one parent. This process was first described in 1984, when two laboratories discovered a mark, or "imprint," that differentiates between certain genes on the maternal and paternal chromosomes and results in the expression of only one copy of those genes in the offspring. The genes in imprinted areas of an organism's genome are expressed depending on the parent of origin. As a result, the inheritance of both the maternal and paternal genes is required for normal development to proceed (McGrath &Solter, 1984; Surani *et al.*, 1984).

To understand how imprinting works, rather than looking at the entire genome, consider the effect of this process on smaller chromosomal regions and single genes. For many diploid genes, even if the copy you inherited from one parent is defective, you have a substitute allele from your other parent. However, in the case of imprinting, even though there are two copies of the gene, it is as if you are haploid for this gene because only one copy is expressed.

In other words, there is no substitute allele, which makes imprinted genes more vulnerable to the negative effects of mutations. Additionally, genes and mutations that might normally be recessive can be expressed if a gene is imprinted and the dominant allele is silenced (Jirtle& Weidman, 2007).

Diseases Related to Imprinting

As you might expect, it is therefore possible for diseases to occur due to deletions or mutations in imprinted genes. Diseases can also result from uniparental disomy, or the inheritance of two copies of a chromosome from one parent and no copy from the other parent, when the involved gene is imprinted. Additionally, diseases are possible when there are mutations in genes responsible for the imprinting process and when the imprint is not set correctly. Some examples of genetic diseases related to errors in the imprinting of specific genes and chromosomal regions include Prader-Willi syndrome, Angelman syndrome, and several types of cancer.

Prader-Willi Syndrome and Angelman Syndrome

Prader-Willi syndrome was first described by John Langdon Down (who also identified Down syndrome) in 1887, and later reported by Andrea Prader, Alexis Labhart, and Heinrich Willi in 1956. This disorder occurs in approximately one in 20,000 births and is associated with behavioral and cognitive problems, including mental retardation, deficiencies in sexual development and growth, hyperphagia, and obesity (Prader *et al.*, 1956; Falls *et al.*, 1999). In 1965, Dr. Harry Angelman was the first to report the symptoms of Angelman syndrome. The disorder occurs in approximately one in 15,000 births, and the syndrome is characterized by developmental deficiencies, mental retardation, sleep disorders, seizures, ataxia, hyperactivity, and a happy disposition with outbursts of laughter (Angelman, 1965; Falls *et al.*, 1999).

Prader-Willi syndrome and Angelman syndrome were the first imprinting diseases discovered in humans. The symptoms of these two disorders are very different, but scientists discovered that both conditions are caused by indistinguishable deletions in chromosome 15 in the 15q11-q13 region (Knoll *et al.*, 1989). What distinguishes these disorders is the parental origin of the affected chromosome. Specifically, Prader-Willi syndrome is caused by the loss

of a group of paternally inherited genes on chromosome 15 (Butler *et al.*, 1986; Nicholls *et al.*, 1989a, 1989b). In contrast, Angelman syndrome is caused by the loss of a maternally inherited gene in the same region of chromosome 15. Researcher Joan Knoll and her colleagues thus concluded that both Prader-Willi syndrome and Angelman syndrome were due to defects in imprinted genes, as the imprinted region of chromosome 15 normally contains genes that are either paternally or maternally expressed (Knoll *et al.*, 1989).

In most cases (60%-70%), Prader-Willi syndrome is caused by deletions of a genetic region that includes the small nuclear ribonucleoprotein polypeptide N gene, the necdin gene, and possibly other genes (Robertson, 2005). In the remaining 20%-30% of Prader-Willi patients, the disorder occurs because the affected individual has two copies of maternal chromosome 15 and no copy of the corresponding paternal chromosome (Robertson, 2005). This condition is called maternal unipaternal disomy. It is not yet known how the loss of expression of these paternally imprinted genes mechanistically results in Prader-Willi syndrome.

The genetic errors associated with Angelman syndrome are, of course, somewhat different. In particular, Tatsuya Kishino *et al.* (1997) showed that Angelman syndrome is due to the loss of expression of a single maternally expressed gene in the region, called UBE3A. The UBE3A gene encodes a protein called E3 ubiquitin ligase, which is involved in targeting proteins for degradation, and it is only imprinted in the brain. The loss of UBE3A may result in abnormalities in normal protein degradation during brain development, thereby causing Angelman syndrome (Kishino *et al.*, 1997). In fact, in most cases (65%-70%), Angelman syndrome is caused by maternally derived deletions of the UBE3A gene (Robertson, 2005). However, this condition can also be caused by paternal unipaternaldisomy (wherein embryos inherit both copies of chromosome 15 from their father), mutations in the UBE3A gene, and imprinting defects, such as the loss of maternal DNA methylation (Robertson, 2005).

Spontaneous Mutation:

- Tautomerism — A base is changed by the repositioning of a hydrogen atom, altering the hydrogen bonding pattern of that base, resulting in incorrect base

pairing during replication.

- Depurination — Loss of a purine base (A or G) to form an apurinic site (AP site).
- Deamination — Hydrolysis changes a normal base to an atypical base containing a keto group in place of the original amine group. Examples include C → U and A → HX (hypoxanthine), which can be corrected by DNA repair mechanisms; and 5MeC (5- methylcytosine) → T, which is less likely to be detected as a mutation because thymine is a normal DNA base.

  Slipped strand mispairing — Denaturation of the new strand from the template during replication, followed by renaturation in a different spot ("slipping"). This can lead to insertions or deletions.

  - Replication slippage

  X chromosome inactivation (XCI) is the process by which the dosage imbalance of X-linked genes between XX females and XY males is functionally equalized. XCI modulates the phenotype of females carrying mutations in X-linked genes, as observed in X-linked dominant male-lethal disorders such as oral-facial-digital type I (OFDI) and microphthalmia with linear skin-defects syndromes. The remarkable degree of heterogeneity in the XCI pattern among female individuals, as revealed by the recently reported XCI profile of the human X chromosome, could account for the phenotypic variability observed in these diseases. Furthermore, the recent characterization of a murine model for OFDI shows how interspecies differences in the XCI pattern between *Homo sapiens* and *Mus musculus* result in discrepancies between the phenotypes observed in patients and mice Consanguinity and its effects.

  In clinical genetics, a consanguineous marriage is defined as a union between two individuals who are related as second cousins or closer, with the inbreeding coefficient (*F*) equal or higher than 0.0156 (Bittles 2001), where (*F*) represents a measure of the proportion of loci at which the offspring of a consanguineous union is expected to inherit identical gene copies from both parents. This includes unions termed first cousins, first cousins once removed and second cousins. In some communities, the highest inbreeding coefficients are reached with unions between double first cousins practiced among Arabs and uncle–niece marriages practiced in South India where (*F*) reaches 0.125 (Hamamy et al. 2011) [Fig. 2].

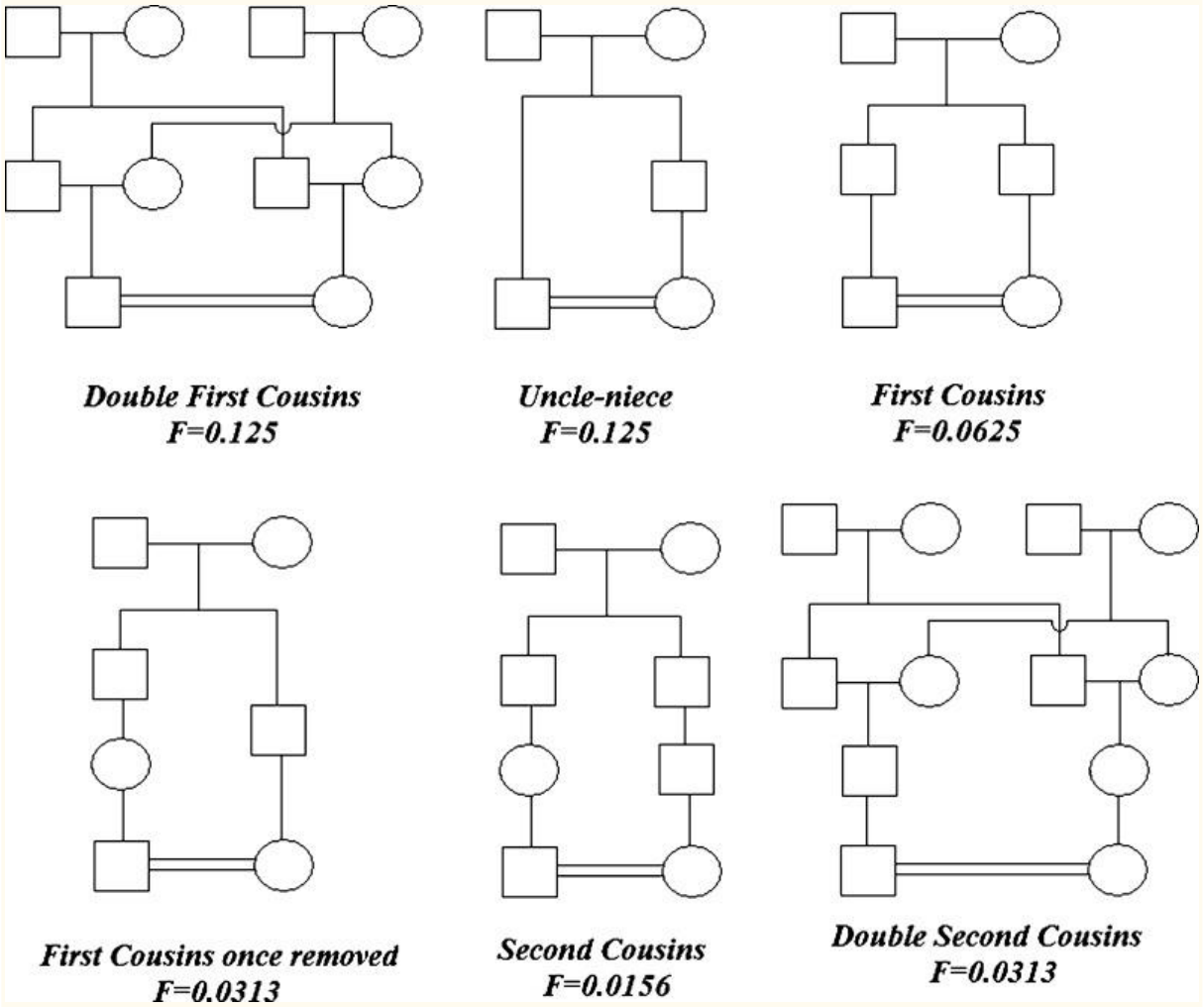


Fig. 2

Categories of consanguineous marriages

In highly consanguineous populations, pedigrees with complex consanguinity loops arising from cousin marriages in successive generations are encountered leading to higher inbreeding coefficients. Reports on consanguinity rates may sometimes include marriages between third cousins or more distantly related individuals. Although this discrepancy affects the total consanguinity rate, the lower coefficients of inbreeding in more remote unions limit a marked alteration of the mean inbreeding coefficient (α). To report and compare consanguinity rates among different populations, the two parameters best used are the mean inbreeding coefficient and the rates of marriages between first cousins.

QUESTIONS FOR PRACTICE

UNIT - 3

6 marks

1. Explain monogenic traits.
2. Differentiate autosomal recessive and autosomal dominant inheritance.
3. What is mitochondrial inheritance?
4. Explain male lethality in OFD1.
5. What are spontaneous mutations? Explain with an example.
6. Write a note on Angelman syndrome.
7. Explain autosomal recessive inheritance
8. Briefly explain X-linked inactivation.

10 marks

1. Explain in detail about autosomal inheritance with examples.
2. What are sex linked inheritance? Explain two disorders which are result of sex linked inheritance.
3. Describe male lethality in detail.
4. Explain about consanguinity and its effects.
5. Write short notes on: a) Mitochondrial inheritance,b) Genomic imprinting, c) Spontaneous mutations.

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF BIO AND CHEMICAL ENGINEERING

DEPARTMENT OF CENTRE FOR MOLECULAR AND NANOMEDICAL SCIENCES

UNIT - IV – COMPLEX TRAITS - (SMB5202)

Unit IV

Complex traits, also known as **quantitative traits**, are traits that do not behave according to simple Mendelian inheritance laws. More specifically, their inheritance cannot be explained by the genetic segregation of a single gene. Such traits show a continuous range of variation and are influenced by both environmental and genetic factors. Compared to strictly Mendelian traits, complex traits are far more common, and because they can be hugely polygenic, they are studied using statistical techniques such as QTL mapping rather than classical genetics methods.[1] Examples of complex traits include height, circadian rhythms, enzyme kinetics, and many diseases including diabetes and Parkinson's disease. One major goal of genetic research today is to better understand the molecular mechanisms through which genetic variants act to influence complex traits.

‘Nature vs nurture’, role of family and shared environment,

The **nature versus nurture** debate involves whether human behavior is determined by the environment, either prenatal or during a person's life, or by a person's genes. The alliterative expression "nature and nurture" in English has been in use since at least the Elizabethan period[1] and goes back to medieval French.[2]

The combination of the two concepts as complementary is ancient (Greek: ἀπό φύσεως καὶεὐτροφίας[3]). Nature is what we think of as pre-wiring and is influenced by genetic inheritance and other biological factors. Nurture is generally taken as the influence of external factors after conception e.g. the product of exposure, experience and learning on an individual.[4]

The phrase in its modern sense was popularized by the English Victorian polymath Francis Galton, the modern founder of eugenics and behavioral genetics, discussing the influence of heredity and environment on social advancement.[5][6][7] Galton was influenced by the book *On the Origin of Species* written by his half-cousin, Charles Darwin.

The view that humans acquire all or almost all their behavioral traits from "nurture" was termed *tabula rasa* ("blank slate") by John Locke in 1690. A

"blank slate view" in human developmental psychology, one that assumes that human behavioral traits develop almost exclusively from environmental influences (sometimes termed "blank-slatism"), was widely held during much of the 20th century. The debate between "blank-slate" denial of the influence of heritability, and the view admitting both environmental and heritable traits, has often been cast in terms of nature *versus* nurture. These two conflicting approaches to human development were at the core of an ideological dispute over research agendas throughout the second half of the 20th century. As both "nature" and "nurture" factors were found to contribute substantially, often in an inextricable manner, such views were seen as naive or outdated by most scholars of human development by the 2000s.[8][9][10][11][12][13]

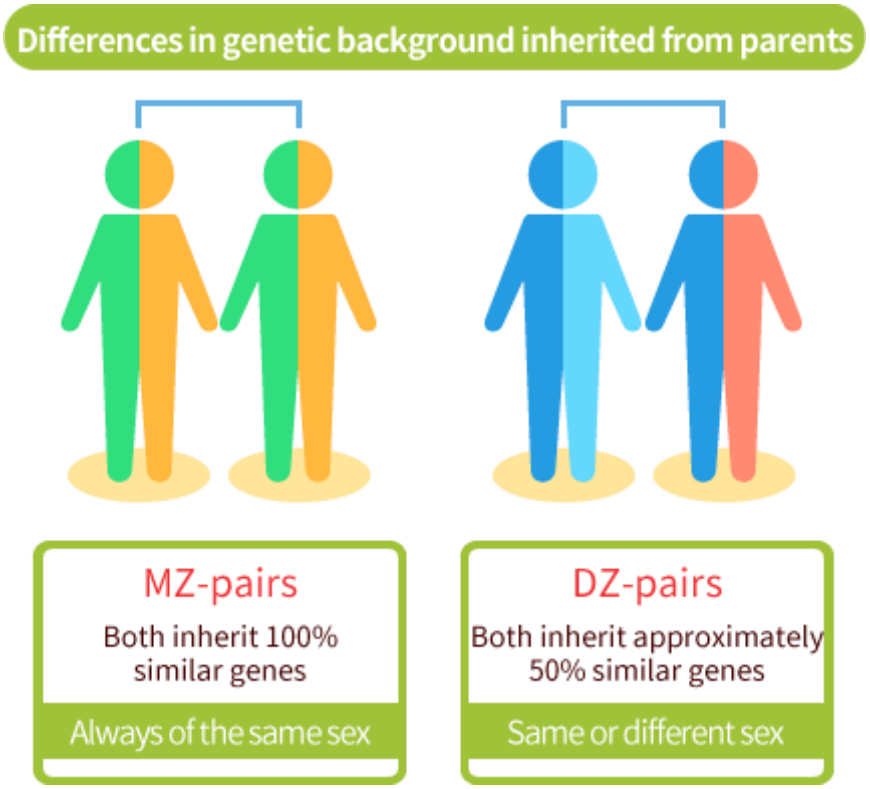
The strong dichotomy of nature *versus* nurture has thus been claimed to have limited relevance in some fields of research. Close feedback loops have been found in which "nature" and "nurture" influence one another constantly, as seen in self-domestication. In ecology and behavioral genetics, researchers think nurture has an essential influence on nature.[14][15] Similarly in other fields, the dividing line between an inherited and an acquired trait becomes unclear, as in epigenetics[16] or fetal development.[17][18]

Monozygotic and dizygotic twins,

There are two kinds of twins: monozygotic and dizygotic.

Monozygotic (MZ) twins are conceived when a single egg fertilized by a single sperm splits into two after fertilization. Such twins have nearly 100% of their genes in common. Both of them invariably have the same blood type and are of the same sex.

Dizygotic (DZ) twins are conceived when two eggs are fertilized by two different sperms. They can simply be thought of as siblings who are born at the same time. They share averagely 50% of their genes, and their blood types and sex can be either same or different.



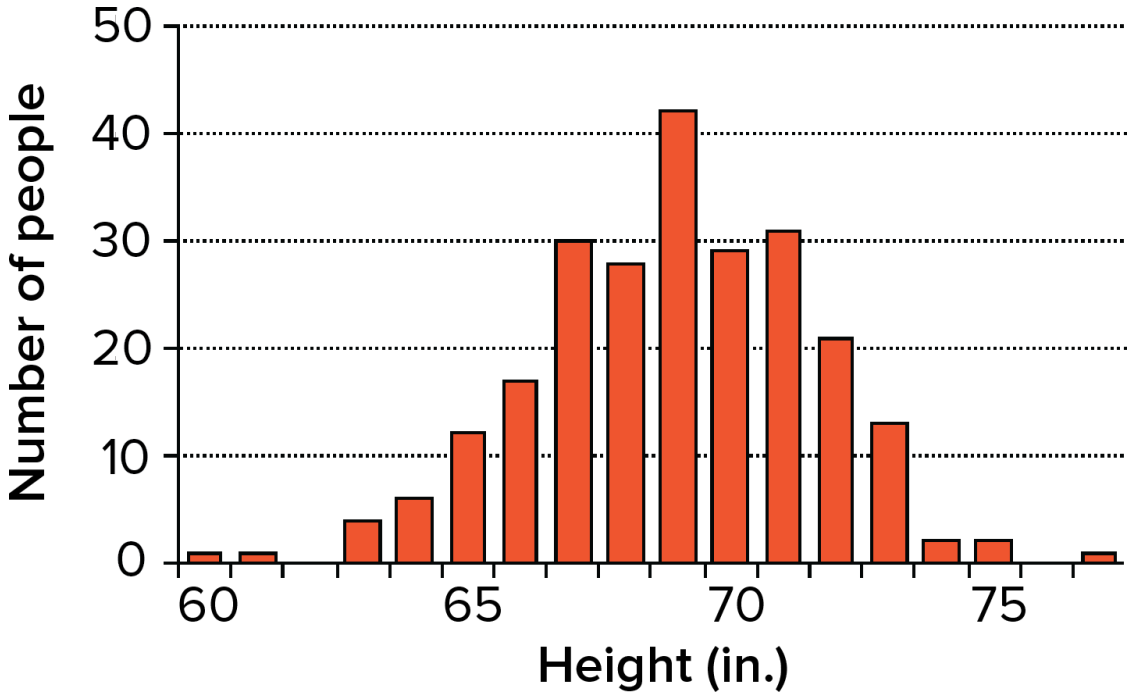
Polygenic inheritance of continuous (quantitative) traits,

How is height inherited?

If what you're really interested in is human genetics, learning about Mendelian genetics can sometimes be frustrating. You'll often hear a teacher use a human trait as an example in a genetics problem, but then say, "that's an oversimplification" or "it's much more complicated than that." So, what's actually going on with those interesting human traits, such as eye color, hair and skin color, height, and disease risk?

As an example, let's consider human height. Unlike a simple Mendelian characteristic, human height displays:

- **Continuous variation.** Unlike Mendel's pea plants, humans don’t come in two clear-cut “tall” and “short” varieties. In fact, they don't even come in four heights, or eight, or sixteen. Instead, it’s possible to get humans of many different heights, and height can vary in increments of inches or fractions of inches^11start superscript, 1, end superscript.



Histogram showing height in inches of male high school seniors in a sample group. The histogram is roughly bell-shaped, with just a few individuals at the tails (60 inches and 77 inches) and many individuals in the middle, around 69 inches.

The heights of a group of male high school seniors. Image modified from "Continuous variation: Quantitative traits," by J. W. Kimball (CC BY 3.0)

- **A complex inheritance pattern.** You may have noticed that tall parents can have a short child, short parents can have a tall child, and two parents of different heights may or may not have a child in the middle. Also, siblings with the same two parents may have a range of heights, ones that don't fall into distinct categories.

  Simple models involving one or two genes can't accurately predict all of these inheritance patterns. How, then, is height inherited?

  Height and other similar features are controlled not just by one gene, but rather, by multiple (often many) genes that each make a small contribution to the overall outcome. This inheritance pattern is sometimes called **polygenic inheritance** (*poly-* = many). For instance, a recent study found over 400 genes linked to variation in height^22squared.

  When there are large numbers of genes involved, it becomes hard to distinguish the effect of each individual gene, and even harder to see that gene variants

(alleles) are inherited according to Mendelian rules. In an additional complication, height doesn’t just depend on genetics: it also depends on environmental factors, such as a child’s overall health and the type of nutrition he or she gets while growing up.

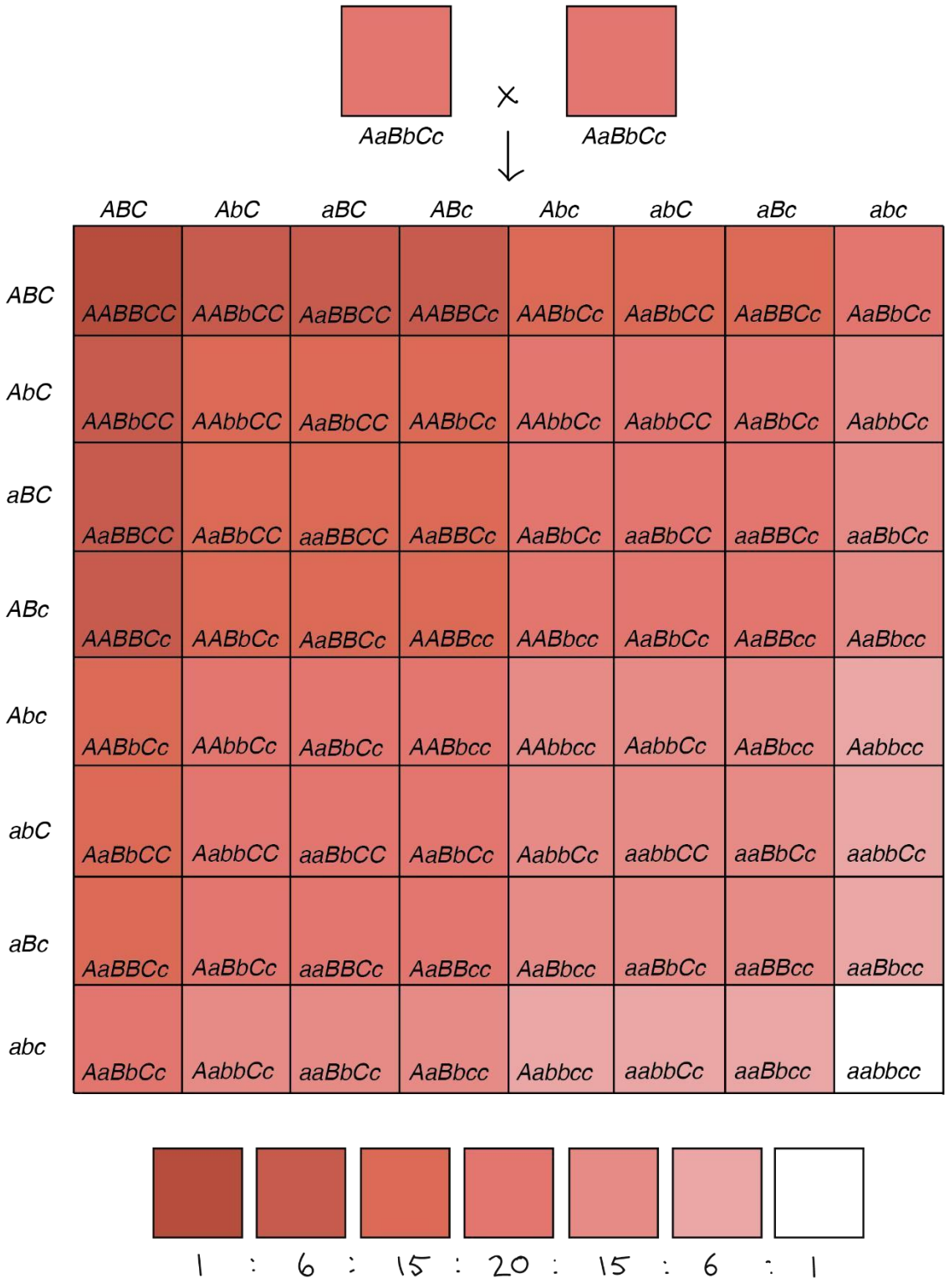
In this article, we’ll examine how complex traits such as height are inherited. We'll also see how factors like genetic background and environment can affect the **phenotype** (observable features) produced by a particular **genotype** (set of gene variants, or alleles).

Polygenic inheritance

Human features like height, eye color, and hair color come in lots of slightly different forms because they are controlled by many genes, each of which contributes some amount to the overall phenotype. For example, there are two major eye color genes, but at least 14 other genes that play roles in determining a person’s exact eye color^33cubed.

Looking at a real example of a human polygenic trait would get complicated, largely because we’d have to keep track of tens, or even hundreds, of different allele pairs (like the 400 involved in height!). However, we can use an example involving wheat kernels to see how several genes whose alleles "add up" to influence the same trait can produce a spectrum of phenotypes^{1,4}1,4start superscript, 1, comma, 4, end superscript.

In this example, there are three genes that make reddish pigment in wheat kernels, which we’ll call *A*, *B*, and *C*. Each comes in two alleles, one of which makes pigment (the capital- letter allele) and one of which does not (the lowercase allele). These alleles have additive effects: the *aa* genotype would contribute no pigment, the *Aa* genotype would contribute some amount of pigment, and the *AA* genotype would contribute more pigment (twice as much as *Aa*). The same would hold true for the *B* and *C* genes^{1,4}1,4start superscript, 1, comma, 4, end superscript.


AaBbCc
×
AaBbCc
ABC AbC aBC ABc Abc abC aBc abc
ABC AABBCC AABbCC AaBBCC AABBCc AABbCc AaBbCC AaBBCc AaBbCc
AbC AABbCC AAbbCC AaBbCC AABbCc AAbbCc AabbCC AaBbCc AabbCc
aBC AaBBCC AaBbCC aaBBCC AaBBCc AaBbCc aaBbCC aaBBCc aaBbCc
ABc AABBCc AABbCc AaBBCc AABBcc AABbcc AaBbCc AaBBcc AaBbcc
Abc AABbCc AAbbCc AaBbCc AABbcc AAbbcc AabbCc AaBbcc Aabbcc
abC AaBbCC AabbCC aaBbCC AaBbCc AabbCc aabbCC aaBbCc aabbCc
aBc AaBBCc AaBbCc aaBBCc AaBBcc AaBbcc aaBbCc aaBBcc aaBbcc
abc AaBbCc AabbCc aaBbCc AaBbcc Aabbcc aabbCc aaBbcc aabbcc
1 : 6 : 15 : 20 : 15 : 6 : 1

64-square Punnett square illustrating the phenotypes of the offspring of an *AaBbCc* x *AaBbCc* cross (in which each uppercase allele contributes one unit of pigment, while each lowercase allele contributes zero units of pigment).

Of the 64 squares in the chart:

1 square produces a very very dark red phenotype (six units of pigment) 6 squares produce a very dark red phenotype (five units of pigment)
15 squares produce a dark red phenotype (four units of pigment). 20 squares produce a red phenotype (three units of pigment)
15 squares produce a light red phenotype (two units of pigment)

6 squares produce a very light red phenotype (one unit of pigment) 1 square produces a white phenotype (no units of pigment)
Diagram based on similar diagram by W. P. Armstrong^55start superscript, 5, end superscript.
Now, let's imagine that two plants heterozygous for all three genes (*AaBbCc*) were crossed to one another. Each of the parent plants would have three alleles that made pigment, leading to pinkish kernels. Their offspring, however, would fall into seven color groups, ranging from no pigment whatsoever (*aabbcc*) and white kernels to lots of pigment (*AABBCC*) and dark red kernels. This is in fact what researchers have seen when crossing certain varieties of wheat^{1,4}1,4start superscript, 1, comma, 4, end superscript.
This example shows how we can get a spectrum of slightly different phenotypes (something close to continuous variation) with just three genes. It's not hard to imagine that, as we increased the number of genes involved, we'd be able to get even finer variations in color, or in another trait such as height.

Dysmorphology

A **dysmorphic feature** is an abnormal difference in body structure. It can be an isolated finding in an otherwise normal individual, or it can be related to

a congenital disorder, genetic syndrome or birth defect. Dysmorphology is the study of dysmorphic features, their origins and proper nomenclature. One of the key challenges in identifying and describing dysmorphic features is the use and understanding of specific terms between different individuals.[1] Clinical geneticists and pediatricians are usually those most closely involved with the identification and description of dysmorphic features, as most are apparent during childhood.

Dysmorphic features can vary from isolated, mild anomalies such as clinodactyly or synophrys to severe congenital anomalies, such as heart defects and holoprosencephaly. In some cases, dysmorphic features are part of a larger clinical picture, sometimes known as a sequence, syndrome or association.[2] Recognizing the patterns of dysmorphic features is an important part of a geneticist's diagnostic process, as many genetic disease present with a common collection of features.[1] There are several commercially available databases that allow clinicians to input their observed features in a patient to generate a differential diagnosis.[1][3] These databases are not infallible, as they require on the clinician to provide their own experience, particularly when the observed clinical features are general. A male child with short stature and hypertelorism could have several different disorders, as these findings are not highly specific.[1] However a finding such as 2,3-toe syndactyly raises the index of suspicion for Smith-Lemli-Opitz Syndrome.[4]

Dysmorphic features are invariably present from birth, although some are not immediately apparent upon visual inspection. They can be divided into groups based on their origin, including malformations (abnormal development), disruptions (damage to previously normal tissue), deformations (damage caused by an outside physical force) and dysplasias (abnormal growth or organization within a tissue).[1][2]

Genetic susceptibility in multifactorial disorders

MULTIFACTORIAL DISORDERS

Researchers are learning that nearly all conditions and diseases have a genetic component. Some disorders, such as sickle cell disease and cystic fibrosis, are caused by mutations in a single gene. The causes of many other disorders,

however, are much more complex. Common medical problems such as heart disease, diabetes, and obesity do not have a single genetic cause—they are likely associated with the effects of multiple genes in combination with lifestyle and environmental factors. Conditions caused by many contributing factors are called complex or **multifactorial disorders**.

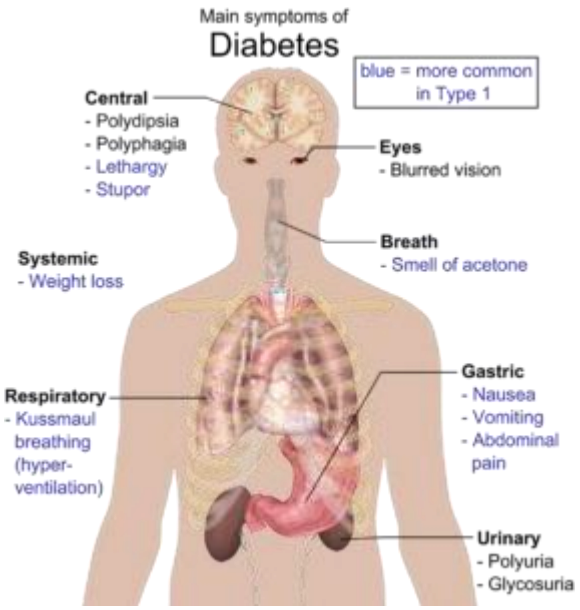


Figure 1 The main symptoms of diabetes, a multifactorial disorder

Although complex disorders often cluster in families, they do not have a clear-cut pattern of inheritance. This makes it difficult to determine a person's risk of inheriting or passing on these disorders. Complex disorders are also difficult to study and treat because the specific factors that cause most of these disorders have not yet been identified. Researchers continue to look for major contributing genes for many common complex disorders.

GENETIC PREDISPOSITIONS

A **genetic predisposition** (sometimes also called genetic susceptibility) is an increased likelihood of developing a particular disease based on a person's genetic makeup. A genetic predisposition results from specific genetic variations that are often inherited from a parent. These genetic changes contribute to the development of a disease but do not directly cause it. Some

people with a predisposing genetic variation will never get the disease while others will, even within the same family.

Genetic variations can have large or small effects on the likelihood of developing a particular disease. For example, certain mutations in the *BRCA1* or *BRCA2* genes greatly increase a person's risk of developing breast cancer and ovarian cancer. Variations in other genes, such as *BARD1* and *BRIP1*, also increase breast cancer risk, but the contribution of these genetic changes to a person's overall risk appears to be much smaller.

Current research is focused on identifying genetic changes that have a small effect on disease risk but are common in the general population. Although each of these variations only slightly increases a person's risk, having changes in several different genes may combine to increase disease risk significantly. Changes in many genes, each with a small effect, may underlie susceptibility to many common diseases, including cancer, obesity, diabetes, heart disease, and mental illness.

In people with a genetic predisposition, the risk of disease can depend on multiple factors in addition to an identified genetic change. These include other genetic factors (sometimes called modifiers) as well as lifestyle and environmental factors. Although a person's genetic makeup cannot be altered, some lifestyle and environmental modifications (such as having more frequent disease screenings and maintaining a healthy weight) may be able to reduce disease risk in people with a genetic predisposition.

QUESTIONS FOR PRACTICE

UNIT - 4

6 marks

1. What are complex traits?
2. Explain Nature vs nurture.
3. Write a note on dysmorphology.
4. Explain genetic susceptibility in obesity.
5. Differentiate monozygotic and dizygotic twins.
6. What are quantitative traits?

10 marks

1. Explain the role of family and shared environment in human genetics.
2. Write a note on polygenic inheritance of continuous traits with examples.
3. Explain genetic susceptibility in multifactorial disorders with three examples.
4. What are principles of complex genetic traits? What is the difference between single gene trait and complex traits?

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF BIO AND CHEMICAL ENGINEERING

DEPARTMENT OF CENTRE FOR MOLECULAR AND NANOMEDICAL SCIENCES

UNIT - V – HUMAN CYTOGENTICS - (SMB5202)

Unit-V

Human cytogenetics, Techniques in human chromosome analysis, Human karyotype: banding, nomenclature of banding, of human chromosomes, Nomenclature of aberrant karyotypes, chromosome abnormalities in cancer, Genetics of fetal wastage, Pharmacogenetics, ecogenetics and teratogenetics.

Human cytogenetics

The field of human cytogenetics was launched in 1956 with this hesitant statement. The serendipitous addition of water to a suspension of human mitotic cells2 , before they were fixed and dropped onto glass microscope slides, caused the chromosomes to spread apart from each other so that Tjio and Levan1 could accurately count the full complement of 46 human chromosomes (FIG. 1). The number 46 was independently confirmed by Ford and Hamerton in the same year3 . The prevailing dogma had held the count at 48 for more than 30 years, ever since the geneticist Thomas Painter had reported on his observations of testicular cells4 . Establishing the correct number and this simple technological advance set off many discoveries that associated specific chromosomal abnormalities with disease in the late 1950s and quickly established the central role of cytogenetics in medicine. In the ensuing years, human cytogenetics has been transformed by technological advances that have combined innovations in molecular biology, chemistry and instrumentation. Cytogeneticists can now extract far more information about the human genome than just chromosome number. Each chromosome can be easily recognized — even in the highly rearranged karyotypes of tumour cells — by colour-coded labels. The resolution and sensitivity of analyses have improved more than 10,000-fold in a very short time, first with the introduction of banding technology and later with fluorescence in situ hybridization (FISH). Extremely subtle alterations in chromosome composition can now be detected and analysed for their association with disease. Cytogeneticists have been freed from their early dependence on mitotic cells by techniques that make it possible to evaluate the karyotype of non-dividing cells. Other approaches yield quantitative information on chromosomal content and structure and allow cytogeneticists to isolate specific chromosomes for molecular analyses. The latest technology

allows genome-wide screens for the loss or gain of chromosomal material to be conducted at unprecedented resolution. Most importantly, the cytogenetic map is cross- referenced to the human draft sequence at thousands of points. These connections greatly facilitate the translation of microscopically visible clues of the molecular basis of disease to the actual genes that are disrupted or altered in dosage. This article outlines the history of the main technological advances that have occurred in human cytogenetics during the past 46 years. It highlights the impact that these advances have on our understanding of the molecular basis of human disease and of the structure, function and evolution of our chromosomes.

A late start, but rapid recovery Flemming and Arnold first observed human chromosomes in the 1880s. It is therefore remarkable that such a fundamental aspect of human biology as chromosome number could have escaped the scientific community until 1956, three years after the structure of the DNA helix was elucidated5 . Friedrich Vogel and Arno Motulsky6 ascribe this delay to both technological and politico-social causes. They assert that most laboratory-based medical scientists at the time were uninterested in human genetics; they considered humans to be far too complex and preferred to focus on simpler model organisms that could be more easily manipulated. Also, many serious geneticists had dissociated themselves from human genetics during the eugenics movement in the early 1900s, which reached its nadir with the horrific practices of the Nazis. However, soon after the number 46 was firmly established, scientists readily applied the new cytogenetic technique to the investigation of phenotype–genotype correlations in humans and began to tap useful information from naturally occurring chromosomal rearrangements. Human cytogeneticists were dealt a good hand by evolution. Had human chromosomes been as morphologically similar as those of mice, or as tiny and numerous as those of most birds, progress in cytogenetics would have been much slower. Fortunately, differences in the relative size of human chromosomes and the position of the centromeric constriction allowed cytogeneticists

to match up the 23 pairs and classify them into seven groups (A to G) with relative ease7 . Although crude, these early karyotypes allowed the discovery that some human disorders result from changes in chromosome number or appearance. In 1959, trisomy 21 was shown to be the cause of Down syndrome8 , and abnormalities in the number of sex chromosomes were shown to cause Turner syndrome (X0) and Klinefelter syndrome (XXY), two frequent disorders of sex differentiation9,10. It also became quickly apparent that most

miscarriages were caused by abnormalities in chromosome number11. Work on the diminutive, but deadly,‘Philadelphia’ chromosome established a new model for using cytogenetic clues to find genes that, when altered, cause human disease. In 1960, cytogeneticists recognized the Philadelphia chromosome as the cause of chronic myeloid leukaemia (CML)12. Thirteen years later, this chromosome was shown by Janet Rowley to be the product of a translocation between chromosomes 9 and 22 (REF. 13). The point at which these two chromosomes break and fuse was the obvious place to look for the molecular explanation of this disease. Indeed, by using the derivative chromosomes in molecular assays, the translocation was shown, in 1985, to create a new hybrid gene of BCR and ABL (breakpoint cluster region and v-abl Abelson murine leukaemia viral oncogene homologue 1)14. Subsequent studies showed that constitutive activation of BCR–ABL, a tyrosine kinase, affects many cellular pathways and leads to the cancer phenotype (reviewed in REF. 15). This understanding in turn led to the development of Gleevec (STI571), a drug that was designed to block the function of the BCR–ABL protein and that has proved to be a highly successful treatment for CML16. The rudimentary chromosome preparations of the early 1960s yielded other breakthroughs in human genetics. Lejeune recognized the first inherited deletion syndrome, Cri du Chat, in 1963; patients with severe mental retardation and a characteristic cat-like cry were all missing a portion of the short arm of chromosome 5 (REF. 17). In the same year, a patient with bilateral retinoblastoma was found to have a deletion of the long arm of a D-group chromosome18. Later work by Cavenee et al.19 provided paradigm-setting proof of KNUDSON’S TWO-HIT HYPOTHESIS20 by showing that the cancer arises owing to the loss of one allele of the RB (retinoblastoma) gene in 13q14 and mutation of the other allele. One of the first autosomal human genes to be mapped, the gene for the DUFFY BLOOD GROUP, was assigned to chromosome 1 because of the consistent way it tracked in families as a visible cytogenetic anomaly near the centromere of chromosome 1 (REF. 21). Chromosomal barcodes The power of cytogenetic analysis redoubled in the late 1960s with TorbjornCaspersson’s development of staining protocols that produced highly reproducible patterns of dark and light bands along the length of each chromosome22. These banding patterns became the barcodes with which cytogeneticists could easily identify chromosomes, detect subtle deletions, inversions, insertions, translocations, fragile sites and other more complex rearrangements, and refine breakpoints (FIG. 2). The bands

appear only in metaphase chromosomes, and cycling cells are therefore required for this analysis. If cells can be caught in prometaphase — when chromosomes are in the very early stages of condensation — up to 2,000 bands can be discerned23; more typically, 400–800 bands are visible. The band-naming convention introduced in 1971 reflects the levels of resolution with which chromosomes can be analysed24. Despite the extensive use of these bands, their cause remains an enigma. They correlate with regional differences in basepair composition, repetitive elements, replication timing and chromatin packaging and can be induced by many methods, but their molecular basis is not understood. Cytogenetic information moved from the bench to the clinic in the late 1960s with the discovery that fetal cells could be obtained through AMNIOCENTESIS and could be checked for chromosomal abnormalities. Methods were quickly developed to induce fetal cells that had been derived from amniotic fluid to divide in culture and to obtain high-quality banded karyotypes. The same procedures are widely used today to provide prenatal diagnostic information to families. Through the painstaking analysis of chromosome banding patterns, thousands of chromosomal abnormalities have been associated with inherited or de novo disorders, generating many leads to the underlying molecular causes of these disorders (see Online links box at the end of the article). Even today, when highresolution genetic linkage analysis can be conducted easily, the discovery of a patient whose disorder is caused by a gross chromosomal abnormality is heralded as a valuable resource for locating the disease gene. Solid tumours also present a myriad of complex chromosomal aberrations — each is a possible clue to tumour initiation and progression. The challenge is to navigate from the visible morphological alteration to the sequence level. The next major advances in cytogenetics facilitated that process. Moving from microscope to molecule Once a rearranged chromosome has been identified, the next step is to position the translocation breakpoints or deletion boundaries relative to genes on molecular maps. This step can be accomplished by using techniques that physically separate abnormal and normal chromosomes so that they can be independently assayed for gene content. Three methods have been particularly useful in achieving this: somatic-cell-hybrid technology, fluorescence-activated cell (chromosome) sorting (FACS) and FISH (all discussed below). These techniques help researchers to zoom in on the defect from the cytogenetic to the molecular level, and, importantly, they have yielded rough maps for navigating the genome and for allowing more detailed molecular

mapping and sequencing. Somatic-cell hybrids are a fortunate quirk of cell biology. When rodent and human cells are fused in the laboratory, human chromosomes are preferentially ejected, but some are retained25,26. This phenomenon was capitalized on by the groups of Weiss and Ruddle, who were the first to use panels of hybrid cell lines, each retaining a different set of human chromosomes, to map genes and anonymous markers to specific chromosomes or portions thereof 27,28. The chromosomal content of each line, established by cytogenetic analysis, is simply correlated with the results of hybridization assays, functional tests or PCR to assign a gene or marker of interest to a chromosome. Much more precise maps, which served as frameworks for the assembly of the human genome sequence, were generated using panels that contain different chromosomal fragments, such as aberrant chromosomes transferred from the cells of patients29 or fragments that were experimentally produced by radiation30. Originally developed for cell analysis and separation, flow cytometry was adapted in 1979 for the quantitative analysis and sorting of human chromosomes by a team of investigators at the Lawrence Livermore National Laboratory in California31. In this technique, chromosomes are released into suspension from mitotic cells and stained with two fluorescent DNA dyes that have different base-pair specificities: this allows all but four human chromosomes (9–12) to be resolved by a flow cytometer32 (FIG. 3). The measurements give quantitative information on the extent of normal variation in chromosome size (some vary by 50% in DNA content) and the amount of DNA that is missing or gained in abnormal chromosomes33,34. Abnormal and normal chromosomes can also be separated for the molecular characterization of DNA-marker retention or loss35. Flow sorting was the key to the production of chromosome-specific DNA clone libraries36,37, which have been important for constructing detailed, marker-dense physical maps of the genome, especially in the days when tackling the whole genome at once seemed too daunting. Flow sorting continues to be the technique of choice for producing chromosome-specific paints38 (see below) and for characterizing sequences that are duplicated on more than one chromosome39. The FISHing trip The next advance to revolutionize cytogenetics, FISH, provided a direct link between microscope and sequence. This technique allows the chromosomal and nuclear locations of specific DNA sequences to be seen through the microscope (FIG. 4). Each probe is a cloned piece of the genome that is conjugated to a reporter molecule, such as biotin. After denaturation, the probe is allowed to seek out its

complement in the chromosomal DNA, and these locations are then marked with a fluorescent reagent, such as avidin-FITC, that binds to the reporter attached to the DNA probe. Although radioactively labelled DNA and RNA probes had been localized to cytogenetic bands since 1969 (REFS 40,41), the field advanced significantly when groups led by David Ward and Mels van der Ploeg replaced the isotopic labels with fluorescent ones42,43. Fluorescent tags are safer and simpler to use, can be stored indefinitely, give higher resolution and opened up prospects for simultaneously locating several DNA sequences in the same cell by labelling them with different fluorochromes. In less than 15 years, the sensitivity of FISH improved 10,000-fold. This remarkable achievement can be attributed to improvements in the probe labels that made them less bulky, simpler to incorporate into the probe and brighter; in the optics for fluorescence microscopy; and in more mundane, but crucial, aspects of the procedure, such as probe fragmentation and slide storage. By 1985, the first single-copy human gene,thyroglobulin, had been localized to a chromosome band by non-radioactive in situ hybridization44. This feat was an important milestone, even though thyroglobulin was one of the largest genes known at the time, and specificity was achieved by fastidiously removing all the interspersed repetitive elements from the probe before its use. Today, localizing segments as small as 10 kb is routine and 1 kb is achievable45. We now exploit the kinetics of DNA reassociation to pre-anneal the repetitive elements, so that only the unique/low-copy portions of the labelled probe are available for hybridization to chromosomes41,46. Using new probes that are based on PEPTIDE NUCLEIC ACID chemistry, the intensity of FISH spots is a reasonable measure of the local amount of complementary target. A good illustration is the study of telomere dynamics in normal and immortalized cells by quantitative analyses of TTAGGG-specific probes bound to the ends of chromosomes47. A clever modification of FISH (called COD- FISH) goes even further to reveal the absolute 3′–5′ direction of a particular sequence on the chromosome48 and to detect inversions and sister-chromatid exchanges49,50 The interplay between genome mapping and cytogenetics escalated in the mid-1980s as FISH technology improved and cloned DNA reagents became available through the efforts of many genome mapping and sequencing groups. Using FISH, cytogeneticists could detect chromosomal abnormalities that involve small segments of DNA — if their probe was situated, fortuitously or by design, in the affected chromosomal segment51. Cytogeneticists were no longer limited to

the resolution afforded by crude banding patterns. Conversely, FISH could be used to establish the order of DNA clones relative to bands, naturally occurring breakpoints and other clones (for example, REF. 52) (FIG. 4). These data were funnelled into the genome project as independent tests of the validity of maps constructed by other techniques53. In turn, as molecular biologists filled in the genome maps, large collections of molecular reagents in the form of cloned, mapped segments of the human genome (cosmids, BACS, PACS AND YACS) became available with which abnormal chromosomes could be characterized by FISH to identify affected genes. For example, FISH analyses identified clones that cross the two breakpoints of the PERICENTRIC INVERSION of chromosome 16 seen in patients with acute myelogenous leukaemia (AML). This finding set the stage for the identification of the two genes (MYH11, smooth muscle myosin heavy chain 11, and CBFB, the β-subunit of core-binding factor) that, when aberrantly fused, cause the leukaemic transformation54,55. Cytogenetic studies in Sam Latt's laboratory were crucial to the discovery that Angelman and Prader–Willi syndromes are disorders of IMPRINTING: rearrangements in 15q11–15q13 were invariably found in the maternal or paternal copy of this region, respectively56, and FISH has been crucial in the identification of imprinted genes in this region (reviewed in REF. 57). The genome-wide view afforded by FISH has also revealed sequences that have been duplicated at distinct sites in the human genome; these sequences light up at more than the two expected sites and can be flagged for special attention during the assembly of the draft sequence53. Furthermore, many of these duplicated blocks have been implicated in chromosomal rearrangements that cause disease and are therefore of biological interest (reviewed in REF. 58). Even more importantly, FISH opened up the nuclei of non-dividing cells to karyotype analysis. Conventional cytogenetics requires the capture of cells in mitosis, and many samples, particularly those from solid tumours, produce few, if any, analysable metaphases. Using FISH and chromosome-specific probes, cytogeneticists can enumerate chromosomes, simply by counting spots in each nucleus59,60. Deviations in spot number also signal gene deletion and amplification. Because DNA is packaged ~10,000-fold more loosely in interphase nuclei than in metaphase chromosomes, abnormalities that are not resolvable by metaphase FISH, such as the 1-Mb duplication that causes CHARCOT– MARIE–TOOTH SYNDROME61, can be detected by interphase FISH (FIG. 5). Shifts in relative spot position reveal structural rearrangements,

such as translocations and inversions62 (FIG. 5). Interphase FISH has also made it possible to determine the relative times at which specific DNA sequences are replicated during the S phase of the cell cycle. Before replication, the probe generates a single dot on each chromosome, whereas two closely juxtaposed dots are visible after replication63. Using this approach, it was found that the order of replication is carefully orchestrated, and, for most loci, that the maternal and paternal alleles replicate in synchrony. By contrast, alleles of most imprinted loci are asynchronously replicated, with the expressed allele replicating earlier than the silenced one64. As the relationship between sequence proximity in interphase chromatin and separation along the DNA helix was elucidated, the order of DNA sequences could be inferred with 50–100-kb resolution by measuring the distances between fluorescent spots that mark DNA sequences of interest65. The ultimate in cytogenetic resolution is reached by wiping out nuclear organization altogether and conducting FISH on DNA fibres that have been affixed to glass (fibre-FISH)66,67. What is condensed to a small spot at the resolution of light microscopy in interphase becomes a long fluorescent line in fibre-FISH. Fibre-FISH is used to resolve ambiguities in the order of genes in a chromosomal region, to analyse the organization of tandem duplications and to detect small-scale rearrangements in chromosomes. Clinical cytogenetics laboratories now make significant use of FISH in both their diagnostic and their research work. FISH is routinely used to augment conventional banding analyses of chromosomal rearrangements. Cytogeneticists have at their disposal various commercially available probe kits that are tailored for specific questions, such as the diagnosis of syndromes caused by chromosomal abnormalities that are too subtle to detect reliably by banding. The FISH test for SMITH–MAGENIS SYNDROME, which uses a probe for a small deleted region of chromosome 17, is an excellent example68. In research, FISH features prominently in the cytogeneticists' process of finding recurrent translocation breakpoints or overlapping deletions among patients with similar phenotypic abnormalities. Chromosome painting with a colourful palette The thrill of seeing a single-copy gene fluoresce in a human cell was soon surpassed by the vivid image of 24 human chromosomes painted in different colours69,70 (FIG. 6). This powerful development, called spectral karyotyping (SKY) or multiplex (M)-FISH, combines three significant advances. First was the production of chromosome-specific 'paints': collections of sequences derived from each chromosome (usually by flow sorting)71,72.

These collections can be generated easily from small numbers of chromosomes using DEGENERATE OLIGONUCLEOTIDE-PRIMED PCR38 or LINKER–ADAPTOR PCR73.When used as a probe, these collections label a chromosome end to end. (Region-specific paints can be generated if microdissected portions of chromosomes are used as a template74.) Second was the combination of fluorochromes to produce 24 colour combinations, one for each chromosome75. Third were the advances in microscopic optics, filters and imaging systems for multicolour analyses. In the SKY system, the spectral characteristics of each pixel in the image are read out by an INTERFEROMETER69. In M-FISH, the spectral characteristics are evaluated by collecting images through a series of excitation and emission filters70. These imaging systems can be taught to classify each chromosomal segment automatically, and they offer the first real hope of automated karyotype analysis. So far, no system can classify banded chromosomes as robustly and accurately as a skilled cytogeneticist, despite the millions of dollars that have been invested in automated karyotype analysis since 1968. SKY and M-FISH have proved to be extremely useful for detecting translocations and other complex aberrations (FIG. 6). For example, SKY has revealed amplification of regions on 11q, 21q and 22q that had not been detected before in AML patients with complex karyotypes; these defects could have a significant role in leukaemogenesis76. Even the karyotypes of tumours in mice can be deciphered77. M-FISH has been especially helpful in the study of radiation-induced damage and chromosome repair78. Although the breaks occur randomly, they are repaired in non- random patterns that reflect the proximity of the breaks in the nucleus during the repair process. So, SKY both has an impact on radiation dosimetry and gives insights into the organization of the human cell nucleus79. M-FISH has also sparked a new industry of probe development to monitor many loci at once for subtle aberrations. The best example is the use of probes that mark the unique sequence near each telomere to detect subtle rearrangements of the ends of chromosomes80. With this technique, as many as 7% of patients with previously unexplained mental retardation have been found to have chromosome abnormalities that had gone undetected in previous analyses81. One of the most thriving areas of cytogenetics today is the study of the chromosomal rearrangements that occurred during evolution82,83. During each speciation event, some cards in the genome deck are moved. These events can be reconstructed with FISH. Such studies have revealed, for example, that the

evolutionary rate of chromosomal translocations is ten times greater between the mouse and the rat genomes than between those of humans and cats or chimpanzees84. Comparative cytogenetics is also crucial for disease-gene mapping. The use of dogs to identify genes that cause human disease is a case in point85. At least half of the inherited disorders that are recognized in various dog breeds correspond to specific human diseases, including various forms of cancer, deafness, heart disease, blindness and epilepsy. With extensive dog pedigrees, it is feasible to genetically map the canine disease to a region of the dog genome. Comparative cytogenetic maps of the human and dog genomes, produced by hybridizing human chromosome paints to dog chromosomes86, show where to dig in the human genome for candidate genes, which can then be tested for mutations in dogs and/or humans. CGH-arrays — a surrogate for chromosomes The next transformation of cytogenetics came with the realization that genome-wide scans for the loss or gain of chromosomal material could be conducted without even looking directly at the subject's chromosomes. The technique that made this possible is called comparative genome hybridization (CGH) and was developed by a team led by Ollie and Anna Kallioniemi, Dan Pinkel and Joe Gray 87. In this approach, the genomic DNA of test and reference samples is isolated, fragmented, labelled in red and green, respectively, and allowed to compete for hybridization sites in sets of normal chromosomes (FIG. 7). As in regular FISH, interspersed repetitive elements are taken out of the picture by pre-annealing the probes with unlabelled DNA that is enriched for repetitive sequences. The ratio of red-to-green fluorescence is measured along the length of each chromosome. The chromosomal regions that are equally represented in the test and reference samples appear orange, but those deleted or amplified in the test sample appear more red or more green. CGH is particularly important in cancer cytogenetics, in which it is used to identify chromosomal regions that are recurrently lost or gained in tumours. For example, CGH led the way to the identification of PIK3CA, the catalytic subunit of phosphatidylinositol 3-kinase (PI3K), as an oncogene in ovarian cancer88. DNA-amplification techniques have also been developed to find genetic alterations in small samples of rare cells89, such as rogue cells found in blood that have escaped a primary tumour and might foreshadow metastasis. The current excitement in cytogenetics revolves around the promise of array-CGH90 (FIG. 7). In this technique, metaphase chromosomes are replaced by an array of thousands of BAC clones, each of which contains an ~150-kb segment of the

human genome. An array of 3,000 BACs can be constructed that samples the genome, on average, once every megabase pair53. Array-CGH is therefore the equivalent of conducting thousands of FISH experiments at once, but without the need to count dots to measure the copy number of each test locus. CGH provides better quantification of copy number and more precise information on the breakpoints of segments that are lost or gained than does conventional CGH. More importantly, each clone is an entry point to the genomic sequence in which affected genes can be identified. Although CGH is insensitive to changes that are present at low frequency in the cells being analysed, it is expected that arrayCGH will enable many groups to evaluate large numbers of tumours for recurrent changes using a common platform. These analyses should generate prognostic markers, identify new tumour-suppressor genes or oncogenes and, ultimately, lead to a better understanding of the cancer process. In addition, I predict that some prenatal diagnostic tests that now rely on banding and conventional FISH will also be supplanted by custom arrays. It is hoped that technological advances, such as array-CGH, will reduce the time and cost of cytogenetic analyses so that they can be accessed by more families.

Human Karyotype

Karyotyping is the process of pairing and ordering all the chromosomes of an organism, thus providing a genome-wide snapshot of an individual's chromosomes. Karyotypes are prepared using standardized staining procedures that reveal characteristic structural features for each chromosome. Clinical cytogeneticists analyze human karyotypes to detect gross genetic changes—anomalies involving several megabases or more of DNA. Karyotypes can reveal changes in chromosome number associated with aneuploid conditions, such as trisomy 21 (Down syndrome). Careful analysis of karyotypes can also reveal more subtle structural changes, such as chromosomal deletions, duplications, translocations, or inversions. In fact, as medical genetics becomes increasingly integrated with clinical medicine, karyotypes are becoming a source of diagnostic information for specific birth defects, genetic disorders, and even cancers.

Preparing Karyotypes from Mitotic Cells

Karyotypes are prepared from mitotic cells that have been arrested in the metaphase or prometaphase portion of the cell cycle, when chromosomes assume their most condensed conformations. A variety of tissue types can be used as a source of these cells. For cancer diagnoses, typical specimens include tumor biopsies or bone marrow samples. For other diagnoses, karyotypes are often generated from peripheral blood specimens or a skin biopsy. For prenatal diagnosis, amniotic fluid or chorionic villus specimens are used as the source of cells.

The process of generating a karyotype begins with the short-term culture of cells derived from a specimen. After a period of cell growth and multiplication, dividing cells are arrested in metaphase by addition of colchicine, which poisons the mitotic spindle. The cells are next treated with a hypotonic solution that causes their nuclei to swell and the cells to burst. The nuclei are then treated with a chemical fixative, dropped on a glass slide, and treated with various stains that reveal structural features of the chromosomes.

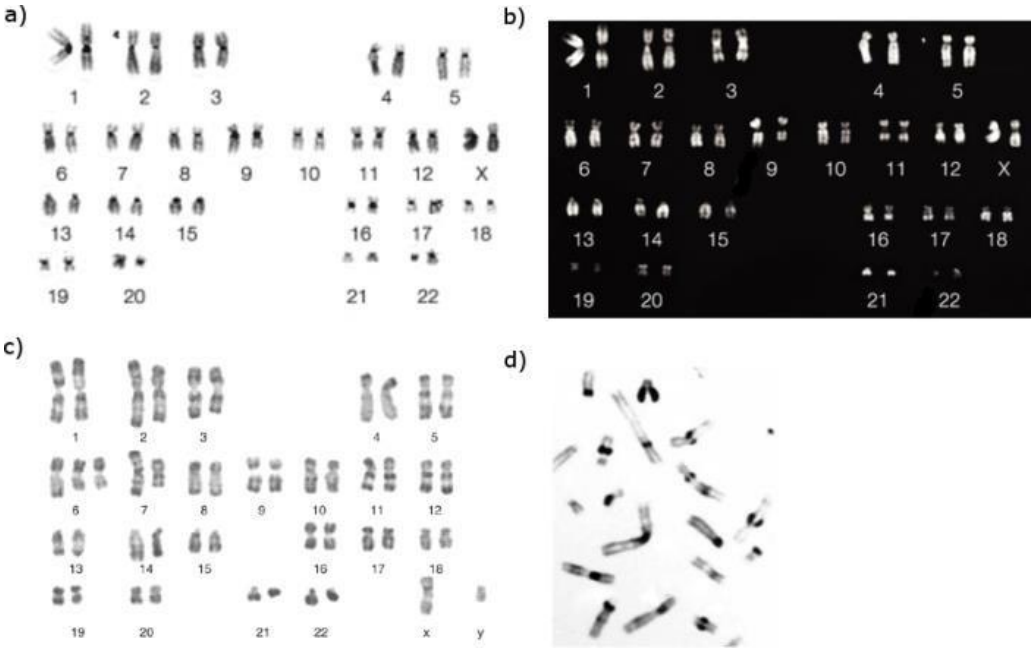
Banding Patterns Reveal the Structural Details of Chromosomes

Without any treatment, structural details of chromosomes are difficult to detect under a light microscope. Thus, to make analysis more effective and efficient, cytologists have developed stains that bind with DNA and generate characteristic banding patterns for different chromosomes. Prior to the development of these banding techniques, distinguishing chromosomes from one another proved very difficult, and chromosomes were simply grouped according to their size and the placement of their centromeres.

This changed in 1970, when TorbjornCaspersson and his colleagues described the first banding technique, known as Q-banding. Q-banding involves use of the fluorescent dye quinacrine, which alkylates DNA and is subject to quenching over time. Caspersson *et al*. demonstrated that quinacrine produced characteristic and reproducible banding patterns for individual chromosomes. Since then, researchers have developed a variety of other chromosome banding techniques that have largely supplanted Q-banding in clinical cytogenetics. Today, most karyotypes are stained with Giemsa dye, which offers better resolution of individual bands, produces a more stable preparation, and can be analyzed with ordinary bright-field microscopy.

The molecular causes for staining differences along the length of a chromosome are complex and include the base composition of the DNA and local differences in chromatin structure. In G-banding, the variant of Giemsa staining most commonly used in North America, metaphase chromosomes are first treated briefly with trypsin, an enzyme that degrades proteins, before the chromosomes are stained with Giemsa. Trypsin partially digests some of the chromosomal proteins, thereby relaxing the chromatin structure and allowing the Giemsa dye access to the DNA.

In general, heterochromatic regions, which tend to be AT-rich DNA and relatively gene-poor, stain more darkly in G-banding. In contrast, less condensed chromatin—which tends to be GC-rich and more transcriptionally active—incorporates less Giemsa stain, and these regions appear as light bands in G-banding. Most importantly, G-banding produces reproducible patterns for each chromosome, and these patterns are shared between the individuals of a species. An example of Giemsa-stained human chromosomes, as they would appear under a microscope, is shown in Figure 1a. Typically, Giemsa staining



produces between 400 and 800 bands distributed among the 23 pairs of human chromosomes. Measured in DNA terms, a G-band represents several million to 10 million base pairs of DNA, a stretch long enough to contain hundreds of genes.

Figure 1: Chromosome banding revealed by different staining techniques.

Different chromosomal staining techniques reveal variations in chromosome structure. Cytogeneticists use these patterns to recognize the differences between chromosomes and enable them to link different disease phenotypes to chromosomal abnormalities. Giemsa banding (a), Q-banding (b), R-banding (c) and C-banding (d) are shown.

Figure Detail

G-banding is not the only technique used to stain chromosomes, however. R-banding, which is used in parts of Europe, also involves Giemsa stain, but the procedure generates the reverse pattern from G-banding. In R-banding (Figure 1c), the chromosomes are heated before Giemsa stain is applied. The heat treatment is thought to preferentially melt the DNA helix in the AT-rich regions that usually bind Giemsa stain most strongly, leaving only the comparatively GC-rich regions to take up the stain. R-banding is often used to provide critical details about gene-rich regions that are located near the telomeres. Yet another method is C-banding (Figure 1d), which can be used to specifically stain constitutive heterochromatin, or genetically inactive DNA, but it is rarely used for diagnostic purposes these days. C-banding is a specialized Giemsa technique that primarily stains chromosomes at the centromeres, which have large amounts of AT-rich satellite DNA. The first method to be used to identify all 46 human chromosomes was Q-banding (Figure 1b), which is achieved by staining the chromosomes with quinacrine and examining them under UV light. This method is most useful for examining chromosomal translocations, especially ones involving the Y chromosome. Taken together, these banding techniques offer clinical cytogeneticists an arsenal of staining methods for diagnosing chromosomal abnormalities in patients.

Organizing Chromosomes in Karyograms for Review

In order to maximize the diagnostic information obtained from a chromosome preparation, images of the individual chromosomes are arranged into a standardized

format known as a karyotype, or more precisely, a karyogram (Figure 1a-c). According to international conventions, human autosomes, or non-sex chromosomes, are numbered from 1 to 22, in descending order by size, with the exceptions of chromosomes 21 and 22, the former actually being the smallest autosome. The sex chromosomes are generally placed at the end of a karyogram.

Within a karyogram, chromosomes are aligned along a horizontal axis shared by their centromeres. Individual chromosomes are always depicted with their short p arms—p for "petite," the French word for "small"—at the top, and their long q arms—q for "queue"—at the bottom. Centromere placement can also be used to identify the gross morphology, or shape, of chromosomes. For example, metacentric chromosomes, such as chromosomes 1, 3, and 16, have p and q arms of nearly equal lengths. Submetacentric chromosomes, such as chromosomes 2, 6, and 10, have centromeres slightly displaced from the center. Acrocentric chromosomes, such as chromosomes 14, 15, and 21, have centromeres located near their ends.

Arranging chromosomes into a karyogram can simplify the identification of any abnormalities. Note that the banding patterns between the two chromosome copies, or homologues, of any autosome are nearly identical. Some subtle differences between the homologues of a given chromosome can be attributed to natural structural variability among individuals. Occasionally, technical artifacts associated with the processing of chromosomes will also generate apparent differences between the two homologues, but these artifacts can be identified by analyzing 15–20 metaphase spreads from one individual. It is highly unlikely that the same technical artifact would occur repeatedly in a given specimen.

Using Karyograms to Detect Chromosomal Abnormalities

Today, G-banded karyograms are routinely used to diagnose a wide range of chromosomal abnormalities in individuals. Although the resolution of chromosomal changes detectable by karyotyping is typically a few megabases, this can be sufficient to diagnose certain categories of abnormalities. For example, aneuploidy, which is often caused by the absence or addition of a chromosome, is simple to detect by karyotype analysis. Cytogeneticists can also frequently detect much more subtle deletions or insertions as deviations from normal banding patterns. Likewise, translocations are often readily apparent on karyotypes.

When regional changes in chromosomes are observed on karyotypes, researchers often are interested in identifying candidate genes within the critical interval whose misexpression may cause symptoms in patients. This search process has been greatly facilitated by the completion of the Human Genome Project, which has correlated cytogenetic bands with DNA sequence information. Consequently, investigators are now able to apply a range of molecular cytogenetic techniques to achieve even higher resolution of genomic changes. Fluorescence *in situ* hybridization (FISH) and comparative genomic hybridization (CGH) are examples of two approaches that can potentially identify abnormalities at the level of individual genes.

Molecular cytogenetics is a dynamic discipline, and new diagnostic methods continue to be developed. As these new technologies are implemented in the clinic, we can expect that cytogeneticists will be able to make the leap from karyotype to gene with increasing efficiency.

References and Recommended Reading

Caspersson, T., Zech, L., & Johansson, J. Differential banding of alkylating fluorochromes in human chromosomes. *Experimental Cell Research* **60**, 315–319 (1970) doi:10.1016/0014-4827(70)90523-9

Gartler, S. M. The chromosome number in humans: A brief history. *Nature Reviews Genetics* **7**, 655–660 (2006) doi:10.1038/nrg1917 (link to article)

Speicher, M. R., Ballard, S. G., & Ward, D. C. Karyotyping human chromosomes by combinatorial multi-fluor FISH. *Nature Genetics* **12**, 368–375 (1996) (link to article)

Strachan, T., & Read, A. P. *Human Molecular Genetics*, 2nd ed. (Wiley, New York, 1999) Tjio, J. H., & Levan, A. The chromosome number of man. *Hereditas* **42**, 1–6 (1956)

Trask, B .J. Human cytogenetics: 46 chromosomes, 46 years and counting. *Nature Reviews Genetics* **3**, 769–778 (2002) doi:10.1038/nrg905 (link to article)

Pathology of Human Chromosome

The pathology of the human chromosomes (the " substance " of heredity) is the subject of cytogenetics, in which the chromosomes are examined and a karyotype established, at the resolution level of the optical microscope.

Cytogenetic analysis is performed by pairing our forty-six human chromosomes in a well- defined order based on characteristic bands inherent to each chromosome. This

analysis is undertaken on a cell undergoing mitosis (metaphase).

A karyotype may be established on any cell line capable of division (mitosis), but readily accessible lineages are generally used i.e.: lymphocytes extracted from whole blood, or, in case of postnatal diagnosis, fibroblasts grown from skin biopsies. For prenatal purposes, trophoblast-derived cell lines (choriocentesis), amniocytes (amniocentesis) or lymphocytes from fetal blood (cordocentesis) are studied. Cell lines grown from curettage material (abortion) may also be used. Sperm and oocyte karyotypes may also be established on the basis of cross-fertilization techniques recently developed.

The cells obtained are processed in the following manner:
Culture: the duration is determined by the type of tissue and can be omitted in the case of cells with a high spontaneous mitotic capacity such as trophoblastic cells of placental origin.
Colchicine treatment: abruptly stops the mitosis in metaphase.

Hypotonic shock: swells the cells and allows the cytoplasm to be eliminated. Fixation.
Production of bands by various denaturing techniques which give rise to patterns specific to each chromosome.
Examination of several dividing nuclei. The chromosomes are paired (23 pairs) and numbered in order of size, the first pair being the longest.
A human male has 44 autosomes and 2 gonosomes (XY)
A human female has 44 autosomes and 2 gonosomes (XX)
In normal chromosome structure the centromere or central core is flanked by two " arms ": the short arm (p) and the long arm (q). The basic structure is the same for all chromosomes, with the exception that the acrocentric chromosomes have little or no short arm. The 23 pairs differ in the length of the arms, and each shows unique banding pattern.

Chromosome aberrations

Numerical aberrations, which can affect either the autosomes or the gonosomes (sex chromosomes), involve the loss or gain of a part of or an entire chromosome. This gives rise in the latter case to a trisomy (a chromosome of

which three copies are present) or in the former to a monosomy (a single copy).

Structural aberrations affect a portion of one or several chromosomes; these can be of several types, and imply in general that a break has taken place and given rise to one of the following rearrangements:

Translocations, the most frequent type of rearrangement between two chromosomes or two arms of the same chromosome, involve an exchange of genetic material; depending on whether the quantity of chromatin is modified, they are either balanced or unbalanced.
Insertions of chromosomal material in an unusual location. Inversions: a portion of a chromosome is upside-down.
Ring chromosomes: annealing of the two extremities of a chromosome after loss of their terminal segments.
Isochromosomes: one arm is present in two copies, separated from each other by a centromere.
Clinical characteristics of unbalanced chromosomal aberrations

Missing or excessive autosomal material (e.g. trisomies) gives rise to two or more of the following, and causes clinically-recognizable syndromes:

Mental retardation.

Dysmorphic features.

Malformation of internal organs.

Retarded growth (of prenatal or postnatal origin).

Although the number and importance of symptoms is variable even within a group of patients with the same anomaly, a certain pattern generally exists, producing a characteristic phenotype which usually permits clinical diagnosis. Mental retardation is the common denominator. Monosomies of entire autosomes are incompatible with life, but partial monosomies due to small or even large deletions can affect all chromosomes, generally producing severe clinical syndromes.

Clinical characteristics of balanced autosomal rearrangements

In this case, the normal quantity of chromosomal material is preserved, but in disorganized fashion either due to an exchange between two chromosomes (translocation), with no net loss or gain of genetic material, or due to a centromere fusion. Whatever the structural defect, there is rarely a major clinical abnormality, but reproductive problems may arise; in women spontaneous abortion is most often seen, whereas in men spermatogenesis may be impaired, resulting in primary sterility.

Clinical characteristics of gonosomal abnormalities

The clinical implications are different from autosomic abnormalities: growth retardation is less frequent (with the exception of X monosomy or Turner's syndrome), and in some sex chromosomal disorders, growth may even be accelerated. Dysmorphism is more subtle, and internal malformations and mental retardation are infrequent. Borderline intelligence is however not unusual in XXY and XXX syndromes, and 45,X women have a higher incidence of internal (heart and renal) malformations than women with other disorders.

X chromosome monosomy is the only example of a monosomy compatible with life but even conceptuses with this defect are for the most part " eliminated " as spontaneous abortions.

Most important are fertility problems secondary to gonadal dysgenesis, which is an almost constant feature in large X chromosome structural defects. In X monosomy, however, the presence of two cell lines (a mosaicism) may lead to fertility, especially if the normal cell line predominates. Reproduction may also remain possible if the affected gonosomal segment is sufficiently small.

When is a karyotype indicated? Prenatal diagnosis

In pregnancies of older women, the incidence of aneuploidy (the presence of an abnormal number of chromosomes) rises steeply after the age of 35 years (about 1/200 births) and reaches 1/50 by 40 years of age. See Table 1 for the incidence of Down Syndrome.

In the case of familial chromosomal aberrations, whether they concern autosomal translocations or maternal gonosomal mosaicisms (45,X/ 46,XX/ 47,XXX).
In familial hereditary disorders of unknown etiology, to exclude a visible chromosomal abnormality, in parallel with molecular studies.
Prenatal diagnosis is sometimes offered to reduce the anxiety of parents of a child with a congenital defect, whether of chromosomal origin or not; this is not, however, based on medical need.

In pediatrics When at least two of the following abnormalities are present:

Internal malformations.

Facial dysmorphism.

Delayed psychomotor development.

Small stature, or retarded growth.

Delayed puberty.

Gynecological indications Recurrent spontaneous abortions: in this case, 1/17 couples will carry a familial aberration in a balanced form, most frequently a translocation. When these are transmitted in an unbalanced form, partial monosomy or trisomy may ensue, often leading to spontaneous abortion.
Sterility in males with azoospermia or oligospermia (sperm count inferior to 10 million/ ml), with or without testicular atrophy. This may be due to gonosomal or autosomal abnormalities, the latter occasionally impairing fertility even in a balanced form (the degree of impairment varying from mild reproductive difficulties to total infertility).
Sterility in females: primary or secondary amenorrhea; premature ovarian failure; gonadal dysgenesis. Usually the X chromosome is present in an abnormal number, in all cells or as a mosaic. See Table 2 for examples of chromosomal abnormalities found in 258 subfertile patients.
It is well to bear in mind that a single chromosome carries several thousand genes, and that monogenic disorders, or even those involving several genes

cannot be determined by cytogenetic techniques alone, the purpose or which is to detect major aberrations, involving chromosome number or structure.

Nonetheless, karyotyping remains the essential tool for analysis of some forms of reproductive difficulties, in prenatal screening, and in the diagnosis of a number of relatively frequent clinical syndromes.

References

Gelehter, T.D., and Collins, F.S. (1990): In: Principles of medical genetics. Williams & Wilkins, Baltimore.

Navarrete, C., and Salamanca F. (1986): Ann. Genet. (Paris), 29:98-103.

Thompson, M.W., McInnes, R.R., and Willard, H.F. (1991): In: Genetics in medicine. W.B. Saunders Co., Londres

Aberrant Karyotype and Chromosomal abnormalities in cancer

It is well known that many cancerous genomes exhibit abnormal karyotypes. The abnormalities found in these karyotypes include numerical aberrations, i.e. changes in chromosome copy number, and structural aberrations, i.e. rearrangements within the genome (see Fig. 1). Some of the malignancies, mostly hematological ones, are associated with specific patterns of aberrations. A classical example of such association is between the “Philadelphia chromosome” abberation (a specific translocation between chromosomes 22 and 9) and chronic myelogenous leukemia [17, 19]. This translocation leads to the formation of the oncogene BCR-ABL [5]. Fig. 1. A schematic view of an aberrant karyotype (produced by the SKYGRAM converter tool [1]). Chromosomes 1,14, and 18 show structural aberrations, and chromosome 18 shows a numerical aberration. (An ISCN description of this karyotypeis 47,XY,der(1)t(1,18)(p36;q21),t(14,18)(q32;q21),+der(18)t(12;18)(p11;q21),+ der(18)t(14;18)) Over the last few decades, intensive research on chromosomal abberations in cancer has led to the accumulation of large amount of data on cancerous karyotypes. The largest available public depository of 2 such data is the Mitelman database [15], which contains over 50,000 karyotypes collected

from over 8,000 publications. In this study we analyze this database. Our goal is to understand the main abberation types and their frequency in different cancers. Our hope is that such studies will provide insights and better understanding of the evolution of karyotypes in specific cancer types. Traditionally, karyotypes have been constructed using chromosome staining methods, mostly G-banding. SKY [22] and M-FISH [25] are relatively new molecular cytogenetic techniques that permit the simultaneous visualization of all the chromosomes in different colors, considerably improving the detection of material exchange between chromosomes. The Mitelman database contains primarily karyotypes based on G- banding. The resolution and the detectable level of details in such karyotypes is lower than what can be observed with SKY and M-FISH or with novel high throughput methods (e.g. array-based CGH [24] and ESP [26]). Nevertheless, we chose to focus on the Mitelman database since it is the largest collection of cancerous karyotypes. Karyotypes are usually described using the ISCN nomenclature [14]. In this system, every aberrant chromosome is described using specific rearrangement and numerical events, e.g., translocations, inversions, deletions, and duplications. Although ISCN attempts to describe the correct set of events leading to the observed karyotypes, it has almost no ability to do so when there are overlapping rearrangements, e.g. a chromosome involved in two translocations, each at a different position. Moreover, while the inference of the events is an easy task for many modestly rearranged karyotypes of hematological disorders, it can be a computationally hard task when the karyotypes are complex, as often happens in solid tumors. There are many computational studies analyzing large data sets of cancerous genomes. Most of these analyses consider a cancerous genome as a collection of chromosomal abberations easily computed from the data. For example, in a series of studies, reviewed in [12], H¨ogland et al. analyzed cytogenetic data from individual tumor types, by inspecting various parameters, including the number of gains or losses of genomic fragments, the number of aberrations, and the frequency at which bands are involved in breaks. In another study [21], Sankoff et al. compared the distributions of cancer-related breakpoints, derived from the Mitelman database, and evolutionary breakpoints, derived from a human-mouse comparative map. Another important branch of computational studies searches for statistical dependencies between chromosomal aberrations, usually in the form of tree or directed acyclic graph, such as [6, 7, 12, 11]. Chromosomal aberrations observed in cancer are by and

large somatic and thus non-inheritable. When a rearrangement occurs in a genome of a germ-line cell, it can be inherited by offsprings. Indeed, the comparison of genomes of related species reveals that genome rearrangements play a significant role during the evolution of species. In a pioneering paper [20], Sankoff raised the problem of computing a shortest sequence of rearrangement operations between two given genomes, when genomes are represented by linear orders of oriented genes. Over the last fifteen years, this problem was intensively studied for many types of rearrangement events and their combinations, including inversions, translocations, block exchanges, deletions and insertions (see [4] for a review). All these studies ignored the ploidy in the genomes, i.e., the number of copies of each chromosome. Since numerical aberrations are prevalent in cancer, every model of cancer rearrangements must contain both numerical and structural events. This makes the reconstruction task more complicated and prevents direct use of results from the rich algorithmic literature on germ- line rearrangements. The main purpose of this study was to estimate the prevalence of specific types of genome rearrangement events in cancer karyotypes. For this purpose we developed a new efficient heuristic for reconstructing a sequence of events that best explain the transformation from the normal karyotype into a given cancer karyotype. We applied this algorithm to over 40,000 karyotypes published in scientific literature, and collected statistics on event frequency across cancer types. The algorithm is deliberately simplistic, mimicking the process of detecting obvious events and "undoing" them, going back from the given karyotype towards the normal. As such, it does not guarantee finding the shortest solution or finding any solution. However, we reasoned that most reported karyotypes are of limited complexity and thus may be amenable to such approach. Reassuringly, over 98% of the karyotypes were solved by this method. Our study provides for the 3 first time a broad picture of event frequency in hematological and solid cancers. Our analysis shows that chromosome gains and losses, reciprocal translocations, and terminal deletions, dominate the evolution of cancer karyotypes. By using the event frequencies in each karyotype as its profile, we show that many different cancer types have clearly distinguishable profiles, which can be meaningful for further understanding of the cancers. This paper is organized as follows. In Section 2 we provide a short background on chromosome aberrations in cancer. In Section 3 we present some basic statistics regarding the complexity of cancer karyotypes. In Section 4 we describe our

heuristic for reconstructing genome rearrangement events for a given karyotype. The analysis of the reconstructed events is reported in Section 5. For lack of space, some details are deferred to an appendix. 2 Background 2.1 Mechanisms for chromosomal aberrations Many molecular mechanisms are involved in the formation of chromosomal aberrations. The following mechanisms are reviewed in [2, 9, 16, 18]. A double strand break (DSB) is one of the frequent lesions in DNA. The repair of DSBs in eukaryotic cells is carried out by two main pathways: non-homologous end joining (NHEJ) and homologous recombination (HR). NHEJ repairs DSBs by directly re-ligating DNA ends, which may create a deletion if sequences surrounding the lesion were lost. Another potential risk of NHEJ is the ligation of two non-matching broken ends, leading to genome rearrangements. HR repairs breaks through interaction of a free DNA end with an intact homologous sequence, which is used as a template to copy missing information prior to religation. Because of the ability to fill in gaps by copying information from a sister chromatid or homologous chromosome, HR runs the risk of generating rearrangements through interaction of similar sequences on non-homologous chromosomes or regions. In particular, HR may extend to the end of a chromosome, resulting in a duplication of the whole "tail" of that chromosome. Another possible lesion to the DNA is the loss of a telomere. The telomeres protect the ends of chromosomes from fusion with other ends. Thus a chromosome end that lacks a functioning telomere tends to be adhesive and may initialize a breakage-fusion-bridge process [13]. Stabilization of the genome occurs only through the net gain of a telomere, either through duplications of protected chromosome ends, or by direct telomere addition. Indeed, telomerase activity has been detected in the majority of malignant epithelial tumors [8]. A direct cleavage through a centromere generates two telocentric (i.e. single-arm) chromosomes, each containing a portion of the kinetochore (the functional component of an active centromere). Non-disjunction of sister chromatids of a telocentric chromosome results in the formation of an isochromosome or isoderivative, i.e. a Chromosome with two identical, mirror-image arms. As elaborated above, DSBs, telomeres dysfunction and centric fissions may lead to structural aberrations. Numerical aberrations may occur when genes involved in chromosome segregation or cytokinesis are deregulated. In particular, failure in cytokinesis (e.g. endomitosis) and multipolar mitoses may alter the ploidy of the genome. 2.2 The Mitelman database The "Mitelman database of chromosome aberrations in

cancer" [15] (henceforth abbreviated MD) contains the description of cancer karyotypes manually culled from the literature over the last twenty years. For our analysis we used the version of March 27, 2007, which contained 53,573 cancerous karyotypes, collected from 8748 published studies. The karyotypes in the database are represented in the ISCN format and can be automatically parsed and analyzed by the software package CyDAS [10]. We shall use here a simplified 4 version of ISCN for representing karyotypes (see Appendix A). We refer to a karyotype as valid if it can be parsed by CyDAS without any errors. According to our processing, 47,045 (87.8%) of the records were valid karyotypes. 2.3 Complex karyotypes When the cytogeneticist analyzes a sample, several cells are checked. Each abberation described in a cancerous karyotype must be present in at least two cells in the described sample. In some cases the cell population may be non-homogeneous, and contain cells with several distinct karyotypes, resulting from evolution of the cell population during the development of the cancer. A homogeneous cell sample is described by a simple karyotype, and a non-homogeneous one has a complex karyotype, which consists of several karyotype species. In this study we derive simple karyotypes from complex karyotypes and analyze each of them independently. About 17% of all valid karyotypes in MD are complex. The total number of simple (valid) karyotypes that we deduced from MD is 57941 (33% of which originate from complex karyotypes). For the rest of this paper we assume that every analyzed karyotype is simple. 3 Basic statistics on karyotype complexity In this section we present some simple statistics based on the MD regarding the complexities of cancerous karyotypes. Human malignancies can be divided into two main categories: hematological disorders and solid tumors. Our first step was to distinguish between hematological malignancies and solid tumors. The type of neoplasia can be identified by its morphology, i.e. the cancer classification based on neoplasm histology, and its topography, i.e. the tumor site (applicable only for solid tumors). Based on the morphology and topography descriptors of each karyotype, we partitioned the karyotypes in the database into three categories: • HEMA: hematological neoplasms, e.g.: leukemia, myeloma, lymphoma. • BENIGN: solid benign tumors, e.g.: meningioma, leiomyoma, lipoma. • SOLID: solid malignant tumors, e.g.:adenocarcinoma, Wilms tumor, malignant melanoma. The HEMA category covers 71.2% of the valid simple karyotypes derived from the MD, while SOLID and BENIGN cover only 22.9% and 5.9% respectively. In the following,

we compare the distributions of simple variables defined on karyotypes between these categories. We define a chromosome as abnormal if it does not match any chromosome in the standard normal karyotype. As expected, the distribution of the number of abnormal chromosomes per karyotype had the longest tail for solid tumors, while benign and hematological karyotypes seldom have more than five abnormal chromosomes (Fig. 5-a). The number of fragments (maximal contiguous interval in the normal) per an abnormal chromosome (Fig. 5-b) had a similar distribution across categories, with less than 1% of the abnormal chromosomes having four or more fragments. We defined karyotype ploidy level as b n+11 23 c, where n is the total number of chromosomes. As expected, solid tumors tended to have higher ploidy, reflecting their higher complexity (Fig. 5-c). Multicentric chromosomes (i.e. chromosomes with more than one centromere) are considered non-stable, as each of the centromeres in these chromosomes may be passed to opposite poles in the mitotic anaphase. Interestingly, all three categories had some 2-4% of karyotypes with multicentric chromosomes (Fig. 5-d). Overall, the difference between the categories are quite subtle. Karyotypes of solid tumors, in particular malignant solid tumors, tend to have more complex abnormal chromosomes and ploidy changes, in comparison to hematological malignancies. Do the statistics above - as well as those we shall report later - reflect the distributions of properties in cancer karyotypes “in the real world”? The answer is probably no. For example, although up to 80% of all human malignancies are

solid, most of the karyotypes in MD belong to hematological malignancies. 5 One major reason for this bias is the difficulty in cytogenetically analyzing solid tumors. Solid tumor genomes often demonstrate poor visual quality during metaphase. Moreover, the karyotypes of solid tumors are often much more complex and thus more difficult to interpret. In addition, the database contains reported karyotypes from the literature, and there is a bias in this reporting. For example, the hematological karyotypes in MD are probably of higher complexity than those simple cases seen regularly in the clinic, which are not deemed publish-worthy as they are too simple or fully understood. While this means that the statistics we are collecting should be interpreted with caution, we believe they can still be useful in understanding how to model cancer evolution on the karyotype level and how different classes and subclasses differ.

1. NCI and NCBI's SKY/M-FISH and CGH Database, 2001. http://www.ncbi.nlm.nih.gov/sky/skyweb.cgi. 2. D.G. Albertson, C. Collins, F. McCormick, and J. W. Gray. Chromosome aberrations in solid tumors. Nature Genetics, 34:369–376, 2003.

3. V. Bafna and P. A. Pevzner. Genome rearragements and sorting by reversals. SIAM Journal on Computing, 25(2):272–289, 1996.

4. G. Bourque and L.Zhang. Models and methods in comparative genomics. Advances in Computers, 68:60–105, 2006.

5. A. de Klein et al. A cellular oncogene is translocated to the philadelphia chromosome in chronic myelocytic leukaemia. Nature, 300:765–767, 1982.

6. R. Desper, F. Jiang, O. Kallioniemi, H. Moch, C. Papadimitrou, and A. Sch¨affer. Inferring tree models for oncogenesis from comparative genome hybridization data. Journal of Computational Biology, 6:37–51, 1999. 1. NCI and NCBI's SKY/M-FISH and CGH Database, 2001. http://www.ncbi.nlm.nih.gov/sky/skyweb.cgi. 2. D.G. Albertson, C. Collins, F. McCormick, and J. W. Gray. Chromosome aberrations in solid tumors. Nature Genetics, 34:369–376, 2003. 3. V. Bafna and P. A. Pevzner. Genome rearragements and sorting by reversals. SIAM Journal on Computing, 25(2):272–289, 1996. 4. G. Bourque and L.Zhang. Models and methods in comparative genomics. Advances in Computers, 68:60–105, 2006. 5. A. de Klein et al. A cellular oncogene is translocated to the philadelphia chromosome in chronic myelocytic leukaemia. Nature, 300:765–767, 1982. 6.

R. Desper, F. Jiang, O. Kallioniemi, H. Moch, C. Papadimitrou, and A. Sch¨affer. Inferring tree models for oncogenesis from comparative genome hybridization data. Journal of Computational Biology, 6:37–51, 1999.

Genetics of foetal wastage

of chromosomes and have ignored the issue of possible effects within the chromosome pairs.

This study has involved measuring heterochromatic regions from chromosomes 1, 9 & 16 in giemsa banded preparations from a control group of 100 fertile but otherwise unselected women and a test group of 54 women with histories of recurrent abortion. The measurements were adjusted for contraction differences between cells. For each chromosome pair, the larger heterochromatic region, the smaller heterochromatic region and the pair difference was computed for each individual. The 3 groups of measurements were then compared between the control and test group for each chromosome, using a Mann-Whitney test. No significant differences were found for chromosomes 1 and 16, but chromosome 9 gave statistically significant results for 2 of the 3 criteria. That is for chromosome 9, the larger chromosomes were significantly larger, and the pair difference was significantly greater.

The implications of this finding will be discussed.

CHROMOSOME 15 DELETION IN THE PRADER-WILLI SYNDROME

J. H. FORD, L. J. SHEFFIELD & D. I. WHITE *Cytogenetics Unit, The Queen Elizabeth Hospital, Medical Genetics Department, Adelaide Children's Hospital, Intellectually Retarded Services, Toorak Gardens*

The Prader-Willi syndrome is a relatively common dysmorphic syndrome characterized by mental retardation, muscular hypotonia, obesity, short stature and hypogonadism. Routine cytogenetic analysis is usually normal; but recently, studies of extended chromosomes (prophase or prometaphase) have suggested that a small deletion in the long arm of chromosome 15 may be found in patients with this syndrome.

Our study is of 9 patients, 4 with a definite diagnosis of P.W. syndrome, and 5 with suspected but unconfirmed diagnosis of the syndrome. Most patients had been karyotyped previously and had been reported as having normal chromosomes. Prometaphase chromosome preparations were examined on each patient and particular attention was given to band 15q12. This band was found to be absent in all patients with a definite diagnosis of P.W. syndrome and in 2 cases with a doubtful diagnosis.

This study then confirms that an absence of the band in the region of 15q12 is found in the Prader-Willi syndrome if suitable cytogenetic analysis is undertaken. Such a finding can be taken as positive confirmation of the diagnosis of Prader-Willi syndrome and may allow for the initiation of early dietary treatment.

A 'NEW' SYNDROME OF CONGENITAL AMEGAKARYOCYTIC THROMBOCYTOPENIA AND MULTIPLE MALFORMATIONS

R. J. M. GARDNER,* P. S. MORRISON† and G. D. ABBOTT‡ **Department of Paediatrics and Child Health, University of Otago, Dunedin, †Timaru Hospital, Timaru and ‡Department of Paediatrics and Child Health, Christchurch Clinical School of the University of Otago, Christchurch*

Three siblings, 2 dying as neonates, one surviving to 22 mth, had congenital amegakaryocytic thrombocytopenia and multiple

consuming a standard dose of alcohol (0.75 g/kg body weight). Frequent measurements of breath and blood alcohol content were taken during this time.

The results have demonstrated considerable population variation in susceptibility to alcohol. There are also large individual differences in aspects of alcohol metabolism monitored by blood ethanol concentrations (BAC) e.g.: (i) time to peak BAC; (ii) peak BAC; and (iii) rate of decline in BAC.

This variation has been investigated in an attempt to partition environmental components specific to the individual, common to members of a family or due to genetic variation.

Analyses of these data with respect to the genetic components of deterioration in performance at some psychomotor tasks will be described in Dr Martin's paper.

THE GENETICS OF FETAL WASTAGE

BARBARA K. HECHT & FREDERICK HECHT *The Genetics Center, Southwest Biomedical Research Institute, Tempe, Arizona 85281, U.S.A.*

Studies of fetal wastage have provided certain facts and raised new questions. For example:

Fact	Chromosome studies of spontaneous abortions show frequent aneuploidy.
Questions	What are the risks for a couple who have had an aneuploid miscarriage to have an aneuploid abortion or a liveborn aneuploid child with the next pregnancy?
Fact	Chromosome studies of couples with 2 or more spontaneous abortions may disclose one parent with an 'unusual' balanced translocation not previously reported.
Questions	How can risks for different translocations be determined? Does a given translocation segregate in the same fashion within 1 family? Or in different families?
Fact	Chromosome studies of couples with 2 or more spontaneous abortions may disclose one parent with a Robertsonian translocation.
Questions	Is the finding of the Robertsonian translocation serendipitous? Or meaningful? Are accepted empiric risk figures applicable?
Fact	Chromosome studies of a woman who had repeated miscarriages show X-chromosome mosaicism.
Questions	What does this mean? Is this significant? Should future pregnancies be monitored cytogenetically?

It is admittedly difficult to monitor every pregnancy in every family with recurrent fetal wastage. It should be done.

SPHINGOMYELIN STORAGE DISEASE: A NEW CLASSIFICATION

A. POULOS,* J. W. CALLAHAN† & A. B. LOWDEN† **Department of Chemical Pathology, Adelaide Children's Hospital and †Department of Neurosciences, Research Institute, Hospital for Sick Children,*

Pharmacogenetics,

Pharmacogenetics is the study of how people respond differently to drug therapy based upon their genetic makeup or genes. Diet, overall health, and environment also have significant influence on medication response, but none are stronger indicators of how you will process medication than your genetics.

From the moment medications enter your body, the body is working to actively process or metabolize them.

All drugs will eventually leave the body by a process called elimination– but the time that they stay active, in your bloodstream working, is often determined by genetic variations that change the way your drug-processing enzymes work.

Link:

How Providers use Pharmacogenetics | History of Pharmacogenetics | Pharmacogenetics Today | How does Pharmacogenetics Work? | Pharmacogenetics and Drug-to-Drug Interactions | Glossary of PGx Terms

How Providers use Pharmacogenetics

Utilizing Pharmacogenetics allows a healthcare provider to choose the right drug and dose that are likely to work best for each individual patient. Tailoring a patient's medication to their unique genetic characteristics may one day replace the one-size-fits-all approach to drug selection and dosing that is commonly used today.

History of Pharmacogenetics

Pharmacogenetics is not new - An article was published in 1957 by a geneticist who noted that adverse reactions to an anti-malarial drug and a muscle relaxant were inherited and linked to deficiencies in the activity of specific liver enzymes which were responsible for the breakdown or metabolism of those particular drugs. This article established the link between genetics and the enzymes that break down medications while simultaneously establishing the link between that process and adverse reactions to the medications themselves.

The history of Pharmacogenetics may well date back to 510 B.C. when

Pythagoras established a link between the eating of fava beans to the development of hemolyticanemia. This observation was later tested with modern scientific equipment and it was found that certain people, usually males, lack an enzyme which is involved in the stability of the red blood cell membrane. This deficiency is further exaggerated by the consumption of fava beans, confirming Pythagoras's initial observations.

prescription medications contain Pharmacogenetic information in their FDA approved labels.

The label's information contains the identification of biomarkers – the primary measurable indicators associated with a patient's specific condition. The labels also identify targeted drug therapy specific for that genetic abnormality as well as list the genetic variations which may influence how a drug is metabolized or broken down in the body and is likely to cause a significant adverse event. The Clinical Pharmacogenetics Implementation Consortium or CPIC has created a searchable list of all medications with known pharmacogenetic implications and prescribing guidelines that can be found here.

Healthcare providers may use pharmacogenetic information to help inform the most appropriate treatment for each individual patient. This includes choosing a drug that is more likely to work, avoiding drugs that may cause side effects, adjusting the starting dose of a drug where appropriate, or determining whether closer monitoring of the drug's effect is needed.

How does pharmacogenetics work?

The cytochrome P450 system are a family of enzymes found throughout the body which are responsible for the synthesis and metabolism of various molecules and chemicals within the cell, most notably including the active ingredient of mostdrugs. Common variations - known as polymorphisms - in the genes that determine cytochrome P450 enzyme activity may affect the function of the enzymes. These are most commonly seen in the breakdown or metabolism of medication. Drugs may be metabolized quickly or slowly. If a cytochrome P450 enzyme metabolizes a drug slowly, the drug remains active longer and a lower dose is

needed to get the desired effect whereas normal doses may cause toxicity. Cytochrome P450 enzymes, particularly CYP2C9, CYP2C19, and CYP2D6, are responsible for approximately 70% of drug metabolism in the body. Additionally, there are other genes outside of the cytochrome-p450 system that affect drug metabolism and as a result - a patient's response to medications.

Pharmacogenetic testing is primarily concerned with variations in enzymes that affect drug metabolism.

These variations are broken down into four categories based on their expected effect on drug metabolism:

- Poor
- Intermediate
- Normal
- Rapid / Ultra Rapid

Pharmacogenetics and Drug-to-Drug Interactions

Inducers are substances that affect gene expression. For example: if a drug is a CYP2D6 inducer, this will increase CYP2D6 activity which changes the way other drugs that rely on this enzyme are metabolized. Inhibitors function in the opposite way, decreasing the activity of that enzyme and potentially altering the metabolism of drugs that rely on that enzyme. This illustrates the need for an all-encompassing look at FDA approved drug-to-gene and drug-to- drug interactions.

Being aware of these inducers and inhibitors as well as monitoring for changes in the cytochrome P450 enzymes helps make pharmacogenetics a powerful force in understanding drug metabolism.

Strong inducers include:

- rifampin (CYP2C19 and CYP3A)
- phenytoin (CYP3)

Strong inhibitors include:

- ciprofloxacin (CYP1A2)
- clopidogrel (CYP2C8)
- fluoxetine (CYP2C19 and CYP2D6)

danoprevir, ritonavir, itraconazole, clarithromycin, and grapefruit juice. (CYP3A)

Glossary of Important PGx Terms

Pharmacogenetics – the study of how people respond differently to drugs based upon their genetic makeup or genes.

Genes – basic units of DNA within the cell that play an important role in heredity like determining physical traits such as eye color.

Adverse Drug Events (ADEs) – an unintended side effect caused by a medication at the time it is used.

FDA (Food and Drug Administration) – government department responsible for protecting and promoting public health through control and supervision of prescription and over-the- counter medications. Also responsible for food safety, dietary supplements, vaccines, cosmetics, and medical devices as well as other products.

HemolyticAnemia – abnormal, early breakdown of red blood cells which may be caused by a medication reaction and lead to a low blood count. Other causes include hereditary abnormalities such as sickle cell disease, cell breakdown by artificial heart valves and very high blood pressure.

G6PD (glucose-6-phosphate dehydrogenase) deficiency – a genetic disorder seen primarily in males which causes the early or premature breakdown of red blood cells leading to hemolyticanemia.

The Clinical Pharmacogenetics Implementation Consortium (CPIC) – an international organization interested in facilitating use of pharmacogenetic tests for patient care.

Cytochrome P450 System – a group of enzymes involved in drug metabolism found in high levels in the liver. These enzymes change drugs into less toxic forms that are easier for the body to eliminate or excrete.

CYP3A4/5, CYP2C9, CYP2C19, CYP2D6 – isoenzyme systems that are part of the cytochrome P450 system and found primarily in the liver and intestines. Responsible for breaking down (metabolizing) nearly 70% of the medications we take.

Inducers – a drug which increases the activity of the enzymes of the cytochrome P450 system resulting in a decrease in the effect of certain other drugs. A dose increase of the affected drug may be necessary.

Inhibitors – a drug which decreases the activity of the enzymes of the cytochrome P450 system resulting in an increase in the effect of certain other drugs. A decrease in the dose of the affected drug may be necessary.

Poor Metabolizers – a person who breaks down (metabolizes) a drug very slowly causing a buildup of the drug within the body and potential toxicity.

Intermediate Metabolizers – a person who metabolizes a drug at a rate somewhere between a poor and extensive metabolizer. May potentially cause a buildup of the drug within the body and potential toxicity.

Normal Metabolizers – the most common type of metabolizer; one who breaks down (metabolizes) a drug at the expected or normal rate.

Rapid Metabolizers – a person who breaks down (metabolizes) a drug so fast that it will not reach optimal blood levels leading to lower than expected drug levels and an inadequate response to the drug.

Ultra-Rapid Metabolizers – a person who breaks down (metabolizes) a drug so fast that they receive no benefit from a standard dose of the drug.

Ecogenetics

Ecogenetics is a branch of genetics that studies genetic traits related to the response to environmental substances.[1] Or, a contraction of ecological genetics, the study of the relationship between a natural population and its genetic structure.[citation needed]

Ecogenetics principally deals with effects of preexisting genetically-determined variability on the response to environmental agents.[2] The word environmental is defined broadly to include the physical, chemical, biological, atmospheric, and climate agents. Ecogenetics, therefore, is an all-embracing term, and concepts such as pharmacogenetics are seen as subcomponents of ecogenetics. This work grew logically from the book entitled Pollutants and High Risk Groups (1978), which presented an overview of the various host factors i.e. age, heredity, diet, preexisting diseases, and lifestyles which affect environmentally-induced disease.

The primary intention of ecogenetics is to provide an objective and critical evaluation of the scientific literature pertaining to genetic factors and differential susceptibility to environmental agents, with particular emphasis on those agents typically considered pollutants. It is important to realize though that ones genetic makeup, while important, is but one of an array of host factors contributing to overall adaptive capacity of the individual. In many instances, it is possible for such factors to interact in ways that may enhance or offset the effect of each other.

Red blood cell conditions There is a broad group of genetic diseases that result in either producing or predisposing affected individuals to the development of hemolyticanemias. These diseases include abnormal hemoglobin, inability to manufacture one or the other of the peptide globin chains of the hemoglobin, and deficiencies of the Embden-Meyerhoff monophosphate.

Liver metabolism Individuals lacking the ability to detoxify and excrete PCB's may have a high risk of total liver failure in conjunction with certain ecological conditions.

Cardiovascular diseases The pathologic lesion of atherosclerosis is a plaque-like substance that thickens the innermost and middle of the three layers of the artery wall. The thickening of the intimal and medial layers results from the accumulation of the proliferating smooth muscle cells that are encompassed by interstitial substances such as collagen, elastin, glycosaminoglycans, and fibrin.

Respiratory diseases There are three genetically-based respiratory diseases that can directly correspond with ecological functions and induce disease. These include lung cancer and the upper and lower respiratory tract associated with a serum Ig A deficiency.

QUESTIONS FOR PRACTICE

UNIT - 5

6 marks

1. Brief note on human cytogenetics and its importance.
2. Brief note on Pharmacogenetics, Ecogenetics and Teratogenetics.
3. Describe chromosomal banding with illustration.
4. Explain the techniques involved in Chromosomal banding.
5. Describe the pathology of human chromosomes.
6. Elaborate the term ‘Chromosome painting’ with illustration.

10 marks

1. Explain the techniques in human chromosome analysis.
2. Explain human karyotyping and nomenclature of chromosome banding.