Population Genetics and Pharmacogenetics

Ken K. Ong^{a,b} and Cathy Elks^a

^aMedical Research Council Epidemiology Unit, Cambridge, and

^bDepartment of Paediatrics, University of Cambridge, UK

Rapid technological advances continue to pave the way for exciting genetic discoveries. Just a handful of years since the appearance of genome-wide association studies that have rapidly expanded the number of common variants robustly linked to multigenic diseases and traits, whole-genome and whole-exome sequencing studies are now demonstrating their power to shed novel insights into human genetic architecture and disease etiology. This year's review starts with four such studies all published during the first few months of 2010.

Genome-wide association studies continue to generate dozens of new loci for common diseases and continuous traits, and examples described below include fasting glucose, insulin resistance and birth weight. Novel analytical approaches to genome-wide association data have described variants with parent-of-origin associations with both type 1 and type 2 diabetes. Extension of genetic studies to African-derived populations with greater power for fine mapping has finally pinpointed the first genome-wide obesity locus to the FTO gene.

However, while the limits of the genome are now clearly definable and we have the tools to see it in its entirety, the percentage of disease susceptibility that can be explained remains surprisingly small. Whether you see this position as a future opportunity or as an indictment against the current efforts, it is clear that we will need new analytical strategies and likely new large family-based studies to decipher the remaining genetic basis of human phenotypic variation and disease risks.

Mechanism of the year The clinical utility of genome sequencing

Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy

Lupski JR, Reid JG, Gonzaga-Jauregui C, Rio DD, Chen DC, Nazareth L, Bainbridge M, Dinh H, Jing C, Wheeler DA, McGuire AL, Zhang F, Stankiewicz P, Halperin JJ, Yang C, Gehman C, Guo D, Irikat RK, Tom W, Fantin NJ, Muzny DM, Gibbs RA

Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Tex., USA

N Engl J Med 2010;362:1181–1191

Background: Whole-genome sequencing may revolutionize medical diagnostics through rapid identification of alleles that cause disease. However, even in cases with simple patterns of inheritance and unambiguous diagnoses, the relationship between disease phenotypes and their corresponding genetic changes can be complicated. This study aimed to assess the usefulness of human whole-genome sequencing for genetic diagnosis in a patient with Charcot-Marie-Tooth disease.

Methods: A family was identified with a recessive form of Charcot-Marie-Tooth disease of yet unknown genetic basis. The whole genome of the proband was sequenced and potential functional variants were genotyped in other family members.

Results: Two compound, heterozygous, causative mutations in SH3TC2 (the SH3 domain and tetratricopeptide repeats 2 gene) were identified and validated. Separate subclinical phenotypes segregated independently with each of the two mutations; heterozygous mutations confer susceptibility to neuropathy, including the carpal tunnel syndrome.

Conclusion: This study of a family with Charcot-Marie-Tooth disease shows that whole-genome sequencing can identify etiological variants and provide clinically relevant diagnostic information.

Personalised medicine is used to include a patient's lifestyle, family history, and environment. This and the following articles illustrate the potential that rapid-sequencing technology could have for personalised medicine. While several whole genomes of healthy individuals have been published, this study is one of the first to apply whole-genome sequencing in a clinical context. Beyond the acquisition of an accurate genome sequence, two major difficulties in interpreting the data may quickly become apparent. Firstly is the realisation of non-classical mendelian inheritance patterns, involving more than one mutation in the same or even different genes; these are compound (as illustrated in this study), digenic, and trigenic inheritance patterns, and the presence of modifying genes.

Secondly, many seemingly deleterious mutations will be found in each individual, and most of these will be silent or 'non-penetrant'. In this patient, whole-genome sequencing unearthed over 3 million variants compared to the human genome reference sequence, of which 1 million were within gene regions. 9,069 were predicted to be amino-acid changing, and 121 were non-sense mutations, i.e. predicted to lead to truncation of the protein. 21 have been described to cause mendelian disease. For example the (male) proband had a mutation in ABCD1 which causes the X-linked disorder adrenoleukodystrophy (which he did not have!). The authors therefore focussed only on the 54 coding SNPs in the 40 genes with previous links to neurological conditions. Two SNPs were in *SH3TC2* and one showed complete segregation with disease status in other family members.

Sequencing the whole genome to find one or two causative mutations may seem like a sledgehammer to crack a small nut, particularly if most of the data is uninterpretable. However this sledgehammer is rapidly becoming lighter due to technological advances. Even within the 6-month duration of this study, the sequence yield increased threefold with no appreciable increase in cost. The estimated cost per genome in April 2009 was USD 50,000, in contrast to around USD 15,000 for a diagnostic panel for Charcot-Marie-Tooth covering only 15 genes. Current costs for whole-exome sequencing are only about USD 4,000, and this approach could have led to the same conclusion – see article below. Other studies described below illustrate further ways to interpret the data.

Exome sequencing identifies the cause of a mendelian disorder

Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ

Department of Genome Sciences, University of Washington, Seattle, Wash., USA

Nat Genet 2010;42:30–35

Background: To date, allelic variants underlying fewer than half of all monogenic disorders have been discovered. This study demonstrates the first successful application of exome sequencing to discover the gene for a rare mendelian disorder of unknown cause, Miller syndrome.

Methods: For 4 affected individuals in three independent kindreds, gene-coding regions were captured and sequenced to a mean coverage of 40× and sufficient depth to call variants at approximately 97% of each targeted exome. Filtering against public SNP databases and eight HapMap exomes for genes with two previously unknown variants in each of the 4 individuals.

Results: A single candidate gene, *DHODH*, was identified which encodes a key enzyme in the pyrimidine de novo biosynthesis pathway. Sanger sequencing confirmed the presence of *DHODH* mutations in three additional families with Miller syndrome.

Conclusion: Exome sequencing of a small number of unrelated affected individuals is a powerful, efficient strategy for identifying the genes underlying rare mendelian disorders and will likely transform the genetic analysis of monogenic traits.

The justification for the smaller 'whole-exome', rather than 'whole-genome', sequencing approach is that the majority of known variants for monogenic disorders disrupt the protein-coding sequence. By focusing on exons, plus splice acceptor and donor sites, only ~5% of the whole-genome sequencing load is required. Yet, this is still vast, and still results in a huge number of potential causal variants to sift through. The authors found non-synonymous, splice site or insertion/deletion variants in ~4,600 genes!

Several factors increased the chances of successfully choosing the causal gene in this landmark study. Firstly, the authors chose a recessive disorder, which meant they could focus on those genes with at least two variants. Secondly, as Miller syndrome is so rare, with only ~30 cases to date, the authors argued that they could safely filter out all the variants found in existing databases. Thirdly, they

sequenced 2 affected brothers, which further narrowed the search to those variants that were present in both exomes. The authors found further mutations in the dihydroorotate dehydrogenase gene (DHODH) in several other probands.

The characteristic features of Miller syndrome include severe micrognathia, cleft lip and/or palate, hypoplasia or aplasia of the posterior elements of the limbs, and coloboma of the eyelids. It was therefore a surprise to locate gene mutations to DHODH, which encodes an enzyme that catalyzes the conversion of dihydroorotate to orotic acid, as the authors point out that inborn errors of metabolism rarely cause birth defects. The paper includes a thought-provoking discussion linking inhibitors of pyrimidine biosynthesis (such as methotrexate) to birth defects via TNF- α , NF- κ B and Shh (the sonic hedgehog homolog). Intriguingly, the brothers were also compound heterozygotes for variants in DNAH5, which encodes a ciliary dynein heavy chain, and they were consequently diagnosed with primary ciliary dyskinesia. While this explained their non-Miller syndrome features of recurrent lung infections and bronchiectasis, it illustrates the potential for surprise genetic findings with clinical, and even ethical, consequences. We used to consider multiple genetic defects unique to consanguineous families. It may be more common than we expect.

Analysis of genetic inheritance in a family quartet by whole-genome sequencing

Roach JC, Glusman G, Smit AF, Huff CD, Hubley R, Shannon PT, Rowen L, Pant KP, Goodman N, Bamshad M, Shendure J, Drmanac R, Jorde LB, Hood L, Galas DJ
Institute for Systems Biology, Seattle, Wash., USA
Science 2010;328:636–639

Background: Whole-genome sequences from 4 members of a family represent a qualitatively different type of genetic data than whole-genome sequences from individual or sets of unrelated genomes.

Methods: Whole-genome sequences of a family of 4, consisting of 2 siblings and their parents, were analyzed. Family-based sequencing allowed the precise delineation of recombination sites, the identification of 70% of the sequencing errors (resulting in >99.999% accuracy), and the identification of very rare single-nucleotide polymorphisms. The directly estimated human intergeneration mutation rate was approximately 1.1×10^{-8} per position per haploid genome. Both offspring in this family have two recessive disorders: Miller syndrome, for which the gene was concurrently identified, and primary ciliary dyskinesia, for which causative genes have been previously identified. Family-based genome analysis narrowed the candidate genes for both of these mendelian disorders to only four.

Conclusions: These results demonstrate the value of complete genome sequencing in families. Such data, along with relevant environmental and medical information, will characterize the integrated medical records of the future.

This paper, from the same group who reported the genetic etiology of Miller syndrome, is based on that same family and now includes their parents' genomes. They show for the first time how whole genomes from the same family permit powerful inheritance analyses to greatly improve the detection and understanding of sequencing errors, and allow precise estimates of human mutation rates and the precise locations of recombination events. The 4 family members showed different genotypes at 3.6 million known variants, and these formed the basis for inheritance analyses. The authors describe how in a nuclear family of 4, each variant position can be allocated to one of four inheritance states with the 2 children receiving (i) the same allele from both the mother and the father (identical), (ii) the same allele from the mother but opposites from the father, (iii) the same allele from the father, but opposites from the mother, or (iv) opposites from both parents. Adjacent variants with the same inheritance states delineate inheritance blocks, and recombination must have occurred between these blocks. In contrast, blocks of variants that do not conform to these inheritance patterns (mendelian inheritance errors) must be either sequencing errors or result from hemizygous deletions.

Furthermore, while the authors identified the gene for Miller syndrome in their earlier paper, the current study illustrates how family-based data would have markedly narrowed the search for the disease-causing variant. They conclude that 'When the cost of recruiting additional families is expensive relative to sequencing costs, sequencing genomes of families will be an economical strategy for the identification of many disease-causing genes.'

Clinical assessment incorporating a personal genome

Ashley EA, Butte AJ, Wheeler MT, Chen R, Klein TE, Dewey FE, Dudley JT, Ormond KE, Pavlovic A, Morgan AA, Pushkarev D, Neff NF, Hudgins L, Gong L, Hodges LM, Berlin DS, Thorn CF, Sangkuhl K, Hebert JM, Woon M, Sagreiya H, Whaley R, Knowles JW, Chou MF, Thakuria JV, Rosenbaum AM, Zaranek AW, Church GM, Greely HT, Quake SR, Altman RB

Center for Inherited Cardiovascular Disease, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, Calif., USA

euan@stanford.edu

Lancet 2010;375:1525–1535

Background: The cost of genomic information has fallen steeply, but the clinical translation of genetic risk estimates remains unclear. This study aimed to undertake an integrated analysis of a complete human genome in a clinical context.

Methods: A 40-year-old man with a family history of vascular disease and early sudden death underwent a routine clinical assessment plus analysis of his full genome sequence. In addition to identifying risk variants associated with mendelian disease, post-test probabilities of disease were estimated by applying likelihood ratios derived from integration of multiple common variants to age-appropriate and sex-appropriate pre-test probabilities.

Results: Analysis of 2.6 million single nucleotide polymorphisms and 752 copy number variations showed increased genetic risk for myocardial infarction, type 2 diabetes, and some cancers. Rare variants were found in three genes that are clinically associated with sudden cardiac death – TMEM43, DSP, and MYBPC3. A variant in LPA was consistent with a family history of coronary artery disease. The patient had a heterozygous null mutation in CYP2C19 suggesting probable clopidogrel resistance, several variants associated with a positive response to lipid-lowering therapy, and variants in CYP4F2 and VKORC1 that suggest he might have a low initial dosing requirement for warfarin.

Conclusions: Although challenges remain, these results suggest that whole-genome sequencing can yield useful and clinically relevant information for individual patients.

While other studies have applied genome sequencing to rare mendelian diseases, this is the first to explore its utility in risk prediction for more complex multigenic disorders. Other than his family history, the patient was healthy and was found to have increased risks for coronary artery disease and sudden cardiac death, and his genome was informative for future drug choices and dosing. Remarkably, this assessment was performed on just 2 ml of whole blood.

The authors relied on disease-specific mutation databases and pharmacogenomic databases to identify genes and mutations with known associations with disease and drug response. There were compelling findings related to drug responses, including 63 known pharmacogenomic variants that could affect the patient's response to commonly used drugs. For example, he is likely to respond well to statins and to be at lower risk of statin-induced myopathy. Were he to need warfarin, his initial doses are likely to be low; by contrast, clopidogrel might be less effective. This approach assumes that all the identified variants are expressed and this contrasts with the findings of other studies, which showed that we each carry many mutations for mendelian disease that are somehow silenced. Nevertheless, with increasing knowledge gained from genome-wide association studies and whole-genome and whole-exome sequencing studies, the clinical value of sequence data will rapidly increase.

However, clinical and ethical dilemmas need to be tackled. The patient was found to have potentially damaging variants associated with haemochromatosis and also parathyroid tumors. He had no evidence of either, but it is uncertain whether surveillance and testing for these conditions should continue. Secondly, duty of care to other family members, by informing them of their disease risks, needs to be balanced against confidentiality. Furthermore, it will be increasingly important to defend the principle that health insurers cannot use our genetic information to calculate our premiums.

New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk

Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, Wheeler E, Glazer NL, Bouatia-Naji N, Gloyn AL, Lindgren CM, Magi R, Morris AP, Randall J, Johnson T, Elliott P, Rybin D, Thorleifsson G, Steinthorsdottir V, Henneman P, Grallert H, Dehghan A, Hottenga JJ, Franklin CS, Navarro P, Song K, Goel A, Perry JR, Egan JM, Lajunen T, Grarup N, Sparso T, Doney A, Voight BF, Stringham HM, Li M, Kanoni S, Shrader P, Cavalcanti-Proenca C, Kumari M, Qi L, Timpson NJ, Gieger C, Zabena C, Rocheleau G, Ingelsson E, An P, O'Connell J, Luan J, Elliott A, McCarrroll SA, Payne F, Roccascaccia RM, Pattou F, Sethupathy P, Ardlie K, Ariyurek Y, Balkau B, Barter P, Beilby JP, Ben-Shlomo Y, Benediktsson R, Bennett AJ, Bergmann S, Bochud M, Boerwinkle E, Bonnefond A, Bonnycastle LL, Borch-Johnsen K, Bottcher Y, Brunner E, Bumpstead SJ, Charpentier G, Chen YD, Chines P, Clarke R, Coin LJ, Cooper MN, Cornelis M, Crawford G, Crisponi L, Day IN, de Geus EJ, Delplanque J, Dina C, Erdos MR, Fedson AC, Fischer-Rosinsky A, Forouhi NG, Fox CS, Frants R, Franzosi MG, Galan P, Goodarzi MO, Graessler J, Groves CJ, Grundy S, Gwilliam R, Gyllenstein U, Hadjadj S, Hallmans G, Hammond N, Han X, Hartikainen AL, Hassanali N, Hayward C, Heath SC, Hercberg S, Herder C, Hicks AA, Hillman DR, Hingorani AD, Hofman A, Hui J, Hung J, Isomaa B, Johnson PR, Jorgensen T, Julia A, Kaakinen M, Kaprio J, Kesaniemi YA, Kivimaki M, Knight B, Koskinen S, Kovacs P, Kyvik KO, Lathrop GM, Lawlor DA, Le BO, Lecoeur C, Li Y, Lyssenko V, Mahley R, Mangino M, Manning AK, Martinez-Larrad MT, McAteer JB, McCulloch LJ, McPherson R, Meisinger C, Melzer D, Meyre D, Mitchell BD, Morken MA, Mukherjee S, Naitza S, Narisu N, Neville MJ, Oostra BA, Orru M, Pakyz R, Palmer CN, Paolisso G, Pattaro C, Pearson D, Peden JF, Pedersen NL, Perola M, Pfeiffer AF, Pichler I, Polasek O, Posthuma D, Potter SC, Pouta A, Province MA, Psaty BM, Rathmann W, Rayner NW, Rice K, Ripatti S, Rivadeneira F, Roden M, Rolandsson O, Sandbaek A, Sandhu M, Sanna S, Sayer AA, Scheet P, Scott LJ, Seedorf U, Sharp SJ, Shields B, Sigurdsson G, Sijbrands EJ, Silveira A, Simpson L, Singleton A, Smith NL, Sovio U, Swift A, Syddall H, Syvanen AC, Tanaka T, Thorand B, Tichet J, Tonjes A, Tuomi T, Uitterlinden AG, van Dijk KW, van HM, Varma D, Visvikis-Siest S, Vitart V, Vogelzangs N, Waeber G, Wagner PJ, Walley A, Walters GB, Ward KL, Watkins H, Weedon MN, Wild SH, Willemsen G, Witteman JC, Yarnell JW, Zeggini E, Zelenika D, Zethelius B, Zhai G, Zhao JH, Zillikens MC, Borecki IB, Loos RJ, Meneton P, Magnusson PK, Nathan DM, Williams GH, Hattersley AT, Silander K, Salomaa V, Smith GD, Bornstein SR, Schwarz P, Spranger J, Karpe F, Shuldiner AR, Cooper C, Dedoussis GV, Serrano-Rios M, Morris AD, Lind L, Palmer LJ, Hu FB, Franks PW, Ebrahim S, Marmot M, Kao WH, Pankow JS, Sampson MJ, Kuusisto J, Laakso M, Hansen T, Pedersen O, Pramstaller PP

Department of Biostatistics, Boston University School of Public Health, Boston, Mass., USA

Nat Genet 2010;42:105–116

Background: Levels of circulating glucose are tightly regulated. The aim of this study was to identify new genetic loci associated with various glycemic traits.

Methods: A meta-analysis of 21 genome-wide association studies was performed for fasting glucose, fasting insulin and indices of β -cell function (HOMA-B) and insulin resistance (HOMA-IR) in up to 46,186 non-diabetic participants. The top 25 loci were followed up in a further 76,558 participants.

Results: Altogether 16 loci were identified associated with fasting glucose and HOMA-B and 2 loci associated with fasting insulin and HOMA-IR. These include 9 loci newly associated with fasting glucose (in or near ADCY5, MADD, ADRA2A, CRY2, FADS1, GLIS3, SLC2A2, PROX1 and C2CD4B) and one influencing fasting insulin and HOMA-IR (near IGF-1). In further analyses, ADCY5, PROX1, GCK, GCKR and DGKB-TMEM195 were also demonstrated to be associated with risk of type 2 diabetes. The likely biological candidate genes within these loci influence signal transduction, cell proliferation, development, glucose-sensing and circadian regulation.

Conclusions: These findings demonstrate that genetic studies of continuous glycemic traits can identify novel type 2 diabetes risk loci. However, other loci are associated with a modest elevation in glucose levels but not overt diabetes.

In addition to large GWA studies, distinct diseases, such as type 2 diabetes (T2DM), studying the underlying continuous traits, such as glucose levels, may uncover further genetic loci that are relevant for disease risks. Indeed, studies for BMI have been more powerful and fruitful than studies of obesity. However, there are caveats in this approach. Firstly, some variants that affect glucose metabolism may raise glucose levels only marginally and not sufficiently to contribute to risks for T2DM, or its complications. This is akin to the rare glucokinase mutations that cause the benign condition

MODY type 2. Secondly, a recent genetic variant for HbA1c levels was surprisingly shown subsequently to influence hemoglobin levels rather than glycemia, despite being located in the hexokinase gene [1]. It is therefore reassuring to see that several of the new loci described in this study were confirmed to influence T2DM risk per se.

The size of the combined effects of the 16 loci associated with fasting glucose was substantial. There was a mean difference of -0.4 mmol/l (5.93 vs. 5.51 mmol/l) in fasting glucose when comparing the highest genetic risk group (5.6% of the sample) to the lowest (2.9% of the sample). This represents a shift of approximately 25 centile points in the distribution of fasting glucose, and corresponded to a relative risk of 1.54–1.73 for future T2DM.

The one locus for insulin resistance (HOMA-IR) was near to the strongly plausible gene IGF-1, however this variant failed to reach significance in follow-up studies of over 50,000 individuals. The lack of further loci for insulin resistance is surprising as the estimates of heritability are similar for insulin resistance and insulin secretion (HOMA-B) in family studies. The authors argue that different traits could have distinct genetic architectures. For example, some such as fasting glucose could be influenced by several common variants each with modest effects. Others, such as HOMA-IR, could be putatively influenced by fewer loci, rarer variants, or by stronger environmental modifications on their genetic effects.

Genetic variant near *IRS1* is associated with type 2 diabetes, insulin resistance and hyperinsulinemia

Rung J, Cauchi S, Albrechtsen A, Shen L, Rocheleau G, Cavalcanti-Proenca C, Bacot F, Balkau B, Belisle A, Borch-Johnsen K, Charpentier G, Dina C, Durand E, Elliott P, Hadjadj S, Jarvelin MR, Laitinen J, Lauritzen T, Marre M, Mazur A, Meyre D, Montpetit A, Pisinger C, Posner B, Poulsen P, Pouta A, Prentki M, Ribel-Madsen R, Ruokonen A, Sandbaek A, Serre D, Tichet J, Vaxillaire M, Wojtaszewski JF, Vaag A, Hansen T, Polychronakos C, Pedersen O, Froguel P, Sladek R

McGill University and Genome Quebec Innovation Centre, Montreal, Que., Canada

Nat Genet 2009;41:1110–1115

Background: Previous genome-wide association studies have identified common variants that only partially explain the genetic risk for type 2 diabetes (T2D).

Methods: 392,365 SNPs were tested for association with T2D in 1,376 French cases and controls. The 5% of variants most significantly associated with T2D (16,360 SNPs) were followed up in an independent sample of 4,977 French individuals using a custom designed array. The best 28 hits from this stage were subsequently selected for replication in 7,698 Danish subjects. Finally, the most significant novel locus was followed up in 14,358 French, Danish and Finnish participants from population-based cohorts with detailed T2D-related phenotypes.

Results: Four SNPs were identified which showed genome-wide significant association with T2D ($p < 5.0 \times 10^{-8}$), one of which (rs2943641, $p = 9.3 \times 10^{-12}$, OR = 1.19) was located adjacent to the insulin receptor substrate 1 gene (*IRS1*). The C allele of rs2943641 was associated with insulin resistance and hyperinsulinemia; this allele was also associated with reduced basal levels of *IRS1* protein and decreased insulin induction of *IRS1*-associated phosphatidylinositol-3-OH kinase activity in human skeletal muscle biopsies.

Conclusions: Genetic variation near to *IRS1* is associated with T2D, insulin resistance and hyperinsulinemia. Unlike previous T2D risk loci, which predominantly associate with impaired β -cell function, this highlights the role of insulin resistance in T2D risk.

In GWA studies of complex traits, efforts to uncover further ‘missing heritability’ have been largely focussed on increasing statistical power by combining studies in large-scale meta-analyses. Here, the authors adopted an alternative multistage strategy whereby a GWA study was first used to test genetic association with T2D in a large well-phenotyped cohort before selecting a fraction of SNPs showing the strongest association to study in well-powered subsequent stages. The advantage of this is that statistical power is maximised cost-effectively since each stage contains a smaller number of SNPs and an increased number of individuals. The authors were rewarded by the discovery of a novel T2D locus in a possible regulatory region of the insulin receptor substrate gene (*IRS1*).

In addition to being associated with T2D risk (OR = 1.19), the C allele at rs2943641 was also associated with insulin resistance and hyperinsulinemia, independently of BMI. In a subset of French individuals,

the variant was associated with a 35% increased risk of T2D and a 14% increased incidence of the development of hyperglycaemia after 9 years of follow-up. In functional studies, the authors described likely regulatory effects of rs2943641 on *IRS1* and insulin signaling. Interestingly, this SNP is in the same LD block as a previously identified coronary artery disease (CAD) risk locus (rs2943634, $r^2 = 0.8$), suggesting that genetic variation near *IRS1* may predispose to CAD through increased susceptibility to T2D, a known cardiovascular risk factor. Alternatively this gene region may have pleiotrophic effects on both phenotypes.

In contrast to previously identified T2D loci, this variant is the first to be associated with insulin resistance and hyperinsulinemia, rather than impaired insulin secretion. This will give us new insight into the contribution of insulin sensitivity in establishing T2D risk.

Variants in ADCY5 and near CCNL1 are associated with fetal growth and birth weight

Freathy RM, Mook-Kanamori DO, Sovio U, Prokopenko I, Timpson NJ, Berry DJ, Warrington NM, Widen E, Hottenga JJ, Kaakinen M, Lange LA, Bradfield JP, Kerkhof M, Marsh JA, Magi R, Chen CM, Lyon HN, Kirin M, Adair LS, Aulchenko YS, Bennett AJ, Borja JB, Bouatia-Naji N, Charoen P, Coin LJ, Cousminer DL, de Geus EJ, Deloukas P, Elliott P, Evans DM, Froguel P, Glaser B, Groves CJ, Hartikainen AL, Hassanali N, Hirschhorn JN, Hofman A, Holly JM, Hypponen E, Kanoni S, Knight BA, Laitinen J, Lindgren CM, McArdle WL, O'Reilly PF, Pennell CE, Postma DS, Pouta A, Ramasamy A, Rayner NW, Ring SM, Rivadeneira F, Shields BM, Strachan DP, Surakka I, Taanila A, Tiesler C, Uitterlinden AG, van Duijn CM, Wijga AH, Willemsen G, Zhang H, Zhao J, Wilson JF, Steegers EA, Hattersley AT, Eriksson JG, Peltonen L, Mohlke KL, Grant SF, Hakonarson H, Koppelman GH, Dedoussis GV, Heinrich J, Gillman MW, Palmer LJ, Frayling TM, Boomsma DI, Davey SG, Power C, Jaddoe VW, Jarvelin MR, McCarthy MI

Genetics of Complex Traits, Peninsula College of Medicine and Dentistry, University of Exeter, Exeter, UK

Nat Genet 2010;42:430–435

Background: Birth weight is a complex multifactorial trait. This study aimed to identify genetic variants associated with birth weight.

Methods: The results of six genome-wide association (GWA) studies ($n = 10,623$ Europeans from pregnancy/birth cohorts) were meta-analyzed. Two lead signals were followed up in 13 replication studies ($n = 27,591$).

Results: rs900400 near LEKR1 and CCNL1 ($p = 2 \times 10^{-35}$) and rs9883204 in ADCY5 ($p = 7 \times 10^{-15}$) were robustly associated with birth weight. Correlated SNPs in ADCY5 were recently implicated in regulation of glucose levels and susceptibility to type 2 diabetes, providing evidence that the well-described association between lower birth weight and subsequent type 2 diabetes has a genetic component, distinct from the proposed role of programming by maternal nutrition. Using data from both SNPs, the 9% of Europeans carrying four birth weight-lowering alleles were, on average, 113 g (95% CI 89–137 g) lighter at birth than the 24% with zero or one alleles ($P_{\text{trend}} = 7 \times 10^{-30}$).

Conclusion: The impact on birth weight is similar to that of a mother smoking 4–5 cigarettes per day in the third trimester of pregnancy.

The birth weight-lowering C allele of rs9883204 is in linkage disequilibrium (LD) ($r^2 = 0.75$) with the A allele of rs11708067, which was independently shown to be associated with increased risk of type 2 diabetes, higher fasting glucose, and reduced insulin secretion. Fetal insulin is a key fetal growth factor, and these metabolic associations suggest that one mechanism explaining the ADCY5 (cyclic AMP-generating adenylate cyclase) association with birth weight might be a direct effect of the fetal risk allele on fetal growth via reduced insulin secretion, consistent with the fetal insulin hypothesis.

The association with ponderal index, relative to those with birth length and head circumference, was particularly strong for the variant near the cyclin gene CCNL1, which encodes a kinase known to phosphorylate histone H1, suggesting that it might have a greater association with fat mass than with skeletal growth. However, for ADCY5, the birth measures showed more proportionate associations.

Birth weight represents a complex phenotype. Beyond the usual gene-environment interactions, the fetal environment may be governed by maternal genes. It is therefore important to clarify whether apparent effects of fetal genes might be proxy markers for maternal genes acting on maternal

metabolism. In the current study, meta-analysis of associations between birth weight and fetal genotype, conditional on maternal genotype, yielded very similar results to the original associations at both loci, showing that these are direct fetal effects. Secondly, fetal growth seems to be one of the main targets of imprinted genes. Future studies need to explore whether these fetal genes demonstrate parent of origin effects.

New paradigms The genetics of epigenetics

Heritable individual-specific and allele-specific chromatin signatures in humans

McDaniell R, Lee BK, Song L, Liu Z, Boyle AP, Erdos MR, Scott LJ, Morken MA, Kucera KS, Battenhouse A, Keefe D, Collins FS, Willard HF, Lieb JD, Furey TS, Crawford GE, Iyer VR, Birney E
Center for Systems and Synthetic Biology, Institute for Cellular and Molecular Biology, Section of Molecular Genetics and Microbiology, University of Texas, Austin, Tex., USA
Science 2010;328:235–239

Background: The extent to which variation in chromatin structure and transcription factor binding may influence gene expression, and thus underlie or contribute to variation in phenotype, is currently unknown.

Methods: Individual-to-individual variation and differences between homologous chromosomes within the same individual (allele-specific variation) in chromatin structure and transcription factor binding in lymphoblastoid cells derived from individuals of geographically diverse ancestry were catalogued. Differences between chromatin structure were compared between parents and children in order to identify whether chromatin signatures are heritable in humans.

Results: Ten percent of active chromatin sites were individual-specific; a similar proportion were allele-specific. Both individual-specific and allele-specific sites were commonly transmitted from parent to child, which suggests that they are heritable features of the human genome.

Conclusion: This study shows that heritable chromatin status and transcription factor binding differ as a result of genetic variation and may underlie phenotypic variation in humans.

Variation in chromatin structure can facilitate or restrict the binding of transcription factors to regulatory regions and is therefore a major component in the regulation of gene transcription. While there is much interest in such epigenetic changes as the possible mechanism to explain ‘long-term programming’ responses to specific early exposures, alterations of chromatin structure are in turn also influenced by genetic variation.

In order to increase our understanding of this allele-specific gene regulation, the authors catalogued variation in transcription factor binding and chromatin structure in a European and an African family (both parents and one daughter from each) from the ‘1000 Genomes Project’ using deoxyribonuclease I hypersensitive (DNase I) site mapping and chromatin immunoprecipitation (ChIP). This allowed them to identify regulatory DNA elements such as promoters, enhancers, silencers and insulators. The study design enabled the investigators to compare related and unrelated individuals and determine whether chromatin structure is heritable. Comparison of the children’s chromatin signals to that of the parents revealed that chromatin sites were much more similar to their own parents’ than those of the unrelated family, demonstrating its strong heritability. Remarkably, only 10% of active chromatin sites were individual-specific and this indicates that one of the major types of epigenetic variation (i.e. chromatin structure) is itself largely influenced by genetic variation. Future studies should explore the specific genetic regulation of epigenetic marks.

Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis

Baranzini SE, Mudge J, van Velkinburgh JC, Khankhanian P, Khrebtukova I, Miller NA, Zhang L, Farmer AD, Bell CJ, Kim RW, May GD, Woodward JE, Caillier SJ, McElroy JP, Gomez R, Pando MJ, Clendenen LE, Ganusova EE, Schilkey FD, Ramaraj T, Khan OA, Huntley JJ, Luo S, Kwok PY, Wu TD, Schroth GP, Oksenberg JR, Hauser SL, Kingsmore SF

Department of Neurology, University of California at San Francisco, San Francisco, Calif., USA

sebaran@cgl.ucsf.edu

Nature 2010;464:1351–1356

Background: Monozygotic twins are often studied to quantify the relative contributions of genetics and environment in human diseases. In multiple sclerosis (MS), disease discordance between monozygotic twin pairs has been interpreted to indicate its important environmental pathogenesis. However, both genetic and epigenetic differences between monozygotic twins have been described, challenging the accepted experimental model in disambiguating the effects of nature and nurture.

Methods: Genomes of one MS-discordant monozygotic twin pair were sequenced and compared. Messenger RNA transcriptome and epigenome sequences of CD4+ lymphocytes from three MS-discordant monozygotic twin pairs were also determined and evaluated.

Results: No reproducible differences were detected between co-twins among approximately 3.6 million single nucleotide polymorphisms (SNPs) or approximately 0.2 million insertion-deletion polymorphisms. Nor were any reproducible differences observed between siblings of the three twin pairs in HLA haplotypes, confirmed MS susceptibility SNPs, copy number variations, mRNA and genomic SNP and insertion-deletion genotypes, or the expression of approximately 19,000 genes in CD4+ T cells. Only 2–176 methylation differences out of approximately 2 million CpG dinucleotides were detected between siblings of the three twin pairs, in contrast to approximately 800 methylation differences between T cells of unrelated individuals and several thousand differences between tissues or between normal and cancerous tissues.

Conclusion: In the first systematic effort to estimate sequence variation among monozygotic co-twins, there was no evidence for genetic, epigenetic or transcriptome differences that explained MS discordance. These are the first female, twin and autoimmune disease individual genome sequences reported.

Genetically identical twins discordant for a particular disease are considered a useful tool for studying a possible contribution of epigenetics to disease pathogenesis. In this study, the authors compared genetic and epigenetic differences in monozygotic twins discordant for multiple sclerosis (MS), using CD4+ lymphocytes due to their known involvement in the autoimmune pathophysiology of MS. This study of unprecedented resolution is the first to systematically characterise both genetic and epigenetic contributions to disease discordance in monozygotic twins using an impressive range of sophisticated sequencing technologies.

Surprisingly, no reproducible genetic, epigenetic or transcriptome differences were identified to explain the twins discordance in MS. So what lessons can be learnt? Firstly, the lack of genetic differences between monozygotic twins at least reassures us of the validity of the monozygotic-dizygotic twins model to study the heritability of disease risk. Secondly, the further lack of epigenetic or transcriptome differences between monozygotic twins is consistent with the recent findings of McDaniell et al. (described above) in illustrating the predominant ‘genetic regulation of epigenetic marks’.

And what could explain the disease discordance? The authors describe that their novel comprehensive approach still had limited coverage by exclusion of low coverage regions and repetitive sequences, by only moderate sensitivity for detection of structural variants of size 50–1,500 nucleotides, and (the most tractable explanation in our opinion) by limited feasibility to detect possible somatic mosaicism in tissues other than the CD4+ lymphocyte.

The imprinted DLK1-MEG3 gene region on chromosome 14q32.2 alters susceptibility to type 1 diabetes

Wallace C, Smyth DJ, Maisuria-Armer M, Walker NM, Todd JA, Clayton DG

Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Department of Medical Genetics, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK

Nat Genet 2010;42:68–71

Background: Genome-wide association (GWA) studies successfully mapped common disease susceptibility loci. To date, over 300 reproducibly associated loci have been reported. However, there is yet no convincing evidence for any susceptibility locus that displays parent-of-origin effects.

Methods: Data were combined from three existing GWA studies comprising a total of 7,514 cases and 9,405 controls of European ancestry. Potential novel T1DM loci were tested in additional case-control and family studies.

Results: There was robust evidence at rs941576 for paternally inherited risk of T1DM; ratio of allelic effects for paternal versus maternal transmissions = 0.75; 95% CI = 0.71–0.79). This marker is in the imprinted region of chromosome 14q32.2, which contains the functional candidate gene DLK1. Our meta-analysis also provided evidence for a further T1DM locus at chromosome 19p13.2. The highest association was at marker rs2304256 (odds ratio (OR) = 0.86; 95% CI = 0.82–0.90) in the TYK2 gene, which has previously been associated with systemic lupus erythematosus and multiple sclerosis.

Conclusion: This study reports two novel T1DM loci, one of which is located in an imprinted gene region and conferred protection against T1DM only when paternally transmitted.

Previous analysis of these data had identified over 40 loci for T1DM [2]. This study took forward three further loci which had not quite reached the p-value thresholds for genome-wide significance ($p < 10^{-8}$), but had strong 'suggestive evidence' for association ($p < 10^{-7}$) with T1DM in additional case-control and family samples. In an overall test of GWA stage and replication data, two of these loci then reached overall genome-wide significance ($p = 4 \times 10^{-9}$ and $p = 1.6 \times 10^{-10}$).

The authors noticed that one of these signals was located in a region characterised by several imprinted genes. They therefore tested the hypothesis that this association could show parent of origin effects, which they indeed showed by analysis of informative parent-offspring trios from family-based studies.

As is typical of imprinted regions, the imprinted region of chromosome 14q32.2 contains several imprinted genes, some of which are only maternally expressed (BEGAIN, MEG3, MEG8 and DIO3OS), and others only paternally expressed (DLK1, RTL1 and DIO3). Notably, they found that the overall protective effect was paternally (and not maternally) inherited, and this then narrowed the list of potentially causative genes. The best candidate, DLK1, is paternally expressed, with high levels seen in human heart, pancreatic islet cells, pituitary tissue, ovaries, placenta and testes. DLK1 encodes a membrane-bound protein that is cleaved to form fetal antigen-1, which is involved in the differentiation of many cell types, including pancreatic β cells and B lymphocytes, and promotes B-cell proliferation in human peripheral blood. Thus, there are a number of ways in which variation in DLK1 expression could alter susceptibility to T1DM.

Parental origin of sequence variants associated with complex diseases

Kong A, Steinthorsdottir V, Masson G, Thorleifsson G, Sulem P, Besenbacher S, Jonasdottir A, Sigurdsson A, Kristinsson KT, Jonasdottir A, Frigge ML, Gylfason A, Olason PI, Gudjonsson SA, Sverrisson S, Stacey SN, Sigurgeirsson B, Benediktsdottir KR, Sigurdsson H, Jonsson T, Benediktsson R, Olafsson JH, Johannsson OT, Hreidarsson AB, Sigurdsson G, Ferguson-Smith AC, Gudbjartsson DF, Thorsteinsdottir U, Stefansson K
deCODE genetics, Reykjavik, Iceland
kong@decode.is
Nature 2009;462:868–874

Background: Effects of susceptibility variants may depend on from which parent they are inherited. However, in genome-wide association studies to date, the impact of parental origin has largely been ignored.

Methods: From a public database of all disease-associated SNPs identified by GWA studies, those independent SNPs ($n = 7$) located within 500 kilobases of a known imprinted gene were selected. Their disease associations were re-examined in 38,167 Icelanders with existing GWA data. Parental origin of the alleles was determined using a combination of genealogy and long-range phasing.

Results: Of the 7 SNPs analysed, 5 showed parent-of-origin specific associations (1 with breast cancer, 1 with basal-cell carcinoma, and 3 with type 2 diabetes). These variants are located in two genomic regions, 11p15 and 7q32, each harbouring a cluster of imprinted genes. SNP rs2334499 at 11p15 was identified as a novel locus for type 2 diabetes; here the allele that conferred increased risk when paternally inherited was protective when maternally transmitted. rs2334499 demonstrated correlation with methylation levels at a novel differentially methylated CTCF-binding site at 11p15.

Conclusions: Genome-wide association studies have so far yielded variants that explain only a small fraction of the heritability of the traits studied. Some of the remainder may be hidden in more complex relations between variants and disease risk.

This paper, published only 1 month prior to Wallace et al. (above), took a unique systematic approach to identify SNP-disease associations that differ according to whether the allele is inherited from your mother or father. The strategy was remarkably successful in that 5 of the 7 SNPs analysed did indeed show evidence of parental origin effects.

Novel mathematical approaches were developed to take advantage of this uniquely characterised population, whereby in the absence of direct genotype information on parents, the national genealogy database was used to identify even distant family relationships with other study participants (their examples even included third cousins, twice removed). This inference of parent of origin was tested in a subset of direct parent-offspring trios and was found to be 99.8% accurate.

They identified a novel T2DM variant at KCNQ1, which is one of the known loci for methylation defects in Beckwith-Wiedemann syndrome. Intriguingly, this variant showed little traditional, or overall, genotype association with disease risk because its influence on increased risk if T2DM when paternally inherited was balanced by a protective effect when maternally transmitted. In the debate over the yet 'missing heritability' of common diseases, this study claims to open new avenues for the discovery of further disease loci. However, it is a path that will require a new generation of large and well-characterised family-based studies.

Finding the missing heritability of complex diseases

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM
National Human Genome Research Institute, Bethesda, Md., USA

manoliot@mail.nih.gov

Nature 2009;461:747–753

Background: Genome-wide association studies have identified hundreds of genetic variants associated with complex human diseases and traits, and have provided valuable insights into their genetic architecture. Most variants confer relatively small increments in risk, and explain only a small proportion of familial clustering, leaving the question how the remaining, 'missing' heritability will be explained.

Methods: These authors examined potential sources of missing heritability and propose research strategies, including and extending beyond current genome-wide association approaches, to illuminate the genetics of complex diseases and enhance its potential to enable effective disease prevention or treatment.

Results: Strategies were identified to explore the contribution of rare, low frequency and structural variants, to investigate the potential contributions of gene-gene and gene-environment interactions and to make the most of existing and future genome-wide association studies.

Conclusion: The search for missing heritability provides a potentially valuable path towards future discoveries.

The debate over the nature of the genetic contribution to individual susceptibility to multigenic diseases goes on. Genetic variants with appreciable frequency in the general population but relatively weak effects, and rare variants with relatively large effects are the likely major contributors to genetic susceptibility. The hype over the high profile results of genome-wide association studies has been countered by criticisms about the weak effect sizes and poor individual predictive ability of common variants, and the low proportion of heritability that they explain. For example, the current 40 loci for adult height together explain only around 5% of its heritability, and the corresponding figure for the 18 loci for type 2 diabetes is only 6%. The authors argue that 'few investigators expected GWA studies immediately to find all of the variants associated with common diseases, or even most of them; the hope was that they would find at least some'. In that regard, GWA studies have been enormously successful, and the remaining or 'missing' heritability should be seen as the opportunity for future genetic discoveries!

The authors present several suggestions. Most imminently, 'mega' or greatly expanded GWA studies for several traits are in progress and the results of these more powerful GWA analyses will likely be published in the coming months. Much attention is also currently focused on the role of low frequency variants with potentially relatively large effect sizes. Currently, GWA chips, and the HapMap database on which they are based, poorly represent variants with minor allele frequencies <5%. Distinct from rare deleterious mutations that cause mendelian disorders, low-frequency variants have alleles in the 0.5–5% frequency range. Systematic discovery of these variants is rapidly progressing through the '1000 Genomes Project' and this information is already being applied in new genotyping study designs for obesity, type 2 diabetes and cardiovascular disease.

Fine mapping of the association with obesity at the FTO locus in African-derived populations

Hassanein MT, Lyon HN, Nguyen TT, Akyzbekova EL, Waters K, Lettre G, Tayo B, Forrester T, Sarpong DF, Stram DO, Butler JL, Wilks R, Liu J, Le ML, Kolonel LN, Zhu X, Henderson B, Cooper R, McKenzie C, Taylor HA, Jr, Haiman CA, Hirschhorn JN

Department of Preventive Medicine, Keck School of Medicine and Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, Calif., USA

Hum Mol Genet 2010;19:2907–2916

Background: Genome-wide association studies have typically been performed in individuals of recent European ancestry. In these populations, many common variants are tightly correlated, which considerably limits the resolution of fine mapping. Populations with recent African ancestry often have less extensive and/or different patterns of linkage disequilibrium (LD), and have been proposed to be useful in fine-mapping studies.

Methods: In populations of predominantly African ancestry, variation at the FTO locus and body mass index (BMI) was explored. Single nucleotide polymorphisms that are correlated with the signal of association in individuals of European ancestry but that have varying degrees of correlation in African-derived individuals were genotyped.

Results: Most of the variants, including one previously proposed as being functionally important, showed no significant association with BMI. However, two FTO variants, rs3751812 and rs9941349, showed strong evidence of association ($p = 2.58 \times 10^{-6}$ and 3.61×10^{-6}) in a meta-analysis of 9,881 African-ancestry individuals.

Conclusions: This study both strongly replicated the FTO-BMI association in African-ancestry populations and narrowed the list of potentially causal variants to those that are correlated with rs3751812 and rs9941349 in African-derived populations. This study illustrates the potential of using populations with different LD patterns to fine map genetic associations and supports the rationale for genetically guided functional studies at the FTO locus.

The vast majority of GWA studies have been performed in studies of White-European populations. These authors argue for the study of African-ancestry populations, firstly to test whether the loci identified in European studies are also relevant in individuals of African ancestry. Second, studies of African-ancestry populations may be more efficient for 'fine-mapping' studies, i.e. the more precise location of the underlying causal gene or causal variant. This is because African-ancestry populations show less linkage disequilibrium, or greater genetic heterogeneity, than European populations. They found that the European obesity-associated intronic variant rs9939609 in FTO showed no association at all with obesity risk in African-derived populations, ruling this out as the causal variant. However, two further FTO variants, rs3751812 and rs9941349, did show strong associations in African-derived populations. While in Europeans the 'FTO-association signal' arguably extended to covered neighbouring genes, such as *RPGRIP1L*, the current study clearly localises FTO as the underlying gene in this locus in African populations.

In addition to simple confirmation of 'European' GWA hits in African populations, the setting of 'discovery' phase studies in non-Europeans may identify new genes or loci for common disease. For example, a recent GWA study in 19,633 Japanese individuals identified *IGF-1* as a major locus for adult height [3]. The 40-plus adult height loci identified in previous GWA studies contained surprisingly few genes implicated in traditional hormone pathways [4]. The minor allele frequency for the height lowering at IGF-1 was 26% in Japanese subjects compared to only 1.7% in Caucasians, which explains why those European studies had insufficient power to detect its effects.

Comprehensive mapping of long-range interactions reveals folding principles of the human genome

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J

Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, Mass., USA

Science 2009;326:289–293

Background: Understanding how chromosomes fold can provide insight into the complex relationships between chromatin structure, gene activity, and the functional state of the cell. Previously, methods have explored interactions between target loci.

Methods: Hi-C is a novel method that probes the three-dimensional architecture of whole genomes by coupling proximity-based ligation with massively parallel sequencing. Spatial proximity maps of the human genome were constructed at a resolution of 1 megabase.

Results: Spatial proximity maps confirm the presence of chromosome territories and the spatial proximity of small, gene-rich chromosomes. An additional level of genome organization was identified, characterized by the spatial segregation of open and closed chromatin to form two genome-wide compartments. At the megabase scale, the chromatin conformation is consistent with a fractal globule, a knot-free, polymer conformation that enables maximally dense packing while preserving the ability to easily fold and unfold any genomic locus.

Conclusions: This study demonstrates the power of the novel Hi-C method to map the dynamic conformations of whole genomes.

In its stretched out form, the human genome is roughly 2 m long. Important questions are ‘How does the linear order of basepairs relate to their spatial arrangement?’ and ‘How are genes regulated by distant elements?’ A surprising, but commonly re-occurring feature of GWA studies is that variants that are highly robustly associated with disease risks can be located very far away from genes or gene regions. For example, the obesity-susceptibility SNP rs17782313 is located 188 kb downstream of its nearest gene *MC4R* [5]. Furthermore, the 9p21 locus is incontrovertibly linked to risk of cardiovascular disease, yet is in an ‘intergenic’ region devoid of protein-coding genes [6]. Such loci are beyond the traditional range for binding sites of activators or regulators of gene expression, although it is possible that such regions might encode functional non-coding RNAs or copy-number variants. This study demonstrates a further possible mechanism of long-range DNA interaction due to genome folding, which could bring into close approximation distant genes and regulatory sites.

In general, the ‘average intra-chromosomal contact probability’ for pairs of loci decreased monotonically with increasing genomic distance between the loci. However, even beyond distances of 200 Mb, intra-chromosomal contact probabilities were higher than the contact probabilities between different chromosomes, and blocks could be identified with either enriched or deleted interactions. Furthermore, inter-chromosomal interactions were detected, particularly between small gene-rich chromosomes (16, 17, 18, 19, 20, 21, and 22), which paves the way for exploration of gene-gene interactions. Although this global arrangement allows insights into chromatin folding principles, at this point it is not yet clear whether spatial proximity of two loci necessarily means that there is a functional relationship.

Reciprocal imprinting of human *GRB10* in placental trophoblast and brain: evolutionary conservation of reversed allelic expression

Monk D, Arnaud P, Frost J, Hills FA, Stanier P, Feil R, Moore GE

Clinical and Molecular Genetics Unit, Institute of Child Health, University College London, London, UK

dmonk@iconcologia.net

Hum Mol Genet 2009;18:3066–3074

Background: Genomic imprinting may have evolved not only to regulate fetal growth and development, but also behaviour. The murine growth factor receptor-binding protein 10, *Grb10* is a potent growth inhibitor. Its gene displays reciprocal imprinting in mice.

Methods: To assess the biological relevance of this reciprocal pattern of imprinting, we explored its conservation in humans.

Results: As in mice, human *GRB10* gene was paternally expressed in brain. Maternal allele-specific expression was seen only in placental villous trophoblasts, an essential part of the placenta involved in nutrient transfer. All other fetal tissues tested showed equal expression from both alleles. As in the mouse, the maternal transcripts originate from several kilobases upstream of the imprinting control region (ICR) of the domain, from a promoter region at which we find no allelic chromatin differences. The brain-specific paternal expression from the ICR also shows mechanistic similarities with the mouse. This conserved CpG island is DNA-methylated on the maternal allele and is marked on the paternal allele by developmentally regulated bivalent chromatin, with the presence of both H3 lysine-4 and H3 lysine-27 methylation.

Conclusions: The strong conservation of the opposite allelic expression in placenta versus brain supports the hypothesis that *GRB10* imprinting evolved to mediate diverse roles in mammalian growth and behaviour.

In addition to methylation defects in the imprinted *H19-IGF2* region, *GRB10* has long been a candidate to explain further cases of the Silver-Russell syndrome (SRS) of low birth weight and postnatal growth failure. Indeed, as the authors point out maternal uniparental disomy (mUPD7) and small duplications of 7p12 encompassing *GRB10* have been reported in ~10% of cases of SRS. However, contrary to the conflict model (that fathers' genes favour large offspring size, and mothers' genes favour small size), *GRB10* shows biallelic expression in most tissues except brain, where this potent growth inhibitor is paternally, rather than maternally, expressed. This study now finds that *GRB10* is indeed exclusively maternally expressed, and specifically in the placental cells responsible for nutrient transfer. Remarkably, as in the mouse model, in human brain *GRB10* shows the opposite pattern of exclusive paternal expression.

It has been suggested that imprinting evolved to allow the expression of parent-specific considerations regarding the balance of fetal survival to cost or risk to the parent. Hence promoters of fetal growth tend to be paternally expressed, while growth suppressors are maternally expressed [7]. The authors suggest that this more complex model of imprinting of *GRB10* might have evolved to allow multiple actions of *GRB10* to be manifested. Hence, overexpression of maternally derived *GRB10* in trophoblast would limit fetal growth, resulting in IUGR, whereas a lack of paternally derived growth-enhancing *GRB10* in brain could explain the weaning phenotype associated with poor suckling. They further suggest that chromatin defects at the *GRB10* DMR leading to lack of paternal *GRB10* expression in brain could potentially result in other cases of non-syndromic postnatal growth restriction.

Genetic evidence for high-altitude adaptation in Tibet

Simonson TS, Yang Y, Huff CD, Yun H, Qin G, Witherspoon DJ, Bai Z, Lorenzo FR, Xing J, Jorde LB, Prchal JT, Ge R
Eccles Institute of Human Genetics, University of Utah School of Medicine, Salt Lake City, Utah, USA
Science 2010; *Epub ahead of print*

Background: Tibetans have lived at very high altitudes for thousands of years, and they have a distinctive suite of physiological traits that enable them to tolerate environmental hypoxia. These phenotypes are clearly the result of adaptation to this environment, but their genetic basis remains unknown.

Methods: 31 unrelated Tibetans were genotyped for 1 million single nucleotide polymorphisms (SNPs) using the Affymetrix Genome-Wide Human SNP 6.0 Array. To pinpoint loci under positive selection, the cross-population extended haplotype homozygosity (XP-EHH) statistic was used to compare highland Tibetans to the combined HapMap Chinese (CHB) and Japanese (JPT) lowland populations.

Results: There was evidence for positive selection in several regions that contain genes whose products are likely involved in high-altitude adaptation. Positively selected haplotypes of *EGLN1* and *PPARA* were significantly associated with the decreased hemoglobin phenotype that is unique to this highland population.

Conclusions: Identification of these genes provides support for previously hypothesized mechanisms of high-altitude adaptation and illuminates the complexity of hypoxia response pathways in humans.

There are yet few specific examples of genetic adaptation. Genetic variants that confer lactase persistence are thought to have evolved with the expansion of agricultural societies, while G6PD deficiency, α -thalassaemia, and hemoglobin C persist by protecting against malaria mortality.

Since the beginning of the Himalayan climbing era, the anecdotal extraordinary physical performance at high altitude of Sherpas has intrigued climbers. These authors describe that the Tibetan highlands are one of the most extreme environments inhabited by humans. In contrast to other non-adapted populations who increase haemoglobin levels and maintain oxygen saturation in response to hypoxia (with often adverse consequences), high-altitude Tibetans show profound arterial hypoxia yet they maintain normal aerobic metabolism. This is purportedly due to more efficient oxygen transport systems, however the specific mechanisms and the genetic basis have been elusive. By comparison of genome-wide SNP frequencies in Tibetans to those in neighbouring lowland populations, this study identified evidence for evolutionary selection of several genes with a priori functional roles in hypoxia response. Remarkably, each additional copy of an advantageous haplotype at either *EGLN1* (a suppressor of HIF target genes and recently implicated in growth plate processes) or *PPARA* (known to play a role in lipid metabolism) was associated with a massive 1.7 g/dl reduction in haemoglobin concentration, which is consistent with the previously suggested mechanism of adaptation. Further studies of these genes may lead to new approaches for the treatment of hypoxia-related diseases.

New paradigms **The other genomes we carry**

The cancer genome

Stratton MR, Campbell PJ, Futreal PA
Cancer Genome Project, Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK
mrs@sanger.ac.uk
Nature 2009;458:719–724

Background: All cancers arise as a result of changes that have occurred in the DNA sequence of the genomes of cancer cells. Over the past quarter of a century much has been learnt about these mutations

and the abnormal genes that operate in human cancers. We are now, however, moving into an era in which it will be possible to obtain the complete DNA sequence of large numbers of cancer genomes.

Methods: This paper reviews the history and recent rapid progress in this field and gives comments on its future directions.

Results: Approximately 100,000 somatic mutations from cancer genomes have been reported in the quarter of a century since the first somatic mutation was found in HRAS. At least 350 (1.6%) of the 22,000 protein-coding genes in the human genome have been reported to show recurrent somatic mutations in cancer with strong evidence that these contribute to cancer development. Over the next few years several hundred million more will be revealed by large-scale, complete sequencing of cancer genomes.

Conclusions: These studies will provide us with a detailed and comprehensive perspective on how individual cancers have developed and will provide new directions for the treatment of cancer.

In addition to our 'own' germ-line genome (i.e. that inherited from our parents), there is increasing evidence of the potential disease impacts of the 'other genomes' that we carry, ranging from the microbiomes in body cavities to somatic genomes in our cells. Intricate DNA repair mechanisms usually reduce the impact of somatic mutations, and germ-line mutations in these housekeeping genes explain the heritability of cancer susceptibility.

One of the most famous somatic genetic disruptions (i.e. acquired post-meiosis) in cancer etiology is the 'Philadelphia translocation' between chromosomes 9 and 22 in chronic myeloid leukaemia. These authors now describe that with the recent advent of whole-genome sequencing of DNA from cancer cells it is now recognised that cancer genomes may carry between 1,000 to 100,000 mutations. These somatic mutations range from single basepair substitutions; insertions or deletions of small or large segments of DNA; DNA rearrangements, and copy number increases and reductions. Furthermore, cancer genomes may have incorporated exogenous viral DNA, and will also have acquired epigenetic changes.

Somatic mutations may be classified into two groups: 'driver' mutations that confer a growth advantage on the cancer cell, and 'passenger' mutations without any growth advantage. Recent studies indicate that there may be as many as 20 driver mutations in individual cancers, considerably more than the 5–7 previously predicted. The authors describe that these studies will generate new insights into specific disease phenotypes, detection and monitoring, prognosis, drug response and chemotherapy resistance.

References

1. Bonnefond A, Vaxillaire M, Labruno Y, Lecoecur C, Chevre JC, Bouatia-Naji N, Cauchi S, Balkau B, Marre M, Tichet J, Riveline JP, Hadjadj S, Gallois Y, Czernichow S, Hercberg S, Kaakinen M, Wiesner S, Charpentier G, Levy-Marchal C, Elliott P, Jarvelin MR, Horber F, Dina C, Pedersen O, Sladek R, Meyre D, Froguel P: Genetic variant in HK1 is associated with a proanemic state and A1C but not other glycemic control-related traits. *Diabetes* 2009;58:2687–2697.
2. Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, Erlich HA, Julier C, Morahan G, Nerup J, Nierras C, Plagnol V, Pociot F, Schuilenburg H, Smyth DJ, Stevens H, Todd JA, Walker NM, Rich SS: Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Gen* 2009;41:703–707.
3. Okada Y, Kamatani Y, Takahashi A, Matsuda K, Hosono N, Ohmiya H, Daigo Y, Yamamoto K, Kubo M, Nakamura Y, Kamatani N: A genome-wide association study in 19 633 Japanese subjects identified LHX3-QSOX2 and IGF-1 as adult height loci. *Hum Mol Genet* 2010;19:2303–2312.
4. Weedon MN, Lango H, Lindgren CM, Wallace C, Evans DM, Mangino M, Freathy RM, Perry JR, Stevens S, Hall AS, Samani NJ, Shields B, Prokopenko I, Farrall M, Dominiczak A, Johnson T, Bergmann S, Beckmann JS, Vollenweider P, Waterworth DM, Mooser V, Palmer CN, Morris AD, Ouwehand WH, Zhao JH, Li S, Loos RJ, Barroso I, Deloukas P, Sandhu MS, Wheeler E, Soranzo N, Inouye M, Wareham NJ, Caulfield M, Munroe PB, Hattersley AT, McCarthy MI, Frayling TM: Genome-wide association analysis identifies 20 loci that influence adult height. *Nat Genet* 2008;40:575–583.
5. Loos RJ, Lindgren CM, Li S, Wheeler E, Zhao JH, Prokopenko I, Inouye M, Freathy RM, Attwood AP, Beckmann JS, Berndt SI, Jacobs KB, Chanock SJ, Hayes RB, Bergmann S, Bennett AJ, Bingham SA, Bochud M, Brown M, Cauchi S, Connell JM, Cooper C, Smith GD, Day I, Dina C, De S, Dermitzakis ET, Doney AS, Elliott KS, Elliott P, Evans DM, Sadaf Farooqi I, Froguel P, Ghori J, Groves CJ, Gwilliam R, Hadley D, Hall AS, Hattersley AT, Hebebrand J, Heid IM, Lamina C, Gieger C, Illig T, Meitinger T, Wichmann HE, Herrera B, Hinney A, Hunt SE, Jarvelin MR, Johnson T, Jolley JD, Karpe F, Keniry A, Khaw KT, Luben RN, Mangino M, Marchini J, McArdle WL, McGinnis R, Meyre D, Munroe PB, Morris AD, Ness AR, Neville MJ, Nica AC, Ong KK, O'Rahilly S, Owen KR, Palmer CN, Papadakis K, Potter S, Pouta A, Qi L, Kraft P, Hankinson SE, Hunter DJ, Hu FB, Randall JC, Rayner NW, Ring SM, Sandhu MS, Scherag A, Sims MA, Song K, Soranzo N, Speliotes EK, Lyon HN, Voight BF, Ridderstrale M, Groop L, Syddall HE, Teichmann SA, Timpson NJ, Tobias JH, Uda M, Scheet P, Sanna S, Abecasis GR, Albai G, Nagaraja R, Schlessinger D, Ganz Vogel CI, Wallace C, Waterworth DM, Weedon MN, Willer CJ, Jackson AU, Tuomilehto J, Collins FS, Boehnke M, Mohlke KL, Wright VL, Yuan X, Zeggini E, Hirschhorn JN, Strachan DP, Ouwehand WH, Caulfield MJ, Samani NJ, Frayling TM,

- Vollenweider P, Waeber G, Mooser V, Deloukas P, McCarthy MI, Wareham NJ, Barroso I: Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nat Genet* 2008;40:768–775
6. McPherson R: Chromosome 9p21 and coronary artery disease. *N Engl J Med* 2010;362:1736–1737
 7. Haig D: Altercation of generations: genetic conflicts of pregnancy. *Am J Reprod Immunol* 1996;35:226–232