5 DISCUSSION

5.1 Summary of Findings

Understanding the genetic causes of DD is a priority of contemporary medical research. Modern rare disease studies rely heavily on exome sequencing, yet prior to the research described in this dissertation, software tools to detect uniparental disomy or structural mosaicism from sequencing data were lacking. This limitation led to the development of UP*D*io and MrMosaic, software tools which have extended the diagnostic reach of sequencing data and have been made freely available. Simulation studies have shown that these tools can detect the large-scale abnormalities identified by karyotyping or microarray in standard clinical testing. Implementation on nearly 5,000 children with undiagnosed diseases has shown that UPD and structural mosaicism are enriched in children with developmental disorders compared with healthy children. The estimated odds ratios compared to apparently healthy population controls suggested that most of the detected abnormalities are likely to be pathogenic. Assessment of the clinical impact of the detected events identified several disease-causing mechanisms, including UPD-associated imprinting and recessive diseases, and genomic disorders associated with large mosaic deletions and duplications. Some pathogenic mechanisms were unexpected and opened new research opportunities, such as UPD associated with triplication and mosaic reversion. The results of the analyses presented here have enabled genetic diagnoses for about 25 children, ending for them and their families, their quest for diagnosis.

5.2 Implications

The new methods described in this dissertation detected abnormalities and enabled diagnoses in approximately 1% and 0.5%, respectively, of the probands enrolled in DDD. The implication of this finding is that UPD and mosaic structural variation are small but important parts of genetic diagnosis in rare disease studies.

Heterodisomy is difficult to detect without genome-wide trio data and no large trio dataset had existed prior to the DDD study. Therefore, some of the outstanding questions in the field related to heterodisomy, such as the prevalence and diagnostic rate of heterodisomy in children with DD, could now be answered. For instance, of the 21 UPD events detected among 4,320 samples, 8 (38%) were entirely heterodisomic and likely to have escaped detection by non-trio-based screening. The implication of this finding is that trio-based methods increase UPD detection by about 50%. About half of the all-heterodisomy UPD chromosomes appear to be diagnostic, suggesting that trio-based analysis increases UPD diagnostic yield by 25%. The 0.49% UPD detection rate (21 of 4,320 samples) is, given assessment of both isodisomy and heterodisomy in this large trio study, and not withstanding the ascertainment bias of children selected for DDD recruitment, the best estimate of UPD frequency in children with DD to date.

Investigation of structural mosaicism identified a disparity in the tissue-distribution of mosaicism since in 8 of 11 cases, mosaicism was not observed in blood but was observed in saliva (likely from buccal epithelium). This tissue-difference may reflect greater negative selection against pathogenic mosaicism in lymphocytes, as suggested in Pallister-Killian syndrome[240]. An alternative possibility is differential rate of generation, but this is less likely, as studies of cadavers have shown that non-pathologic somatic CNVs are commonly found in many tissue types[241,242]. The tissue disparity observed in this study lends support for the assessment of saliva in disease studies, as, other factors equal, this tissue yielded greater numbers of mosaic diagnoses. There are several additional arguments supporting the collection of DNA from saliva rather than blood for high-throughput analysis, including that it is less invasive, less expensive[243], easier to store and ship[244], and genotyped equally well as blood[243]. Arguments against the use of saliva may include the absence of biomarkers present in blood that may also be of interest[245], lower DNA yield compared to blood[243], increased contamination of foreign (i.e. bacterial) DNA[246], or that higher rates of mosaicism in saliva may make it theoretically more challenging to assess genotype. However, for the purpose of high-throughput genetic analysis in studies of rare disease, DNA extraction

is the primary concern over biomarkers and mosaicism, and can increase diagnostic yield. Therefore, saliva sampling may become more popular for future research studies, and diagnostic testing.

An implication of high-throughput assays, such as WES and WGS, in connection with variant detection software, such as the algorithmic techniques developed in this work, is that the discovery of genomic variation has outpaced its interpretation. In the near term, the interpretation gap is likely to widen as WGS provides the resolution to detect smaller structural variants, whose significance will be unknown, and may add diagnostic anxiety[247]. This pressure highlights the importance of collaborative efforts, such as DECIPHER, and continued aggregation of genomic variation across centres to facilitate pathological assessment of structural mosaicism and UPD.

The most common trisomy in pregnancy is trisomy 16[248], and the most common UPD-generating mechanism is trisomy rescue[124]; but UPD 16 is observed less often than UPD of chromosomes 15, 11, 7, and 14 (descending order of observed frequency)[124]. Ascertainment bias almost certainly plays a role in this discrepancy, as these higher-frequency UPD chromosomes are involved in imprinting disorders, and are observed following scrutiny from characteristic phenotypes in children. While UPD 16 is controversially implicated in imprinting disorders, it is known that constitutive 16 trisomy is lethal, and that trisomy rescue is often incomplete, resulting in mosaic trisomy; perhaps lower levels of UPD16 reflects the fact that trisomy rescue is often incomplete and children with mosaic 16 rarely survive.

5.3 Limitations

5.3.1 Estimates of prevalence

Only about one third of the full DDD sample set was available for the work presented in this dissertation. Therefore, the assessment of UPD and mosaicism frequency is less precise than will be possible when the study is complete. Nevertheless, UPD frequency in the first-stage 1,000 trios was not significantly different from either the second-stage 3,000 trios or from estimates of UPD frequency in other DD studies; these pieces of evidence suggest limited benefit of acquiring additional samples for the purpose of improving the genome-wide estimate of UPD frequency in DD children. There was a non-significant lower frequency of mosaicism from nearly 4,000 additional children

beyond the first analysed 1,000 trios so it is conceivable that collecting greater number of samples will lower our frequency estimate of structural mosaicism. The trio set available in DDD enabled frequency estimates of heterodisomy, but the lack of trio data in the WTCCC dataset hindered heterodisomy frequency estimates in that dataset and relied on extrapolation from the identification of UPD with mixed heterodisomic and isodisomic regions.

The DDD population is not representative of all children with DD but reflects a pre-screened population as recruitment is generally only offered to children for which prior investigation of genetic abormalities failed to yield diagnostic abnormalities. Since many UPD and mosaic structural variants lead to phenotypically evident, syndromic manifestations, some children with such abnormalities and DD may be excluded from recruitment. Therefore, DDD likely has an ascertainment bias that lowers the estimate of UPD and mosaicism compared to the full population of children with DD. Children in DDD are unlikely to have large high-clonality mosaic events, unless perhaps, if such mosaicism is limited to tissue not analysed. Thus, it likely that the frequency estimates made in this work of UPD and structural mosaicism are underestimates compared to children in the general DD population.

DDD is primarily an exome-driven study. Exome read-coverage varies substantially across the genome by design, to maximize limited sequence resources for the genomic locations most likely to disrupt genes. However, whilst such exonic read-coverage enrichment is desired for identifying genic point mutations, it is not necessarily optimal for the detection of large-scale abnormalities. Abnormalities may be harder to detect in genes with widely spaced exons or genes with fewer exons, although, this limitation is mitigated by the target size of event detection (2 Mb and greater). Indeed, analysis for mosaicism of approximately one thousand samples by SNP and exome platforms showed that exome analysis missed two of ten events detected by the SNP platform. Thus, it is likely that exome-based calculation of frequency would produce a slight underestimate because of platform differences.

5.3.2 Algorithmic

Uniparental disomy describes two homologous alleles originating from the same parent and reflects an inheritance aberration. UP*D*io detects abnormal inheritance as an enrichment of uniparental trio genotype configurations on a single chromosome and data for proband and both parents are required to assess inheritance. There are two

failure modes that disrupt UPD detection: 1) missing genotypes and 2) missing parental samples.

Extending the method to account for the first failure mode is fairly straightforward. This approach could work by phasing parental haplotypes and then imputing the genotypes that have failed genotyping. On a practical level, this would likely make little difference for UPD detection because the genotyping error rate is low and UPD events are sufficiently large to be detected even in the context of missing genotypes.

However, for DDD probands now not analysed for UPD because full trio data are not available, the development of a proband single-parent UPD software tool should be possible. The approach might first phase the child's haplotypes and the known parent's haplotypes, and then determine which known parental haplotype the child has inherited. Based on the child's genotypes and the available haplotypes in the population, the other parent's haplotypes could be assessed. Each of the child's haplotypes should derive from a different parent and a discrepancy could reflect UPD or inheritance by descent, the latter distinguished by occurrence on multiple chromosomes.

MrMosaic uses a backbone of autosomal polymorphic di-allelic point mutations from which heterozygous sites are extracted for B_{dev} and C_{dev} calculations. There are three ways to improve the number of assayed sites: first, the number of assayed sites could be increased by adding to this backbone rare and private polymorphisms in each patient; second, the C_{dev} information from non-heterozygous (i.e. homozygous) sites can still be used in detecting deviation in copy number, even though the B_{dev} is not informative; third, gonosomal sites can be included.

MrMosaic has not been tested on the gonosomes but this extension should be possible. Mosaicism of chromosome X will detect the genetic aneuploidies associated with mosaic Klinefelter Syndrome and Turner Syndrome, diseases identified with high frequency in the Conlin *et al*[36] study. Implementing MrMosaic on gonosomes requires an ADM score generated on a sex-specific pool of samples. Mosaicism of the chromosome Y may be less useful, as the XYY karyotype in itself does not result in abnormal phenotypes[249], although mosaicism involving Y may signal other pathogenic events, such as complex aneuploidy involving multiple chromosomes, or chimerism.

Interpreting the output of MrMosaic is fairly labour-intensive because at the Mscore cut-off (8) chosen to be sensitive to mosaic events of 2 Mb and despite filtering based on event detection frequency and exclusion of peri-centromeric regions, approximately one putative detection is made per sample. In this large experiment presented of 4,911 probands, manual curation of 4,643 putative detections was undertaken, which required approximately 12 hours. The full data set will involve approximately three times the number of samples. The number of putative detections for review can be reduced by increasing the Mscore threshold, but is likely to lower the sensitivity of detecting smaller events.

5.3.3 Number of diagnoses

In about half of the cases for which a UPD or mosaic structural event was detected, a direct association between that event and the child's pathology could not be determined. UPD has a prevalence in the general population of about 1 in 3,500 and should therefore appear at least once among the nearly 5,000 studied children here in a benign form. However, given the enrichment of UPD and mosaicism in children with DD compared to generally healthy children, it is reasonable to suspect that the majority of the detected events are pathogenic, although diagnosis has only yet been possible for about half of those with detected abnormalities.

The diagnostic workup differs for UPD events compared with large mosaic abnormalities. For UPD events, the main pathological mechanisms are imprinting disorders, recessive diseases, and incomplete trisomy. The detection of UPD events on imprinting chromosomes in children with manifestations of known imprinting disorders provides definitive diagnosis. The majority of UPD events detected in this study did not lie on chromosomes vulnerable to imprinting, nor were they implicated in incomplete trisomy rescue. Instead, many resulted in regions of isodisomy, which can result in conversion to homozygosity of a deleterious allele inherited from a carrier parent. Assigning pathology to such homozygous variants is challenging and requires at least three broad categories of evidence: the variant causes disruption in the gene, pathology results when the gene is disrupted, and that this pathology matches the phenotypes in the child. This is fairly straightforward when the identified homozygous variant is predicted to be loss-of-function (such as a nonsense mutation), loss-of-function mutations in that gene have been closely associated in a specific disease, and the child's phenotypes match the manifestations of that disease. Knowledge gaps in gene function

and disease-gene associations hinder pathogenic analysis and require further investment in gene function.

The diagnostic workup for structural mosaicism is similar to the assessment of structural variation as a cause of genomic disorders and relies heavily on disease databases. Genetic diagnosis is fairly straightforward if the copy-number event in the child has been observed in other children who share the same phenotypes as the proband. Partially clouding diagnostic assignment in mosaic structural abnormalities is the effect of clonality on physiological disruption; this requires the assumption that an abnormality in mosaic state causes phenotypes similar in quality (but perhaps less severe) than the corresponding constitutive state. The assessment of mosaic UPD is slightly more complicated because incomplete aneuploidy often coexists with imprinting or recessive defects.

UPD and mosaicism are only detected in about 1% of children in the DDD study, and even after comprehensive assessment of constitutive copy-number analysis and other genetic abnormalities detected in the exome, genetic diagnosis still lacks for the majority (69%) of children in DDD. Improvements in understanding of gene function and variant ascertainment are essential and will hopefully lead to substantial reductions in the number of undiagnosed children.

5.4 Future work

Given the limitations above and the increasing trend for larger datasets, there are exciting opportunities for improved methods, which invariably will expand our understanding of DD.

Future trends may benefit from increasing integration of datasets and algorithms. With respect to integration of data, many of the analyses presented in this dissertation have made direct comparisons of the use, suitability, and performance of SNP vs. exome array. However, studies often use multiple platforms to assay genetic variation given unique advantages offered by each platform. In DDD, SNP, exome and aCGH data were generated for thousands of probands. Therefore, it is reasonable to consider the development of a tool that can integrate data gathered by multiple platforms. For example, mosaic analysis using SNP and exome platforms could increase the number of sites by including both common and rare variation, inside and outside of coding regions. Trio data facilitate the possibility of a haplotype-aware version of

MrMosaic, which is challenging given the sparse distribution of exome data, but should be possible for WGS analyses.

With respect to integration of algorithms, UP*D*io and MrMosaic were designed to detect constitutive UPD and structural mosaicism but it may be possible to integrate these two functions into one software tool as subroutines or "plug-ins" that function in a larger part of pipeline. Next-generation sequencing technology provides a substrate for simultaneously assaying a wealth of genomic variation, including structural variation, uniparental disomy, and mosaicism. In addition, there are likely statistical methods that can be learnt from transcriptomics, as this field must deconvolute signals of expression or transcript-assembly from heterogenous collections of tissue-types. Joint analysis of mosaicism and disruptions in expression could yield fascinating insight.

One of the limitations of MrMosaic is the number of putative detections that require manual review and future work could better automate the filtering strategy. A hurdle in such an approach is the lack of a strong positive-control training set, relative to the negative-control dataset. It may suffice to create the positive-control dataset using simulations, and then real mosaic events could be incorporated dynamically as they are discovered. Approximate Bayesian Computation is a Bayesian statistical technique that can be used in the absence of a known underlying likelihood model but when the sampling distributions of parameters are available; this approach may be useful for this automated filtering application as simulation analyses can generate the sampling distributions needed for multiple parameters (number of probes, strength of signal, event frequency, distance to centromere) underlying putative detections.

Regions of heterodisomy on non-imprinted chromosomes without evidence of mosaic aneuploidy are not predicted to be damaging. Despite this, eight examples of such heterodisomic chromosomes were found in this dataset. This invites speculation that many of these heterodisomic events may be pathogenic, perhaps by mechanisms already known, such as hidden trisomy-rescue, or by entirely new mechanisms. Maybe UPD is incompletely penetrant for some chromosomes, or results in highly variable phenotypes, as suspected for chromosome 16. Experiments that investigate the effect of heterodisomy on expression may yield interesting insights.

Decreasing sequencing costs have enabled acceleration in DNA sequencing data availability. Whilst whole-genome sequence data is still expensive to generate and were not available for analysis, such data are likely to be available in future studies of children with DD. Such data will enable unprecedented discovery of smaller mosaicism.

The somatic point mutation rate is approximately 0.3e-9 per site per cell division[250]; therefore mosaicism arises *de novo* with nearly every cell division. Despite this ubiquity, mosaicism is elusive, only detected when present in at least approximately 3,000 cells (based on: standard microarray input requirements require 200 ng (about 30,000 'genomes-worth' of DNA assuming 6 pg per cell) and mosaicism minimal detection threshold is 10% clonality). Future work will benefit from the use of single-cell sequencing or high-depth sequencing to detect mosaicism of lower levels of clonality tissue-specific mosaicism. Intuition suggests that mosaic abnormalities may often result in an intermediate phenotype (i.e. are less severe) than constitutive abnormalities and that mosaic events with greater tissue involvement are more pathogenic. These assumptions are difficult to assess empirically because tissue-sampling resolution is poor, often limited to blood or saliva. Study of mosaic trisomy 21 has found that mosaicism was more frequent in epithelial-derived tissue compared to lymphocytes and that phenotypic severity is linked to mosaic clonality in a tissue-specific manner[251]. These findings highlight the importance of developing a greater understanding of the distribution of mosaicism for diagnostics (identifying the mutation) and prognostics (interpreting its severity and outcome).

Analysis of one structural mosaic abnormality predicted that the most likely generative mechanism was LOH-mediated mosaic reversion, a mechanism previously reported[252]. Recently, chromothripsis has been implicated as an additional reversion mechanism[253] and it is reasonable to hypothesise that additional reversion mechanisms may be uncovered. It is speculative but interesting to consider that reversion may be fairly common; the disconnect between the theoretically-predicted commonality of mosaicism and the poor ascertainment of such events lends credence to this possibility. Several questions for reversion remain for future study: How common is reversion? Are most reversion events triggered by genomic instability? Are reversion events 'in response' to an underlying physiological disruption or an indication that stochastic genomic instability is commonplace? Do other reversion mechanisms, such as single codon deletions, exist? Do reversion clones have a common ancestor? Is the age-related dissipation of epidermal neoplasms (skin moles) immunologic or genetic (reversion)? Nature uses LOH and chromothripsis as reversion mechanisms; can man harness these mutational events therapeutically?

5.5 And then...

Forecasting the future of genomics is a useful exercise for planning but can be challenging. James Crow stated about prediction, "for the near future, I can follow the principle...that tomorrow's weather is best predicted by today's...for a somewhat longer future we can extend current trends. But for the long-term future, we can only guess"[254].

5.5.1 Achieving a higher fidelity genome

There is tremendous societal investment in genomics with an estimated 796 billion US dollars investigated in genomics between 1988 and 2010[255]. Such investment has empowered technological innovation, leading to a 100-fold decrease in sequencing costs within the period between 1991 and 2001[256], and an accelerated 1000-fold decrease between 2008 and 2014[257]. Yet, the cost of sequencing a human genome by WGS today is still expensive, more than $1,000[257], which also does not account for ancillary costs, such as data storage and interpretation[258]. Illumina® "has essentially monopolized the high-throughput sequencing market"[259], controlling 75% of the general genomics market share and 90% of high-throughput sequencing. It is reasonable to predict that continuing investment in genomics will spur industry competition, which will continue to drive down sequencing costs. Additional sequencing methods, such as those that measure changes in electrical current[260] or pH[261] avoid the overhead of optics, are extremely fast, and seem likely to rise in popularity. Inevitably, sequencing costs and technological advances will produce a portable, inexpensive, fast, high-fidelity whole-genome & whole-epigenome sequencing tool, perhaps within 15 years.

The technical implications of this new sequencing era will be profound: 1) long read-length sequencing will enable *de novo* assembly as the primary form of genome reconstitution; 2) reduction of mapping artefacts and sequencing errors will identify genomic variation with greater confidence and will reduce the computational complexity of assembly; 3) high-confidence genotyping will lead to more efficient storage[262], as less intermediate data need to be stored; improved knowledge of population haplotypes will enable an even more compressed haplotype-reference version of storage; re-sequencing a sample will be sufficiently inexpensive if long-term storage is not possible.

5.5.2 Having achieved a higher fidelity genome

The development of third generation (long-read single-molecule) sequencing[56] will especially have important consequences on the assessment of structural variation. Long

read-lengths will greatly facilitate the detection of structural variation via *de novo* reconstruction of the genome[263]. The resulting genome-wide frequency-map of structural variation will provide an empirical catalogue of all haploinsufficient genes and greatly reduce the number of CNVs of unknown clinical significance. More broadly, as sequencing becomes routine, catalogues of all forms of genomic variation will begin to saturate with all possible combinations of non-lethal mutations; this will identify which gene knock-outs are tolerated[142] and improved allele frequency data will facilitate interpretation of mutations in children with DD.

In contrast to constitutive structural variation, the detection of *mosaic* structural variation may prove challenging for some time to come because of sampling difficulties. The detection of mosaicism requires increasing read- and tissue- sampling, but low error rates may reduce the impetus to sequence the genome to high-depth, and accessing multiple tissue types is invasive and therefore not likely to become commonplace. High-depth sequencing is likely to be a continued priority of the cancer genetics community and may yield important insights of distribution of mosaicism throughout the body. Perhaps, sequencing can one day be performed non-invasively, as seen with *in vivo* magnetic resonance spectroscopy[264] for metabolomics, which would profoundly improve the ease of tissue sampling.

Large collections of WGS data are likely to come from healthcare settings, and eventually from domestic and municipal sources. In the Cold Spring Harbor Laboratory Biology of Genomes conference in 2013, Dr. Mike Snyder presented research (a lecture entitled "Integrative personal omics profiling for monitoring healthy and disease states") demonstrating that the distribution of his microflora fluctuated in a consistent and characteristic pattern each time he had 'a cold'. Toilet sensors, in the form of 'smart plumbing', may provide a method to detect early infections (microbiome sequencing) and cancer (detection of new mutations previously characterised as cancer driver mutations). Analysis of sewage microbiota can demonstrate the viruses circulating in the community and inform on community diet[265] (some viruses are endemic to certain types of plants only, for example). Analogous to telemetry used in the clinical setting to identify arrthymias remotely, it may be in the public interest to screen municipal sewers to identify epidemics, for example.

The majority of detected genetic variation today has unknown biological significance. Yet, complex disease studies operate with the assumption that a great

number of variants exhibit low-level effects on phenotype. Higher resolution phenotyping is needed to better understand low-effect variants with better granularity. Currently phenotyping is largely restricted to external traits and standardised human terms[266] but phenotyping is likely to become increasingly molecular, quantitative, and comprehensive ('phenomics'). Computational interpretation of facial dysmorphology is beginning to overtake human performance[267] and the integrated analysis of deep phenotyping data, such as transcriptomics and metabolomics, is likely to exacerbate this gap. The detection of UPD events may one day more appropriately be detected directly, using disruptions in epigenetics and alterations in expression, than indirectly by genotype. It also may be the case that detection of altered transcription or metabolic products will trigger the investigation of low-clonality mosaicism in children with DD.

Further ahead, widespread use of genomics and phenomics perhaps may mean that computational representation of each person's genome and phenome is recorded. Family studies could be performed quickly, entirely using stored data. Social media may allow contact with others who are most genetically similar (yielding interesting implications in genealogy, such as tracing ancestry or finding relatives), or metabolically similar, perhaps finding those who share similar disease states.

5.5.3 Challenges further ahead

Despite the battle cry of exuberant contemporary research papers[268], *determining* the genetic cause of Mendelian disease is not the same as *solving* Mendelian disease. Recent advances have treated some metabolic deficiencies using enzyme replacement and gene therapy[269], and others suggest that reversion of phenotype in children with Rett syndrome and Down syndrome may indeed be possible[270,271]. Nevertheless, a cure for the vast majority of DD has not been found.

Some treatments for DD may require intervention during early embryonic life. Non-invasive prenatal testing (NIPT) is now widely used in the United States, with 90% of pre-natal genetic counsellors having integrated NIPT into their clinical practice[272]. Currently NIPT is limited to detection of foetal aneuploidy and large structural variation but advances in genomics will inevitably lead to the incorporation of whole-genome sequencing in NIPT and the detection of pathogenic variation.

Many of the challenges in medical genetics ahead will be ethical. Intervention on human embryos has already generated substantial ethical debate, with respect to selective abortion[273,274], the right to access a child's genome[275], and whether gene

editing of human embryos[276] should be allowed[277]. It seems inevitable that genomic editing will be eventually welcomed, even by pro-life activists, as a method to cure a child's disease, in a way that preserves the child's life. The privacy implications of databasing and reporting of personal genomics are certain to become contentious but likely to become adopted given the potential impact on medicine and health.

Challenging questions ahead relate to analysis, thorough space and time, of transient and tissue-dynamic components of genomic activity, such as transcriptomics, metabolomics, and 3-dimensional chromatin architecture. The new concept that the 3D layout of the genome is informative[278] is exciting and throws dirt over the grave to the concept that non-exonic genomic regions are 'junk'[279] (although I sympathise with the somewhat unpopular view that much of the genome probably has little biological function[280], despite the widely publicised claim to the contrary[281]). Notwithstanding technical limitations to High-C technology[282], the field now appreciates that intergenic regions hold regulatory value[283] and the way chromatin is spaced is important[284]. It should be possible to quantify how important each DNA base is in terms of the spacing and positioning of regulatory elements beside their targets, a 'white-space' metric of the genome. For aneuploidy, in addition to disruption of gene dose, what proportion of pathogenesis is contributed by the disruption of long-range interactions and regulatory spacing?

DNA, like the heavens, once had complexity seemingly beyond reach. A breakthrough in cosmology research, the construction of a three-dimensional map of our local galactic neighbourhood, has just been completed[285]. Efforts to create a 3D map of the genome may benefit from a cross-disciplinary collaboration involving the mapping techniques of astronomers, the expertise of physicists in electrostatic interactions, and the biological experience held by genomicists. Eventually such maps of our genome will be available and if fortune grants me the opportunity, I would be eager to explore them.