



On the Trail of Genomic Pioneers



Meet Dr. James R. Lupski, M.D., Ph.D.
Department of Molecular and Human Genetics
Baylor College of Medicine

1) Would you tell us a bit about your research interests?

My research focuses on using human genetics, genomics, and molecular biological techniques to elucidate biological mechanisms for disease. We have concentrated over the last couple of decades specifically on neurological diseases and trying to understand their bases using the techniques delineated above. What we have learnt is that structural changes of the genome can be extremely important to nervous system development and function. This is evident by both developmental abnormalities and neurodegeneration being associated with structural changes of the human genome that lead to copy number variation or CNV. Thus, in many neurological diseases, it is not the classic model of mutations within a gene that are responsible for disease, but instead just the change in copy number of a specific segment of the human genome.

2) To what extent are genomic changes and variation responsible for diseases?

This is a good question, and in fact it is only time and future experimental observations that will truly inform us. What I can say is that structural changes, and in particular CNVs, may indeed be responsible for a lot of sporadic diseases, given the fact that locus specific mutation rates for CNV mutations can be 100 to 10,000 times more frequent than locus specific mutation rates for single base pair changes or SNPs (single nucleotide polymorphisms). It may also be that the

nervous system in particular is more susceptible to copy number changes than other biochemical pathways, or networks of interacting proteins in different physiological systems, within the body. We do believe that CNV may be responsible for a lot more disease than currently believed or anticipated. The focus to date in approaches to human genetic traits has been, whether by virtue of the Mendelian model or a hypothesis generating search through genome-wide-association studies (GWAS), the focus has been on base pair changes as being responsible for the variation leading to the genetic cause or susceptibility to disease traits. However, if we look just in the past year or two alone we have witnessed that when researchers open their mind to CNV as a mutational mechanism, and analyze their data with that thought in mind, one can find CNV being responsible for conditions such as autism and schizophrenia for which we previously have had a difficult time identifying specific genetic etiologies. Thus, I believe the future is very exciting and such future experiments will delineate the extent to which CNV is responsible for human disease phenotypes. This may be very different for different portions of the genome and some of that difference may reflect local genome architectural features that foment genome instability.

3) How are copy number variation and GWAS studies helping in understanding genomic disorders?

GWAS that assay for CNV versus SNP have found specific loci in the genome in which CNV appear to cause autism and schizophrenia. This now has given us a "window of opportunity" to



On the Trail of Genomic Pioneers



Meet Dr. James R. Lupski, M.D., Ph.D.
Department of Molecular and Human Genetics
Baylor College of Medicine

go and dissect those regions of the genome and determine how the CNV contributes specifically to that condition and how that CNV mediates its functional consequences in the genome to bring about the phenotype. Thus, I believe that CNV may be important for many other conditions in which there appears to be genetic susceptibility. In particular, complex traits and complex trait analyses must consider the fact that human genetic variation often consists of not only SNP, but there is a great extent of human genetic variation that is due to CNV.

4) Which research study or work has strongly influenced your thought and research goals?

I would have to say that a lot of work in the bacterial genetics field has influenced my thinking specifically as it relates to genomic rearrangements, in particular duplications of the genome, and movable genetic elements or transposable elements. The mutability, especially with respect to structural changes, has been beautifully experimentally demonstrated in bacteria. In fact, my early scientific training and Ph.D. were in the area of bacterial genetics studying bacterial chromosomal DNA replication as well as transposable genetic elements of bacteria. More recently, (i.e. the last decade) I have certainly been influenced and follow the entire emerging field of genomics in particular the human genome project. As both a medical geneticist and human being our genome is of immense interest! I would say that our work has been influenced by, and

has influenced, the work of others in the human genomics field related to architectural features of the human genome; in particular segmental duplications or low copy repeats and their role in susceptibility to disease and human and other non-human primate genome evolution.

5) Where do you see your research leading in the future?

I feel that that our research will continue to try and elucidate the molecular mechanisms for genomic rearrangements associated with human disease phenotypes and try to translate some of this mechanistic information into clinically useful tools that apply high resolution human genome analysis in the clinic. Such technologies really extend the reach and the vision of classic conventional clinical cytogenetics. With this extended reach and high resolution, we can potentially find genetic etiological causes for many of our patient's specific ailments and sporadic birth defects. Thus, I believe there will be a continued need to develop an understanding of the mechanisms of rearrangements and the architecture of the human genome to be able to better understand what regions of the genome undergo genomic instability and how to assay such regions in a clinical context.

Furthermore, I also believe that we will be able to use such techniques in the screening of patient populations to actually identify the functional consequences of CNV and as such take a forward genomics and reverse genomics approach to defining such functional consequences of CNV. What I mean by this is for the previous century of genetics, the focus was on



On the Trail of Genomic Pioneers



Meet Dr. James R. Lupski, M.D., Ph.D.
Department of Molecular and Human Genetics
Baylor College of Medicine

forward genetic screens to map a locus specifically involved in the genetic contribution to a given phenotype and then by the 1970s and 1980s, when we really had robust recombinant DNA and molecular biological techniques, we could make mutations and then in a reverse genetics approach study, the functional phenotypic consequences of those mutations. We can now use high resolution human genome analysis in a clinical setting in a "forward genomics approach" to identify regions of the genome that correlate with the clinical phenotypes; and then using reverse genomics, collect many patients with the same region of the human genome rearranged and look at the functional consequences of those specific rearrangements. In essence, start to robustly establish "genomotype/phenotype" correlations rather than "genotype/phenotype" correlations.

The focus of the previous century has been very, very "genocentric". However, the functional unit now perhaps should change from the gene to the genome. The genetic code was a remarkable triumph for genetics, but remember coding sequences in the human genome account for <2% of all sequences. Thus, for 98% of the genome, we are not going to be able to use only the genetic code to figure out what are the functional consequences of changes to the genetic material. In fact, if we really look at where we are now, of the 20,000-25,000 human genes in our genome that it takes to make us human, we do not even understand the biological role, or the functional consequences of mutations, in 5-10% of

them. So, perhaps the forward genomics and reverse genomics approaches may enable the functional consequences of a significant percentage of the remaining 90% of genes that need to be "figured out" or understood. Perhaps of equal importance, then we have to try to get at the remaining 98% of the genome, less than 2% of which is coding and understood in terms of the genetic code, that makes us what we are and this indeed will be enabled by studying structural changes of the genome. So, in fact, what I think our future research and many other laboratories throughout the world's research will try to get at is: What is the genomic code? Again, we know the genetic code, but coding sequencing account for a very small percentage of the human genome.

6) A great deal of your work focuses on the study of human genomic rearrangements. What would you say are the most important things that have been discovered about this over the years and what will be the role of Genomics in it in the years to come?

I would say that the most important things discovered regarding human genomic rearrangements during the last decade and a half is the fact that rearrangements that only affect the dosage of a specific gene, that is, they do not in anyway perturb the coding sequence of a gene, can still cause disease. Furthermore, CNV can be responsible for many, many diseases, and the *de novo* locus specific mutation frequencies can be very different around the genome and can be incredibly frequent and probably responsible for a lot of sporadic traits and birth



On the Trail of Genomic Pioneers



Meet Dr. James R. Lupski, M.D., Ph.D.
Department of Molecular and Human Genetics
Baylor College of Medicine

defects. I would also say that I think it has been extremely important to determine the mechanisms for structural rearrangements. The many years that we and others spent on elucidating the mechanism that has been come to be termed NAHR (non-allelic homologous recombination) have been extremely helpful because by learning the rules this enabled specific predictions about the regions of genomic instability within the human genome.

The epitome of understanding of the rules has been the beautiful work by the Evan Eichler laboratory in Seattle. After the rules had been determined for the mechanism of NAHR: including the requirement for low copy repeats or segmental duplications of a certain size, and of a certain percent identity, and within a given distance from each other, with direct orientation leading to deletion/duplication and inverted orientation leading to inversion when utilized as substrates for NAHR; applying that knowledge computationally to the draft and then "finished" human genome sequence, led to predictions of what might be regions of the human genome that underwent genomic instability via this NAHR mechanism. Indeed Eichler and colleagues designed arrays to interrogate such regions and applied such "informed arrays" to the study of patient populations with different phenotypes and were thereby able to elucidate many new genomic disorders in a very short time. This approach of "going genome first" then to disorder has been indeed a very powerful approach by the Eichler laboratory and a

vindication of the importance to knowing mechanism for genomic rearrangements and understanding the rules of how they occur to then enable predictions. Such predictions are now being incorporated into high resolution human genome analysis and array designs for those assays that are being implemented in clinical medicine.

The FoSTeS/MMBIR is another example, but a very new example, whereby getting at a mechanism for genomic rearrangement changes one's thinking about the problem and approach to the problem. Eventually, hopefully, we will elucidate enough of the rules regarding this new DNA replication mechanism that may enable us to make predictions regarding which regions of the human genome are perhaps more likely to undergo genomic instability due to that mechanism because of unique architectural features that characterize that region of the human genome. Also, as we recently showed in a paper that appeared in the July issue of *Nature Genetics*, this DNA replication mechanism occurs mitotically (as anticipated) and the mechanism can have implications for genetic counseling regarding recurrence risk.

7) What are genomic disorders? Which genomic approaches are most promising in finding treatment for complex human diseases?

Genomic disorders have been clearly defined and I would certainly refer anyone with interest to the original description and elucidation of the concept of genomic disorders that was published in *Trends in Genetics* in 1998. For a more recent



On the Trail of Genomic Pioneers



Meet Dr. James R. Lupski, M.D., Ph.D.
Department of Molecular and Human Genetics
Baylor College of Medicine

perspective and overview I just recently published an opinion piece in *Genome Medicine* that describes some of our thought processes leading up to the elucidation of genomic disorders. In essence, there are two specific important corollaries to the genomic disorders concept: i) these are conditions that are due to genomic rearrangements and not sequenced based changes or DNA base pair alterations, and ii) that genome architecture incites genome instability. Thus, this is a class of disorders that didn't necessarily follow the major paradigm of the time with respect to genetics and inheritance of genetic information and alterations (i.e. mutations) and how alterations in genetic information occur. The prevailing concept, since the elucidation of the Watson - Crick model, has been that base pair changes to a gene and alterations in the genetic code are the predominant genetic variations that can have functional consequences and lead to phenotypes including disease phenotypes. That is, THE major driver of mutation is Watson-Crick base pair changes of the DNA molecule. However, in genomic disorders you don't have to have any base pairs changed, you don't have to change the structure of any specific gene, you don't have to alter the genetic code and yet you could have a phenotype purely based on a structural change that can cause a copy number variation of a gene or genes, or even regulatory elements in different genomic regions. Thus, CNV is what causes genomic disorders and certainly there was not much of an appreciation that humans even have a lot of CNV until work over the past five

years revealed the tremendous extent to which CNV is responsible for human genetic variation. With respect to the question: What are some of the most promising potential treatments for complex human diseases and genomic approaches to diseases? I would say that there are two major areas. The first would be actually using genetic approaches to get at the etiology of complex traits and this in and of itself will result in a better understanding of disease; with understanding comes thought processes as to how you intervene with respect to therapeutics. I also think the CNV findings are very exciting because if a disease is due to a CNV, and furthermore, you can identify and elucidate perhaps a specific gene within the CNV segment that is a dosage sensitive gene for which change in copy number may elicit the clinical phenotype, then our thinking has to be very different from the way we previously thought. With respect to genetic disease, with the genocentric focus and the singular focus of mutation being base pair changes, we have previously thought of therapy in terms of the gene is broken, let us fix the gene or put in a new gene to deal with the problem. However, with disease due to CNV, there is often no gene broken so there is no need to put in another copy of a gene to compensate for a loss-of-function allele, or some other perhaps gain-of-function or dominant-negative. If the disease is indeed due to dosage, then correcting the dosage may be enough to mitigate the functional consequences of that CNV. Thus, we may be able to open whole new avenues of thinking and approaching disease by addressing the question: How do we mitigate the consequences of expression differences? This can be done potentially through epigenetic modifications as a



On the Trail of Genomic Pioneers



Meet Dr. James R. Lupski, M.D., Ph.D.
Department of Molecular and Human Genetics
Baylor College of Medicine

way to alter gene dosage. Perhaps for diseases due to slight over expression, due to duplications of segments of the genome, siRNA could have a major role in treating such conditions. For others dosage diseases, chromatin remodeling and other epigenetic alterations could be used. Some recent exciting work by Michael Sereda and colleagues in Germany with an animal model for Charcot-Marie-Tooth neuropathy due to the CMT1A duplication uses progesterone antagonist as an epigenetic means to down regulate PMP22 gene expression. I do think that this next decade will be extremely exciting when we can try to better understand what are the genetic contributions to disease and how does this perturb networks and systems and what ways can we try to get those networks and systems back into some kind of equilibrium or biological balance. Often, one may only require very subtle changes that could have profound consequences for millions of patients. For instance, very subtle changes in the pathological consequences that could lead to age-related macular degeneration (AMD) or Alzheimer disease may extend the onset from 65 years of age to 75 years of age and this would dramatically change the number of individuals affected with such clinical phenotypes.

8) Can you please talk us through the DNA replication-based mechanism of fork stalling and template switching (FoSTeS) for studying complex genomic rearrangements?

Indeed this is a bit of a complex concept. The important point is that we proposed this mechanism because of our wanting to explain our experimental observations. These observations were that often genomic rearrangements, in particular duplications, were not simple tandem duplications of a segment of the human genome, but instead had complexities wherein pieces of the genome have been brought together from different and disparate locations and even fused in different orientations direct, inverted, etc. Thus, this mechanism proposed a template driven juxtaposition of different genomic segments from discrete locations. At the join points, it was noticed that one often (if not always) found evidence for microhomology consistent with a microhomology mediated process. Thus, the mechanism proposed that this microhomology mediated process was reflecting priming of the transferred strand when there was a switching of templates. The specific details remain to be further elucidated and it's not really certain whether the fork is actually stalling at a specific location or perhaps the fork proceeds through a single stranded nick, resulting in either a collapsed fork or a one-ended, double-strand DNA as opposed to a two-ended, double-strand break. Furthermore, it is not absolutely clear if there is a single-strand transfer at a stalled fork, or after there is a collapsed fork does one then have a break induced replication (BIR). In fact, a more general mechanism that seems to be more applicable, and accounts for lots of experimental observations from other model organisms including humans, is the so called MMBIR mechanism (Microhomology Mediated Break Induced Replication) recently published with Phil



On the Trail of Genomic Pioneers



Meet Dr. James R. Lupski, M.D., Ph.D.
Department of Molecular and Human Genetics
Baylor College of Medicine

Hastings and Grzegorz Ira, both colleagues here at Baylor College of Medicine, in *PLoS Genetics*. In human genetics we had to try to explain the phenomena that we observed. Our experimental observations were not explainable by the then current models of NAHR, nor by nonhomologous end joining (NHEJ); unless one invoked some complicated (i.e. less parsimonious) idea that many different double-strand breaks occurred each with a NHEJ mechanism, but with a precise fusing of these together (often with different copy number of the constituent rearranged genomic intervals; e.g. duplications or triplications) and observing microhomology at the "break point" or what we call "join point". Such a "multiplex NHEJ model" seems much less plausible than the FoSTeS/MMBIR mechanisms.

9) According to you what are the molecular mechanisms for genomic rearrangements?

You can go to the literature and find that we have recent reviews that we hope better elucidate and make more understandable some of these mechanisms; in particular reviews in the *Annual Reviews of Genomics and Human Genetics*, *Trends in Genetics*, and *Nature Reviews Genetics*, each focusing on different aspects of mechanism for structural variations and CNV. At this point I would say that there are at least four mechanisms by which human genomic rearrangements occur. There are indeed some simple (in particular deletion) rearrangements that appear to

occur by NHEJ or nonhomologous end joining. There are also many simple deletions and duplications that occur through another recombination based mechanism which we termed nonallelic homologous recombination or NAHR. The fact that the human genome is laden with segmental duplication (SD) or low-copy repeats (LCRs) that can act as substrates for NAHR makes humans particularly susceptible to this rearrangement mechanism. The NAHR mechanism certainly seems to be the predominant mechanism for recurrent rearrangements (i.e. same genomic interval rearranged in different patients) in the human genome and we are currently at a level of understanding that enables us to make predictions about which regions of the human genome will undergo genomic instability potentially due to this mechanism. For many nonrecurrent rearrangements, in particular duplications, and more specifically those that may be complex in their nature, it appears that FoSTeS/MMBIR is a very plausible mechanism and to my knowledge there have not been many competing mechanisms put forward that can explain all the observations regarding complex human genome rearrangements. A fourth mechanism is transposition and movable genetic elements in the human genome. If I had to make a guess at this juncture, I would suggest that both NAHR and FoSTeS/MMBIR may be more important mechanisms in real time, that is with respect to new mutation CNV and sporadic traits such as birth defects; whereas some of the transposition mechanisms may be acting over evolutionary time scales as opposed to "real time". Thus, for human genomic rearrangements, there is substantial evidence for at least four



On the Trail of Genomic Pioneers



Meet Dr. James R. Lupski, M.D., Ph.D.
Department of Molecular and Human Genetics
Baylor College of Medicine

mechanisms and those include: NAHR, FoSTeS/MMBIR, NHEJ, and transposition. These mechanisms may act at different frequencies and to different extents in different areas of the genome and may be influenced by the local or regional human genome architecture. Furthermore, the different mechanisms may play more prominent roles in either evolutionary genomic rearrangements versus new and recent germline and somatic mutational events in the human genome.

USA: +1-800-436-3564 (Toll free) | India: +91-40-6698-6700

www.genelogic.com

www.ocimumbio.com

✉ outreach@ocimumbio.com

Ocimum Biosolutions Ltd. ©2009-11. All Rights Reserved.

The information contained in this document is proprietary to Ocimum Biosolutions Ltd. All other registered trademarks are the property of their respective owners. Any forward looking statements regarding future product direction are for information purposes only and Ocimum Biosolutions makes no express or implied commitments or representations concerning the timing / content of any future functionality or releases.

www.ocimumbio.com