

Final version, as accepted for publication in Molecular and Cellular Endocrinology

Published version at <http://dx.doi.org/10.1016/j.mce.2008.06.005>

Perspective:

From molecular endocrinology of mouse to genomic endocrinology of animals

Marc Robinson-Rechavi^{1,2}

1: Department of Ecology and Evolution, Biophore, Lausanne University, CH-1015 Lausanne, Switzerland; e-mail: marc.robinson-rechavi@unil.ch

2: Swiss Institute of Bioinformatics, CH-1015 Lausanne, Switzerland

This short perspective explores some ways in which new genomic methodologies impact the study of endocrine signaling. Emphasis is put on the impact on studying species which are not molecular biology models. This opens the door to using knowledge molecular endocrinology in areas of biology as distant as conservation biology, as well as enriching endocrinology with information from biodiversity and natural variation.

Endocrinology was from its birth an interdisciplinary field of research, combining notably physiology with organic chemistry. Molecular endocrinology resulted from the impact of the molecular biology technological revolution on this field [1]. Molecular biology focused the attention of researchers on the small number of species which were most easily amenable to these techniques. Now a new revolution is ongoing, from molecular biology to genomic biology. One difference between the previous revolution and the new one is the number of species which can be reasonably studied. Molecular biology techniques provide small parts of the picture at a time. To make sense of a result, it needs to be put in the broader context of previous molecular biology results. This means that an experiment will be much more powerful if performed on the same species which everyone else is studying. Moreover, some of the most powerful experimental techniques (e.g. gene Knock-out) are difficult to set up for new species, providing an additional incentive to work in established models. Hence a focus on very few species, such as the lab mouse, fruit fly or nematode worm. This has affected endocrinology like other fields, and molecular endocrinology is in large part molecular endocrinology of the lab mouse.

Given the huge amount of information gathered for these molecular biology models, they will obviously stay with us. But genomic techniques offer the opportunity of a revival of the study of endocrinology in other species for three reasons: (i) an immediate global view, which means that a large body of pre-existing information in the species of interest is less critical; (ii) easy transfer of technology between species; and (iii) the exponential decrease in sequencing costs [2,3]. Starting with the last point, data in biology seems to be increasingly divided into two categories: that which can be reduced somehow to DNA sequencing, and that which cannot. Examples of the first are many and increasing. Let us cite: primary DNA sequences (obviously); levels and patterns of expression (ESTs, SAGE, MPSS); transcription factor binding sites (ChIP-seq); polymorphism, of sequence but also of copy number; epigenetic information [e.g. 4]; and even biodiversity (genetic bar code projects). Examples of non reducible data include protein tertiary sequences or mutant phenotypes.

The most obvious information provided by the new high throughput techniques is sequence. Why is that important? It was possible to sequence an hormone receptor in a frog before any genomic revolution. A first point is that sequence based homology inference increases in power with the abundance of already known sequences. When more sequences are known, orthology inference is more accurate, patterns of conservation of key residues are better characterized, distant homologs can be detected with greater sensitivity, and constraints on genomic organization are better understood. Thus in a way, the more we sequence, the more we can learn from a new sequence. This explains the large impact of the increasing number of genome sequencing projects: as of February 2008, there are genome sequences from 39 animal species in the UCSC genome browser [5], and the GOLD database [6] lists 328 animal genome projects. This means that increasingly we do not have to choose between studying biodiversity, and investigating complete endocrine systems. For a detailed discussion of the application of comparative genomics to endocrinology, see Markov et al. in this issue. In brief, we already have the possibility to list not only all the receptors of species whose endocrinology has been hardly studied previously, but also to infer which ligands may or not be synthesized *in vivo*, by comparing their enzymatic pathways.

In a single gene study, we will often ask first: where and when in the gene expressed? Transcriptome techniques, especially microarrays, have greatly changed the way we consider this question. But the impact has been again especially felt in model species, for which relatively complete and specific oligonucleotide arrays can be designed based on high quality genome sequences, with costs reduced by the large number of potential users of each array. EST spotted arrays are already feasible for non model species, but have important limitations: data is difficult to compare between experiments and has low reproducibility [7], and a biased subset of genes are covered, depending on the source of ESTs. Diminishing costs of HT sequencing provide a new opportunity to study the transcriptome of non model species. Indeed, while methodological investigations are still at an early stage [8], among the first applications are studies of transcriptomes of organisms for which we have no genome sequence and little to no molecular biology: the transcriptome of a wasp [9] and of a butterfly [10]. In the wasp study, notably, insulin related genes were found to be differentially regulated between the brains of individuals which show maternal care behavior (foundresses and workers) and those of individuals which do not (queens and gynes) [9]. Such a result, linking insulin signaling with a complex eusocial behavior in a non model organism, would have been impossible without the conjunction of high throughput sequencing technology (in this case 454), and related genome sequences for comparison, in this case the bee genome.

Moving back to the sequence, an important aim of molecular endocrinology is to link genetic variants with differences in endocrine signaling. Success has been limited in mammalian models, even mouse, and indeed most of the literature concentrates on reports concerning small samples of human patients. The new HT sequencing techniques offer interesting opportunities in this area. Indeed, they have been called "re-sequencing techniques" [2], because of the combination of short reads, ill suited to de novo assembly of a genome, and of large sequencing depth, well suited to discovering polymorphisms. Until now, most polymorphism data outside human has been done on domestic varieties or lab strains [see dbSNP, 11]. But these are a rather biased sampling, which does not reflect selective pressure on genes in natural conditions. Moreover, to make the best of polymorphism data, it is preferable to integrate it with information on population structure and history. The latter is often studied in species which are not molecular biology models, but are good candidates for the new sequencing techniques. Progress in bioinformatic treatment of HT sequencing data opens the door to detecting not only single nucleotide polymorphisms (SNPs), but also copy number variation (CNV) [e.g 12]. Although this remains to be shown conclusively, it is probable that such variation plays a more important role in phenotypic diversity. Abundant data from large samples from the wild will provide the opportunity to not only list variations, but also quantify selective forces, such as positive selection or balancing selection (maintaining variability for functional reasons). With sufficient statistical power, it should be possible to link selection to specific loci on the genome. Comparative genomics can then provide the further link to functional annotation. For example it would be interesting to explain quantitatively the variance in onset of metamorphosis in wild amphibian populations by variations over the whole cascade of thyroid hormone signaling, not limited to a few target genes (e.g. TRs).

An aspect of natural variation that has been largely under-studied, yet may be the most relevant data to link genotype and phenotype, is gene expression variation [13,14]. The limiting factors of such studies have been similar to those on classical transcriptome studies, with an added issue of cost. A survey of wild populations for e.g. microsatellite variation must sample tens of individuals in several populations. Microsatellites may be cheap to characterize, but they do not inform us about functional changes in the genome. At present, even the new technologies remain too expensive to sample tens of individuals for variability in gene expression with high quality. But the exponential decrease in costs [3] means that this is within reach. Thus there is hope that in the next years, we will be able to scan wild populations of interest for variability in gene expression, and identify relevant signaling

pathways with good reliability. Considering the relevance of hormones to many life history traits, such as age at reproduction, metamorphosis, rates of growth, or homeostasis under varying conditions (e.g. temperature, water salinity), it is probable that such non targeted scans of wild populations will considerably enrich our understanding of endocrine signaling in nature.

A specific application of this approach is Conservation Biology, the study and management of endangered species. It is also an area where the intersection of molecular endocrinology, genomics and biodiversity has a large potential. For example, many species of fishes suffer from recently distorted sex ratios. A sex ratio bias will diminish efficient population size, thus depleting genetic diversity. An excess of males is especially damaging, as it leads to a population size crash [reviewed in 15]. These distorted sex ratios may be caused by endocrine disruptors or by changes in temperature (including effects of global warming), and the study of specific cases is complicated by the diversity of sex determination mechanisms in fishes [16,17], which are often poorly understood. In any case, it is clear that endocrine signaling is central to the mechanisms which lead some salmonid populations to have more than 90% of males (Claus Wedekind, pers. comm.). Some genomic data is emerging in aquaculture fishes [e.g. 18], but so far molecular biology has been slow in providing the tools to understand the life cycle of endangered species. In an example of what can be done at present, with a heavy experimental investment, Kidd et al. [19] showed over a 7 year study that low concentrations of estrogenic substances could lead to a near collapse of a lake population of fathead minnows. Molecular variations were recorded, but limited to one candidate gene, vitellogenin.

A future protocol could start for example with HT sequencing of male and female transcriptomes, in young individuals which can be sexed phenotypically. This could provide candidate genes to follow by qRT-PCR during development of many individuals, from small biological samples (e.g. scales). Then we might be able to determine which pathways (e.g. estrogen signaling) are involved in early and late sexual differentiation, and whether the biased ratio is present at birth, dependent on maternal life conditions, or whether a shift occurs during development. Finally, these results could be integrated with data on transcriptome variation between wild populations, related to their demographic status and the quality of their environment. This may provide more mechanistic, predictive models of species management, integrating molecular, phenotypic and populational approaches in ways which were out of reach up to now. By not depending on the construction of specific resources for each species, but combining as in the wasp study [9] de novo sequencing of

transcriptome with comparison to a reference genome, this type of protocol may be applied to many different species. Thus specific molecular information may be gathered in each case, across the diversity of sex determining mechanisms and of environmental factors (such as pollution, temperature, or inbreeding).

In conclusion, a major effect of new genomic technologies could hopefully be to jump in large part the "molecular and cellular endocrinology" step for many non mammalian species, and bring them directly from traditional phenotype centered endocrinology and physiology into the realm of global molecular studies of endocrine signaling. This holds the potential of enriching both our understanding of biodiversity, and of endocrine signaling itself.

Acknowledgements

Research in the laboratory of MRR is supported by the EU program Crescendo, Swiss National Science Foundation grant 116798, Etat de Vaud, the Swiss Institute of Bioinformatics, and the Decryphon program.

- [1] Tata, J.R. (2005) One hundred years of hormones. *EMBO Rep* 6, 490-6.
- [2] Bentley, D.R. (2006) Whole-genome re-sequencing. *Current Opinion in Genetics & Development* 16, 545-552.
- [3] Service, R.F. (2006) GENE SEQUENCING: The Race for the \$1000 Genome. *Science* 311, 1544-1546.
- [4] Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S. and Crawford, G.E. (2008) High-Resolution Mapping and Characterization of Open Chromatin across the Genome. *Cell* 132, 311-322.
- [5] Karolchik, D., Kuhn, R.M., Baertsch, R., Barber, G.P., Clawson, H., Diekhans, M., Gardine, B., Harte, R.A., Hinrichs, A.S., Hsu, F., Kober, K.M., Miller, W., Pedersen, J.S., Pohl, A., Raney, B.J., Rhead, B., Rosenbloom, K.R., Smith, K.E., Stanke, M., Thakapallayil, A., Trumbower, H., Wang, T., Zweig, A.S., Haussler, D. and Kent, W.J. (2008) The UCSC Genome Browser Database: 2008 update. *Nucleic Acids Res* 36, D773-9.
- [6] Liolios, K., Mavromatis, K., Tavernarakis, N. and Kyrpides, N.C. (2008) The Genomes On Line Database (GOLD) in 2007: status of genomic and metagenomic projects and their associated metadata. *Nucl. Acids Res.* 36, D475-479.
- [7] Kuo, W.P., Liu, F., Trimarchi, J., Punzo, C., Lombardi, M., Sarang, J., Whipple, M.E., Maysuria, M., Serikawa, K., Lee, S.Y., McCrann, D., Kang, J., Shearstone, J.R., Burke, J., Park, D.J., Wang, X., Rector, T.L., Ricciardi-Castagnoli, P., Perrin, S., Choi, S., Bumgarner, R., Kim, J.H., Short, G.F., Freeman, M.W., Seed, B., Jensen, R., Church, G.M., Hovig, E., Cepko, C.L., Park, P., Ohno-Machado, L. and Jenssen, T.-K. (2006) A sequence-oriented comparison of gene expression measurements across different hybridization-based technologies. *Nat Biotech* 24, 832-840.

- [8] Torres, T.T., Metta, M., Ottenwalder, B. and Schlotterer, C. (2008) Gene expression profiling by massively parallel sequencing. *Genome Res.* 18, 172-177.
- [9] Toth, A.L., Varala, K., Newman, T.C., Miguez, F.E., Hutchison, S.K., Willoughby, D.A., Simons, J.F., Egholm, M., Hunt, J.H., Hudson, M.E. and Robinson, G.E. (2007) Wasp Gene Expression Supports an Evolutionary Link Between Maternal Behavior and Eusociality. *Science* 318, 441-444.
- [10] Vera, J.C., Wheat, C.W., Fescemyer, H.W., Frilander, M.J., Crawford, D.L., Hanski, I. and Marden, J.H. (2008) Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Molecular Ecology* 0, ???-???
- [11] Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M. and Sirotkin, K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 29, 308-11.
- [12] Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S.F.A., Hakonarson, H. and Bucan, M. (2007) PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* 17, 1665-1674.
- [13] Whitehead, A. and Crawford, D.L. (2006) Variation within and among species in gene expression: raw material for evolution. *Molecular Ecology* 15, 1197-1211.
- [14] St-Cyr, J., Derome, N. and Bernatchez, L. (2008) The transcriptomics of life-history trade-offs in whitefish species pairs (*Coregonus* sp.). *Molecular Ecology* in press.
- [15] Cotton, S. and Wedekind, C. (2007) Control of introduced species using Trojan sex chromosomes. *Trends in Ecology & Evolution* 22, 441-443.
- [16] Devlin, R.H. and Nagahama, Y. (2002) Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. *Aquaculture* 208, 191-364.
- [17] Volff, J.N. (2004) Genome evolution and biodiversity in teleost fish. *Heredity* 94, 280-294.
- [18] Denise Vizziano, G.R.D.B.C.C.Y.G. (2007) Characterization of early molecular sex differentiation in rainbow trout, *Oncorhynchus mykiss*. *Developmental Dynamics* 236, 2198-2206.
- [19] Kidd, K.A., Blanchfield, P.J., Mills, K.H., Palace, V.P., Evans, R.E., Lazorchak, J.M. and Flick, R.W. (2007) Collapse of a fish population after exposure to a synthetic estrogen. *PNAS* 104, 8897-8901.