Comisión Nacional de Investigación
Científica y Tecnológica – CONICYT

FONDAP CENTERS OF EXCELLENCE IN RESEARCH PROGRAM

ANNUAL PROGRESS REPORT

Center for Genome Regulation (CGR)

2011

Comisión Nacional de Investigación Científica y Tecnológica – CONICYT

I. PRESENTATION

PERIOD REPORTED: 1st Year [X] 2nd Year [ ] 3rd Year [ ] 4th Year [ ] 5th Year [ ]

PERIOD COVERED: From: December 23, 2010 To: December 31, 2011

NAME OF THE CENTER **Center for Genome Regulation (CGR)**		**CODE** 15 09 00 07
DIRECTOR OF THE CENTER Miguel L Allende	**E-MAIL** **allende@uchile.cl**	**SIGNATURE**
DEPUTY DIRECTOR Martín Montecino L.	**E-MAIL** **mmontecino@unab.cl**	**SIGNATURE**
SPONSORING INSTITUTION Universidad de Chile		
ASSOCIATED INSTITUTION(S) (if applicable) Universidad Andrés Bello; Pontificia Universidad Católica de Chile		
CENTER WEBSITE ADDRESS www.centrogenomica.cl		

DATE: 01/01/12

RESEARCH LINES

N°	Research Line	Objective	Principal Researcher*	Associated Researcher(s)
1	Epigenetic control of differentiation and stem cell biology	To contribute to the understanding of genetic and genomic mechanisms that control the transformation of stem cells and progenitor cells into differentiated cell types, with the aim of influencing developmental choices and exploring new alternatives to regeneration and therapy.	Martín Montecino, Miguel Allende	Verónica Palma, Tomás Egaña, Giancarlo de Ferrari, Alvaro Glavic
2	Regulation and comparative biology of the stress response.	To understand how species from diverse taxa are able to cope with an ever-changing environment, from the level of the organism and its changes in gene expression to the long-term adaptations that lead to genome evolution.	Ariel Orellana, Rodrigo Gutiérrez, Mauricio González	Andrea Miyasaka, Juan Francisco Miquel, Silvana Zanlungo, Verónica Cambiazo
3	Mathematical modeling of gene networks in response to signals	To contribute to the development of new bioinformatics and mathematical methodology to integrate heterogeneous omics data into a regulatory or/and a metabolic network and study their dynamical behavior	Alejandro Maass	
4**	Discovery through OMIC analysis	Developing platforms and engaging in sequencing of new genomes and those of strategic interest to the country.	All CGR PRs	All CGR ARs

*Several of the Principal and Associate Researchers are involved in more than one research line but for simplicity they have been assigned to only one of the lines.
**A strategic objective that is the collective responsibility of the Center with participation of all Principal and Associate Investigators (PIs and AIs), and which is related to the three scientific aims.

II. EXECUTIVE SUMMARY

The FONDAP Center for Genome Regulation (CGR), launched one year ago, has already become a recognized center of excellence in the areas of Cell and Molecular Biology, Genomics, Bioinformatics and Mathematical Modeling. This has been achieved through continued production of leading edge scientific papers, organization and participation in scientific conferences, recruitment of new associate investigators in key areas of research, establishment of a worldwide collaborative network, graduate and postgraduate training, course and symposia organization and a strong national presence in landmark projects. Our original plan to become a leading Center in Chilean and Latin American science is being realized and the diversity of research lines being implemented in collaboration with colleagues here and abroad has allowed us to broaden the impact of the CGR to unanticipated areas of inquiry. This enhancement in the scope of our interests has been accompanied by a parallel success in establishing a state of the art platform in Genomics by our investigators together with the National Genomics and Bioinformatics Center, which is working closely with the CGR to materialize the installation of advanced sequencing technologies in Chile. Furthermore, this partnership has allowed us to begin several sequencing projects associated with our objectives and to bring into the CGR new Associate Investigators and collaborators working in human medicine, evolutionary ecology, biomining, nutrition and animal health. This is precisely the vision we had for the CGR within the Chilean scientific context: a Center with the flexibility to explore highly relevant biological questions (even those outside the area of expertise of the principal investigators) by generating synergistic interactions with leaders in those fields. It is our belief that the impact of a Center should be measured by more than its scientific production; rather, it should show meaningful contributions to the scientific community at large and to the country. We can modestly say that this key strategic aim is on its way to fulfillment, even at this early stage in our development as a FONDAP Center.

The key research highlights among the three principal aims, are described in the following sections. While we have made significant advances in these objectives, we have also initiated several projects that are “Center Projects”, collective efforts that have been launched to exhibit the potential that our group has for analysis of large omic data sets that can translate into meaningful biology. Among this year’s projects, one of the most important was to begin the sequencing and annotation of human genomes of representative Chilean nationals. The aim is to generate a framework based on sequence that will facilitate mapping of disease predisposition or causative loci, focusing on ailments that are common in the Chilean people such as gastric cancer. Given the sensitive nature of this project, we first carried out a workshop that dealt with the ethical concerns raised by this type of research with invited speakers from countries that have already gone through this experience. The Chilean Human Genome sequencing project (as we have called it) has already begun and the results of this effort, to be made public during 2012, will have a significant impact on society and it will incorporate our country fully into the era of the human genome; we expect the influence of this type of approach on medicine to follow suit.

Furthermore, we have begun to acquire genomic data on a number of Chilean species that are of interest to ecologists and evolutionary biologists. As part of our second aim (genomics of the stress response), we selected an environment that is extremely hostile to plant, animal and bacterial life: the Chilean northern desert. Among the organisms selected are numerous plant

species of diverse taxa, which are distributed in highly specific fashion along an altitudinal gradient. We have begun sequencing RNA and genomic DNA of several of these plants to determine their relationship to species found in less stressful environments and thus to identify peculiarities in their genomes. Likewise, we are examining genomic sequence of a species of fish that lives in the high altitude salt lakes of the *altiplano*, enduring very high salinity and low oxygen. These fish, of the genus *Orestias*, are undergoing rapid speciation and we have detected the beginnings of this process in two populations of *O. ascotanensis* living in two different salt marshes in the same general area. DNA sequence if these fish will reveal what molecular mechanisms are undergoing changes in response to this extremely harsh environment and we expect to find clues as to which physiological processes make these animals tolerant. Finally, we are also exploring a group of plants that are highly ephemeral: the Chilean desert is characterized by occasional blooms of flowers that take advantage of the little rain that falls in the area in some years. The seeds of these flowering plants are very resilient to drying and can survive in dormancy for several years. It will be of interest to examine the genomes of some of these plants (which we have already collected) and to find out how they are able to sustain this very flexible life cycle. All three of these projects are carried out in collaboration with experts in the biology and natural history of these organisms.

Several other CGR projects related with genomes of species of economic interest have also captured our attention during this year. We were directly involved in the sequencing/annotation of the potato genome which was published this year and, likewise, work is concluding with the Atlantic salmon genome, which will be published next year with several CGR members as authors. We initiated new projects with microorganisms, including the sequencing of the genomes of virulent and non-virulent strains of the salmonid pathogen *Piscirickettsia salmonis* and the genomes of six relevant bioleaching extremophiles. We have also invested efforts in the sequencing and annotation of the *Thompson seedless* strain of table grape.

Besides our scientific endeavours, we attained visibility within the Chilean scientific community with the organization of a seminar series, having both in-house speakers as well as invited guest speakers. The series, termed *Interactomics*, was highly successful and will continue during 2012. We also invited Dr. Nadia Rosenthal, Head of EMBL in Rome, to give the CGR lecture at the Chilean Cell Biology meeting, and we hosted several foreign participants in two CGR courses, one in Plant Systems Biology and the other in bioinformatics for the analysis of metabolomic data. We have also implemented a CGR webpage for dissemination of our work.

During 2011, several distinctions and awards have contributed to increase the prestige of CGR investigators. Dr. Rodrigo Gutiérrez, Principal Investigator of the CGR, has been recently awarded a Howard Hughes Medical Institute position, the only Chilean to be distinguished in such fashion in this competition. Dr. Ariel Orellana has been named to the Superior Council of CONICYT, the main Chilean scientific funding agency. Of the four CGR postdoctoral researchers applying for grants this year, all four were awarded fellowships and grants from FONDECYT.

Finally, our intention of implementing facilities that are independent of the host institutions were materialized as we are occupying a building housing our administrative offices and a bioinformatics cluster, with around 10 people. Together with our efforts in communication and outreach, we are confident that we have achieved a strong identity as a one of the leading Chilean research centers.

III. ADMINISTRATIVE ASPECTS

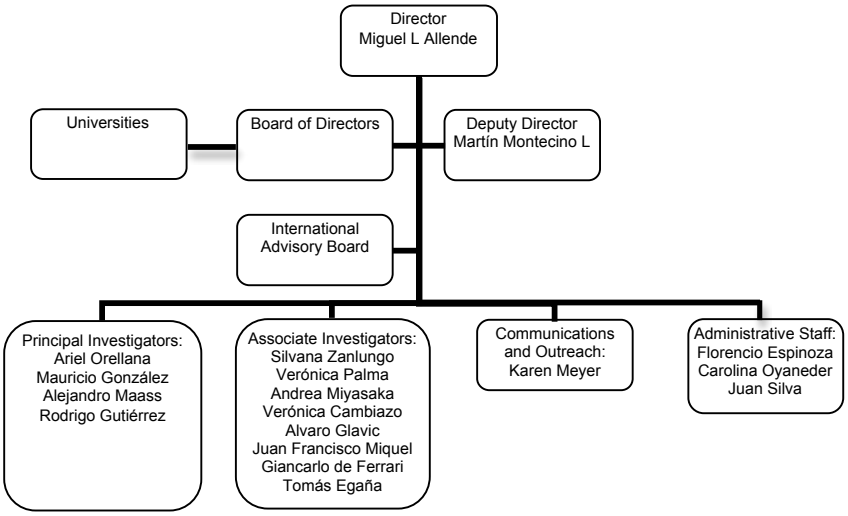
1. **Budget execution**: Describe and justify any budgetary modifications (itemized) of the original proposal.

The original proposal included the purchase of an advanced sequencer so that we could generate our own sequence data. However, given the simultaneous award that the National Genomics and Bioinformatics platform obtained from the Chilean government (Large Equipment Program of CONICYT), which included this type of technology, we decided to invest these funds in the sequencing projects themselves. This allowed us to focus on expanding our original projects and to add new ones, such as the Chilean Human Genome project and those related to plants inhabiting the Chilean desert. Other budget changes were minor and, for the most part, involved transferring funds from personnel to operational expenses due to the fact that we were not able to recruit all of the scientists and administrative assistants at the outset of the project as we had originally planned. In turn, some funds were used for equipment that was unanticipated, such as a high capacity computer server, individual computers for the bioinformatics cluster, parts for updating microscope components and other small lab equipment.

2. **Accomplishment of institutional commitments:** describe any difficulty (ies) encountered regarding this aspect.

In general, the relationship with the principal and associated institutions has been without difficulties. It is neoteworthy that this project functions within three different universities, with project funds alloted proportionally to each one, a situation that required tight coordination of expenses and budget management. Fortunately, we have administrators looking after this aspect in each of the three institutions and we were able to keep track of the progress of the spending and account rendering. The principal institution provided information in a timely manner, supported the project with critical administrative help and was very helpful with all matters regarding the efficient use of the funds. All three institutions were also very forthcoming with the information required for the economic reports and have complied with their obligation to provide an economic leverage to the project.

3. Organizational Chart: A chart showing the internal organization of the CGR



A chart showing principal institutions interacting with the CGR

4. **Personnel:** Provide a table indicating all personnel involved in the operation of the Center during the reported period, including names, position within the center (e.g. associate researcher, post doc, student, technician, etc.) and the number of hours committed to the Center.
   In addition, in no more than one page, provide a brief academic biography for each new researcher recruited by the Center.

Table indicating position and hourly commitment of all CGR personell (scientific and administrative) regardless of their funding source. A commitment of 44 hours refers to the weekly commitment during the period in which they belonged to the Center.

Name	**Position at CGR**	**Commitment (hours)**
Miguel L Allende	Director	44
Martín Montecino	Deputy Director	30
Ariel Orellana	Principal Investigator	26
Alejandro Maass	Principal Investigator	26
Rodrigo Gutiérrez	Principal Investigator	26
Mauricio González	Principal Investigator	26
Silvana Zanlungo	Associate Investigator	15
Verónica Palma	Associate Investigator	15
Andrea Miyasaka	Associate Investigator	15
Verónica Cambiazo	Associate Investigator	15
Alvaro Glavic	Associate Investigator	15
Juan Francisco Miquel	Associate Investigator	15
Tomás Egaña	Associate Investigator	15
Rosario Villegas	Post doc	44
María Laura Ceci	Post doc	44
Alejandro Zuñiga	Post doc	44
Christian Hödar	Post doc	44
Ricardo Nilo	Post doc	44
Paula Vizoso	Post doc	44
María Francisca Blanco	Post doc	44
Elena Vidal Olate	Post doc	44
Diana Grass	Post doc	44
Andrea Vega	Post doc	44
Ricardo Matute	Post doc	44
Adriana Batias	Post doc	44
Henriett Pál-Gábor	Post doc	44
Luis Milla	Post doc	44
Catalina Prieto	Post doc	44
Hoong Soonjo	Post doc	44
José Aliste-Prieta	Post doc	44
Mahsa Allahbakhshi	Post doc	44
Rodrigo Assar	Post doc	44

Leonardo Valdivia	Post doc	44
Anna K Reckhenrich	Post doc	44
Marcela Low	Post doc	44
Angelica Reyes	Post doc	44
Talia del Pozo	Post doc	44
Calixto Dominguez	PhD student	44
Adrián Moreno	PhD student	44
Susana González	PhD student	44
Ignacio Moreno	PhD student	44
Maribel Donoso	PhD student	44
Claudia Muñoz	PhD student	44
Cecilia Rodriguez	PhD student	44
Daniela Doñas	PhD student	44
Eleodoro Riveras	PhD student	44
Pamela Naulin	PhD student	44
José Miguel Álvarez	PhD student	44
Tatiana Kraiser	PhD student	44
Viviana Araus	PhD student	44
Tomas Moyano	PhD student	44
Tomas Puelma	PhD student	44
Orlando Contreras	PhD student	44
Rosario Villegas	PhD student	44
Leonardo Valdivia	PhD student	44
Cristian Undurraga	PhD student	44
Mario Sánchez	PhD student	44
Constanza Martinez	PhD student	44
Joao Botelho	PhD student	44
Gabriela Zavala	PhD student	44
Sebastián Donoso	PhD student	44
Alexander Frank	PhD student	44
Maiximilinao Velardez	PhD student	44
Andrés Aravena	PhD student	44
Diego Rojas Benitez	PhD student	44
Consuelo Ibar	PhD student	44
Jorge Zuñiga	PhD student	44
Marcelo Alarcon Lozano	PhD student	44
Miguel Avila Rivas	PhD student	44
Matias Medina	PhD student	44
Bernabe Bustos B.	PhD student	44
Eduardo Perez	PhD student	44
Giorgia Ugarte	PhD student	44
Emilio Díaz	PhD student	44
Adriana Rojas	PhD student	44
Hugo Sepulveda	PhD student	44

Rodrigo Aguilar	PhD student	44
Graciela Argüello	PhD student	44
Mauricio Latorre	PhD student	44
Leonardo Pavéz	PhD student	44
Henry Temple	Master's student	44
Omar Sandoval	Master's student	44
Gonzalo Cisternas	Master's student	44
José Pablo González	Master's student	44
Macarena Greve	Master's student	44
Oscar Peña	Master's student	44
Javiera de la Paz	Master's student	44
Camila Mardones	Master's student	44
José Moya	Master's student	44
Marjorie Alvarez	Master's student	44
Pablo Lois	Master's student	44
Claudio Cortés	Master's student	44
Carolina Ortiz	Master's student	44
Guillermo Rodríguez	Master's student	44
Fernando Abarca	Master's student	44
Jaime Espina	Master's student	44
Flavia Roman	Master's student	44
Pablo Leon Medina	Master's student	44
Francisco Altimiras	Undergraduate student	44
Leandro Farías	Undergraduate student	44
Hernán Salinas	Undergraduate student	44
Jonathan Peralta	Undergraduate student	44
Cristian Farías	Undergraduate student	44
Bernardo Pollak	Undergraduate student	44
Oscar Peña	Undergraduate student	44
Marjorie Alvarez	Undergraduate student	44
Simón Carrillo	Undergraduate student	44
María del Sol	Undergraduate student	44
Natalie Espinoza	Undergraduate student	44
Samuel Martínez	Undergraduate student	44
Andrea Arros	Undergraduate student	44
Sebastián Donoso	Undergraduate student	44
Italo Cipriano	Undergraduate student	44
Angel Pardo	Undergraduate student	44
Vicente Cataldo	Undergraduate student	44
Paulina Salazar	Undergraduate student	44
Tatiana Opazo	Undergraduate student	44
Daniel Meza	Undergraduate student	44
Paulina Rudolfi	Undergraduate student	44
Nicole Reynaert	Undergraduate student	44

Florencio Espinoza M	Administrator, accounting	20
Karen Meyer B	Journalist, communications	44
Mónica Alvarado	Secretary	44
Carolina Oyaneder	Secretary	44
Juan Silva	Janitor and messaging	10

In the following paragraphs, we present short Biographical sketches of Associate Investigators incorporated in 2011. Full CVs can be found in the Appendix.

Dr. Alvaro Glavic is a Biochemist graduated from the University of Chile in 1997 and he obtained his PhD in Cell and Molecular Biology and Neuroscience from the same institution in 2002. He did postdoctoral work at the Centro de Biología Molecular "Severo Ochoa", Madrid, Spain, with Dr. Antonio Garcia-Bellido. He returned to Chile in 2005 and spent one year at the Faculty of Medicine, university of Chile before obtaining his position as Assistant Porfessor at the Department of Biology, Faculty of Sciences, in 2006. He was a junior investigator of the Millennium Nucleus in Developmental Biology (2006-2007) and senior investigator of the Center for Genomics of the Cell (2007-2010). He has obtained several merit awards including the best thesis in biology prize, the Hermann Niemeyer medal and had the prestigious Fundación Andes fellowship for doctoral studies. He is a member of the Graduate Committee for PhD program in Cell and Molecular Biology and Neuroscience at the Faculty of Sciences and has participateds in many national and international courses, both at the graduate and undergraduate levels.

Dr Tomás Egaña graduated as an Engineer in Molecular Biotechnology at the University of Chile in 2003. Afterward he obtained a PhD in Human Biology at the University of Lübeck, Germany (2008), and a PhD in Pharmacology at the University of Chile (2009). In 2008 he joined the Technical University of Munich as postdoctoral fellow in the Department of Plastic and Hand Surgery at the Faculty of Medicine. Since 2010 he is Group Leader of the Laboratory of Tissue Engineering and Regeneration at the same institution. During this time his research has been focused in the development of new strategies to improve tissue regeneration.

Dr. Juan Francisco Miquel (MD) studied Medicine at the Universidad Austral de Chile, specializing in gastroenterology at the Pontificia Universidad Católica de Chile in 1991. He was awarded a Presidential Scholarship and an Alexander von Humboldt (Germany) Fellowship, for a 2-year stay at the Ludwig-Maximilian Universität in München, Germany, under the direction of Prof. Dr. Gustav Paumgartner (1992-1993). He returned to the Universidad Austral de Chile as an academic and a year later moved to the Faculty of Medicine at the Pontificia Universidad Católica de Chile, where he remains to this day as a

Full Professor at the Dept. of Gastroenterology. He teaches at the Medical School as well as the Faculty of Biological Sciences. He has obtained regular grants from FONDECYT since 1994 as a principal investigator. His research has focused on digestive system diseases that are prevalent among the Chilean population, especially vesicular lestasis and vesicular cancer, studying ranging from basic pathogenic mechanisms to environmental and genetic risk factors. During the past few years, he has concentrated on seeking which genetic determinants are responsible for heritable metabolic diseases among Chileans, especially those linked to indigenous populations. He maintains active collaboratrions with groups in Germany and Holland, forming part of an interdisciplinary group of investigators researching the causes of colelitiasis. He has contributed in more than 70 international publications, many in high impact journals.

Dr. Verónica Cambiazo received her PhD from the University of Chile in 1998, working under the direction of Dr. Ricardo Maccioni on the role of a new regulator of microtubule cytoskeleton during *Drosophila* embryogenesis. During 1998-1999 she was a Marie Curie postdoctoral fellow at Laboratorie de Hormones et Cell Differentiation INSERM, France. She was appointed Assistant Professor at INTA-University of Chile in 1999 and Associate Professor in 2003. In 2003, together with her colleague, Dr. Mauricio González, they created the laboratory of Bioinformatics and Gene Expression (LBGE) at INTA with the aim of establishing a research program on fundamental and applied aspects of gene expression regulation at the cellular and organismal levels. Since the creation of the LBGE, she have mentored 6 graduate and 10 undergraduate students. Dr. Cambiazo's research group has been investigating the differential gene expression patterns that underlie early stages of *D. melanogaster* embryogenesis. By combining functional genomics strategies, analysis of temporal and spatial patterns of gene expression and gene perturbation assays, they have been able to identify essential genes that have escaped detection in standard genetic screens and have discovered a new set of signaling molecules that are developmentally relevant. In addition, the Cambazo group has contributed with their expertise in genomics and bioinformatics to confront problems that affect the Chilean productive sector. In collaboration with the biotech company Aqua INNOVO (partially owned by U. of Chile) they are seeking solutions to problems such as the intolerance of Atlantic salmon to feed containing a high proportion of vegetable oils and proteins and the identification of Atlantic salmon families that are more resistant to an infection with the pathogen *P. salmonis*.

5. **Changes in research personnel:** Describe any changes in the principal and associate researchers relative to the original project.

During 2011, we incorporated four Associate Investigators that had not been identified in the original proposal: Dr. Alvaro Glavic, Dr. Juan Francisco Miquel, Dr. Verónica Cambiazo and Dr. Tomás Egaña. Their short biographies were provided in the preceding section. All of the other Investigators were included in the application. One potential Associate Researcher named in the application did not end up joining the CGR (Dr. Andrew Hart).

We must note that we decided to carry out a rather unorthodox incorporation, as Dr. Egaña is a young Chilean researcher working in Germany, at the Technical University of Munich. Dr. Egaña spends two months each year at the CGR, actively collaborating with Drs. Allende and Palma. We will consider the incorporation of other international Associate Investigators as they enrich the Center with important contributions and the possibility of hosting our students in their labs, as has already happened in Dr. Egaña's case.

6. **Advisory committee:** describe its tasks, the frequency of meetings, and usefulness of the advice provided to the Center. Also, report on the availability of the committee to assist the Center.

We have two advisory committees:
The Board of Directors is composed of the Vicechancellors for Research of the three participating universities, in addition to two National Science Prize winners, Dr. Cecilia Hidalgo and Dr. Pablo Valenzuela, and the six CGR Principal Investigators. We had one organizational and one social meeting this year. The contribution of this group is fundamental for designing the strategy that the CGR will follow in its relationship with the host institutions and society. They have stressed the importance of maintaining a high scientific standard without abandoning key activities such as undergraduate and graduate teaching, collaboration with other centers and colleagues, and outreach to society, especially with school children. The Board has also discussed how to achieve a definitive infrastructure that would physically bring together the CGR researchers, in a setting that could involve a joint venture between the three universities and the government. We are exploring this option and hope to have some advances in 2012.

The International Advisory Board is composed of three prominent scientists: Dr. Joseph Ecker (Director, Salk Institute Genomic Analysis Laboratory, member of the US Academy of Sciences, President International Society for Plant Molecular Biology), Dr. Brandon Wainwright (Director of the Institute for Molecular Bioscience at the University of Queensland, Brisbane, Australia), and Dr. Bernard Prum (Director of the Statistics and Genome Group, Genopole-Paris, France). While we weren't able to coordinate a date in which they could all be here in Chile to meet with us, they have agreed to come in 2012.

IV. OBJECTIVES AND RESULTS ATTAINED

1. RESULTS OBTAINED RELATIVE TO CENTER OBJECTIVES

i. Main research findings

Aim 1. Epigenetic control of differentiation and stem cell biology.

1.1. Epigenetics of cell differentiation. During this first year, we have assessed epigenetic mechanisms regulating gene transcription during cell differentiation in a number of meaningful biological systems and focusing on specific key regulatory genes that are critical during cell-tissue lineage commitment. Among the principal biological systems utilized we find the osteoblast and neuronal differentiation processes. Additionally, as an extension of this work, we have collaboratively addressed epigenetic control of genes coding for membrane receptors and ion channels that are critical components during the pain response in mammals. Our main findings for this first year of work are summarized as follows.

Runx2 and Osterix transcription factors are essential for bone development and osteoblast differentiation as they control the expression of a significant number of bone-related target genes. Runx2 is considered the master regulator in osseous cells as is also required for Osterix expression during early bone development. Through the work of two Ph.D. thesis students in the Montecino lab (Adriana Rojas and Hugo Sepúlveda) we have investigated epigenetic mechanisms that control transcription of Runx2 and Osterix genes during commitment of mesenchymal and osteoprogenitor cells to the osteoblast phenotype as well as during differentiation of mesenchymal cells to non-osteoblastic lineages. We had previously determined that transcription of the Runx2 gene involves a SWI/SNF-independent chromatin remodeling process, associated with a specific pattern of post-translational modifications in histones H3 and H4 at the P1 proximal promoter region of the Runx2 gene, which controls transcription of the Runx2/p57 isoform in osteoblasts (Cruzat et al. 2009). The contribution of C/EBPb to transcriptional up-regulation of the Runx2 gene in pre-osteoblastic mesenchymal cells as well as in cells committed to the osteoblast lineage was established in a manuscript published this year in the Journal of Cellular Physiology (Henriquez et al. 2011). The pattern of histone modifications at the Runx2 gene promoter includes elevated histone H3 acetylation (H3ac), in particular H3K9ac and H3K14ac, as well as increased histone H3 tri-methylation on lysine 4 (H3K4met3) and reduced H3K27met3. Interestingly, we also find that the post-translational modification pattern changes during differentiation to non-bone cells (e. g. neuronal cells) observing enrichment of H3R3met2, H3K9met3, H3K27met3 and H3K4met1 and reduction in H3K4met3 and H3K9ac. Similarly, total H3ac and H4ac levels are significantly reduced at the P1 promoter in young and adult mouse brain tissue (thesis work of the Ph.D. candidate Rodrigo Aguilar). To begin defining the enzymes responsible for this specific epigenetic pattern and their contribution to Runx2 transcription during osteoblast lineage commitment, we have addressed the expression patterns exhibited by the lysine methylases WDR5 and EZH2 as well as the lysine demethylases NO66 and UTX, in C2C12 mesenchymal cells stimulated to engage an osteoblastic differentiation process by incubation with BMP2 (Doctoral thesis work of Adriana Rojas, co-directed by Drs. Allende and Montecino). Our results indicate that these enzymes can contribute significantly to Runx2 transcription activation by establishing an epigenetic environment at the P1 promoter that facilitates binding of the transcription machinery. Similar analyses indicate that the Osterix gene promoter exhibits reduced H3ac and decreased H3K4met3, together with elevated levels of promoter bound Histone Deacetylases (HDACs) and reduced association of RNA polymerase II and Runx2, when the gene is not expressed in early mesenchymal phenotypes (thesis work of the Ph.D. candidate Hugo Sepulveda). Together, our results support a model where epigenetic post-translational modifications set the key Runx2 and Osterix gene promoters in early mesenchymal cells for a rapid and efficient transcription upon activation of the osteoblast differentiation program. During the second year, we will continue these studies but now analyzing the presence and contribution of these epigenetic patterns on cell-lineage commitment with a more global approach that is,

combining chromatin immunoprecipitation and deep sequencing strategies. Initial analyses using this approach have been carried out through a collaborative effort between the group of Dr. Montecino and the research team led by Drs. Andre van Wijnen and Gary Stein at University of Massachusetts Medical School, USA. In these studies, the total number of target genes directly recognized by the transcription factor Runx2 in osteoblastic cells and epigenetic mechanisms controlling cell-cycle regulated genes in Human Stem Cell lines have been recently established (van der Deen et al. Epub 2011; Medina et al. Submitted).
During this first year we have also assessed epigenetic mechanisms regulating cease of proliferation of osteoprogenitors and their engagement to the osteoblast differentiated state. We have focused part of this analysis on the Ric-8B gene, which is essential for cell proliferation during early embryogenesis of several model systems (e. g. Drosophila, Xenopus and Mouse) as it controls symmetric/asymmetric cell division. Our results support a model where C/EBPβ-LAP* recruits SWI/SNF to the Ric-8B promoter leading to nucleosome mobilization and Ric-8B gene repression due to dissociation of the transcriptional machinery, during osteoblast differentiation. These results are the basis of a recently published article in Molecular and Cellular Biology (Grandy et al. 2011), one of the most important journals in the field of transcriptional regulation. Importantly, our results establish for the first time, that a single transcription factor, C/EBPb can epigenetically promote osteoblast differentiation by both activating (e.g. Runx2, Henriquez et al. 2011) and repressing (e. g. Ric-8B, Grandy et al. 2011) critical genes that control osteoblast differentiation and osteoprogenitor proliferation, respectively. During the second year we will extend these studies to establish the specific mechanism that allows C/EBPb to simultaneously activate and repress target genes during cell differentiation, defining whether specific signaling pathways, including cAMP/PKA and Wnt, contribute to these mechanisms.
Previous results from a collaborative effort between the laboratory of Dr. Montecino and the group of Dr. Brigitte van Zundert (CIB-UNAB) indicated that increased expression of the NMDA receptor anchoring protein PSD95 is primordial in the restricted dendritogenesis observed in the adult Central Nervous System. Therefore during this first year, we have collaboratively analyzed (Doctoral thesis work of Fernando Sepúlveda, co-directed by Drs. van Zundert and Montecino), whether epigenetic mechanisms involving specific covalent histone tail modifications are associated with PSD95 promoter activity in hippocampal tissue during rat development. These results will be included in a manuscript that is currently prepared for publication by Drs. van Zundert and Montecino. Another collaborative project led by the laboratories of Dr. Montecino and Dr. van Zundert, has been focused on establishing whether specific epigenetic mechanisms are regulating the level of response of pain-sensitive neurons (Nociceptors). Our results support a functional and coordinated role for Runx1 and C/EBPβ transcription factors during activation of P2X3 gene transcription and they have been included in a manuscript recently published in Journal of Cellular Physiology (Ugarte et al. 2011) and allowed a Ph.D. student (Giorgia Ugarte) to graduate during 2011. During next year we will extend these studies to determine, using now animal models (Mouse), whether the response to neuropathic pain in vivo is related to Runx1-C/EBPb mediated epigenetic changes at target genes.

1.2. Stem cells in the development, differentiation and regeneration of sensory cells and neurons. A different model that we use for the analysis of stem cell function is the zebrafish embryo and larva. A population of progenitor cells that is highly convenient for genetic ad molecular analysis in this animal is the one responsible for the generation of mechanosensory hair cells, such as those in the inner ear and the lateral line system. We have analyzed the transcritpome of these cells (Gallardo et al., 2010) and discovered a group of genes that control proliferation, migration and the immune response, both during normal developmental events and during repair of damaged tissue. A gene of particular interest is the Lef1 transcription factor, a Wnt pathway effector, which showed a novel role in controlling proliferation of migrating progenitor cells that was key for maintaining tissue homeostasis (Valdivia et al., 2011). Besides

examining how stem cells operate in normal development, we have also developed strategies that allow us to examine regeneration of these cells, which in fish occur robustly and in a matter of hours. Damaging existing sensory cells with heavy metals such as copper (Hernández et al., 2011) induces their rapid replacement with newly born cells that arise from proliferating progenitors lying within the organ. Investigating how these cells are stimulated to respond to damage, we discovered an unexpected inflammatory response that appears to be needed for debris clearance and stimulation of regeneration (d'Alençon et al., 2010). Innate immune cells rapidly infiltrate the damaged sensory organ and express a number of proteins that suggest tissue remodelling, including cytokines and extracellular matrix proteases. These findings opened a new line of research in the Allende lab that involves examining the role of imunity in stem cell activation under a paradigm of neural cell damage. We have now extended these studies towards neurons and we have performed work on axonal regeneration and its dependance on immune cells and other cellular and humoral factors (Villegas et al., 2012; submitted). Again, we plan on using transcriptomics to identify genes that are activated during the regenerative response in neurons and to try and link these transcriptional events to epigenetic changes in their promoters.

1.3. Role of the Sonic hedgehog pathway in neural progenitor cell proliferation. Previous studies of Shh signaling in the mesencephalon have uncovered a role for the pathway during early embryogenesis in inducing ventral cell types, in addition to a more general role in proliferation and cell survival. This, and the finding that Shh regulates the behavior of cells with stem cell properties in the developing embryonic neocortex led us to investigate its possible involvement in the control of stem cell lineages in the dorsal midbrain (optic tectum) at mid and late embryonic stages using a comparative approach. Importantly, our findings choosing zebrafish (Feijóo et al., *EJN* 2011), chicken (Rapaciolli et al, submitted *BMC Neuroscience*) and murine (Martínez et al., submitted *MOD*) models suggest that the mitogenic activity of Shh is likely conserved from teleosts to mammals. Targets of the Shh pathway were investigated using transcriptomic approaches (Milla et al., 2008). Some of the candidate genes from our genome-wide screens performed through a collaboration between the labs of Drs. Palma, Cambiazo and Allende, such as *neo1* and *cmyc* (Milla et al., under review in *BMC Genomics*) are being verified functionally and have shown unanticipated relationships between the Shh pathway and cancer. Medulloblastoma are very aggressive tumors of the cerebellum that most commonly occur within the first few years of life. The elucidation of signaling pathways involved in medulloblastoma pathogenesis could substantially improve the clinical management of these neoplasms, a more accurate prediction of the disease risk could be achieved and new-targeted treatments developed. Current experiments, done in collaboration with Dr. Brandon Wainwright (University of Queensland, Australia) are exploring the potential function *neo1* in the genesis of medulloblastoma using conditional mouse models. Furthermore, we have found that the Shh pathway modulates maturation of chondrocytes *in vitro* through *neo1*. Recent evidence points for a role of *neo1* in chondrogenesis but its potential function, under hedgehog control, has not been explored. For this purpose, we have established a primary culture of chondrocytes obtained from the ribs of chicken embryos that will provide the biological material necessary to perform microarrays, RNA-seq and ChIP-seq, in a collaboration between the labs of Drs. Palma and Montecino. Because osteogenesis has principally been studied in mammals, our work in birds will be crucial to unravel the evolution of the skeletal cell-specific regulatory network in tetrapod vertebrates.

1.4. Tissue regeneration through stem cell and biomaterial technologies. The potential use of the body's own capacity to recover tissue functions after injury is the current challenge of regenerative medicine. Isolated cells cannot restore their respective tissue architecture because they do not have a scaffold to guide their growth. The development of such scaffolds is one of the major topics in the field of tissue engineering and biomaterials. Although the use of scaffolds is promising, clinical results have been disappointing and the

lack of vascularisation is one of the main problems in scaffold dependent tissue regeneration. In the absence of blood vessels oxygen, nutrients and immune cells cannot reach the wounded area. As a consequence, low regeneration and high infection rates are among the most common problems in tissue engineering. To enhance the tissue regeneration process, the group of Dr. Egaña is combining the use of scaffolds and therapeutic vascularisation strategies. Regeneration models have been established in the mouse (in collaboration with Dr. Palma) and the zebrafish (in collaboration with Dr. Allende). In addition, his group is exploring other key areas such as: (i) the use of stem cells to improve tissue regeneration; (ii) development of new scaffolds for tissue engineering; (iii) Use of gene-activated biomaterials for tissue engineering and regeneration; (iv) the role of aging in tissue regeneration.

1.5. Control of proliferation in developing systems. Another line of research aims to understand the role of the CSRNP (cystein-serine rich nuclear proteins) in progenitor cell expansion and survival. This is a collaborative effort between Drs. Glavic and Allende. Previous work showed that in *Drosophila* and in Zebrafish a member of the CSRNP gene family (*DAxud1* and *csrnp1,* respectively) regulates the proliferation of progenitors in larval tissues as well as in the CNS (Glavic et al., 2009; Feijoo et al., 2009). Ongoing research has extended these observations showing that a zebrafish paralog (*csrnp1a*) controls the expansion and survival of primitive hematopoietic progenitors. To acquire a deeper understanding on the function of this group of proteins and its conservation, our plan for the next few years is to undertake a high throughput analysis using *D.melanogaster* and zebrafish to define the genes controlled by these proteins in order to get insight on the common mechanisms used by different progenitors to preserve their condition.

1.6 Target gene discovery through transcriptome perturbations and enhancer predictions in *Drosophila melanogaster*. In *D. melanogaster*, dorsal patterning is controlled by an extracellular gradient of the morphogen Decapentlaplegic (Dpp). The transcriptional effectors of this pathway, Mad/Medea, exert their regulation in a graded-manner eliciting at least three threshold responses: high, intermediate and low. However, the mechanism underlying the differential response to Dpp is poorly understood, due in part to the insufficient number of well-studied target genes. In ongoing work in the Cambiazo lab, we combined genetic transcriptome perturbations and computational analyses to identify a set of target genes of transcription factors Mad/Medea, which includes known dorsal patterning genes and several novel genes. We have validated many of them by *in situ* hybridizations on wild type and Dpp signaling pathway mutants. The validated set of novel target genes exhibited functional enrichment of signaling molecules. *In vivo* validations of the predicted enhancers revealed that co-regulated enhancers that are activated by intermediate levels of Dpp contain similar arrangements of Mad/Medea binding site clusters. Our results contribute to decipher the underlying *cis*-regulatory code controlling dorsal ectoderm patterning.

Aim 2. Regulation and comparative biology of the stress response.

2.1. TGA1 and TGA4 transcription factors control nitrogen responses in *Arabidopsis thaliana* roots. Nitrogen (N) and its metabolites regulate plant growth and development and act as potent signals to control gene expression in *Arabidopsis*. Using an integrative bioinformatics approach the Gutiérrez group identified TGA1 and TGA4 as putative regulatory factors that mediate N responses in *A. thaliana* roots. We showed that both *TGA1* and *TGA4* mRNAs accumulate strongly and quickly after nitrate and nitrite treatments in root organs. Phenotypic analysis of *tga1* and *tga4* double mutant plants indicated that TGA1 and TGA4 are necessary for both primary and lateral root growth in a nitrate dependent manner. Global gene expression analyses revealed that 97% of the genes with altered expression in the *tga1/tga4* double mutants are regulated by nitrate treatments indicating these transcription factors have a specific role in nitrate responses in *Arabidopsis* roots. Among the nitrate-responsive genes that depend on TGA1/TGA4 for normal regulation of

gene expression, we found the nitrate transporters *NRT2.1*, *NRT2.2* and the nitrite reductase (*NIR*) genes. Specific binding of TGA1 to its cognate DNA sequence on the target gene promoters was confirmed by chromatin immunoprecipitation assays. These results identify TGA1 and TGA4 as important regulatory factors of the nitrate and nitrite response in *Arabidopsis* roots.

2.2. Gene network of *Enterococcus faecalis* associated to iron and copper homeostasis. During the first year, our aim was to use a system-level approach to integrate data from microarray experiments into a regulatory network that describes copper and iron homeostasis. We defined a "Gene regulatory network" as a set of links that connect transcription factors with their target genes. Therefore, a "gene regulatory network" controls the synthesis of transcripts and their encoded proteins in response to fluctuations in copper and iron availability. In the lab of Dr. González, we are using the bacterium *E. faecalis* as a model organism. In this bacterium copper metabolism depends on the activities of proteins encoded by the cop operon, they are the transcriptional regulator CopY, the Cu chaperone CopZ and the Cu-efflux transporter CopA. The cop operon appears to be conserved in all members of the *Lactobacillales* order. An additional component, CutC, was recently described by us and is likely involved in the efflux of copper (Latorre et al., 2011). Regarding iron metabolism, *E. faecalis* contains five operons encoding proteins mainly associated with the uptake of iron. In our microarray experiments we compared the gene expression changes between *E. faecalis* grown in control medium and metal exposed cells, under the conditions previously defined. The results indicated that a number of genes were up- or down-regulated in response to copper or iron. The Venn diagram (Figure 1, see Annex) shows that most of the genes that were differentially expressed in the presence of one of the metals were specific for that metal and only a few genes responded to the presence of both metals. In order to connect the predicted TFs with their target genes, we recovered the information of TF binding sites already reported in the literature and in different public databases. The list of TFs with known binding sites was compared with the list of TFs of *E. faecalis* using bidirectional best blast analysis; this allowed us to recover 61 orthologs TFs in *E. faecalis*. The binding sites for those TFs with orthologs in the *E. faecalis* genome were used to generate apositional weight matrix (PWM), which then was used to search for allpromoter regions of *E. faecalis* using appropriate software. With this information we were able to reconstruct a network in *E. faecalis* that consists of 648 operons, 23 TF families with 1227 potential interactions. This network represents a first hypothesis of regulatory interactions in *E. faecalis*. The gene regulatory network activated by copper and iron is shown in Figure 2. To verify the target genes for some of the predicted TFs, we started by analyzing the target genes that were regulated only by Fur or CopY. With this aim we generated mutant strains and compared the expression of target genes between wild type and the mutant. The quantitative PCR results of comparing WT and mutant strains revealed that only 3 operons are up-regulated in the absence of the repressor Fur, while only one is activated by the absence of CopY. This analysis allowed us to refine the Fur and CopY binding sites in *E. faecalis* and this information will be incorporated in our protocols to search for new binding sites (Figure 3). With these results we modified the network, in this new network Fur was separated from the other member of the Fur family, Zur and Per, for which we are also generating mutant strains. CopY, on the other hand, was separated from the network, since only cop operon was under the regulation of this TF. As a consequence of this observation, we proposed that a different set of TFs might mediate the transcriptional changes induced by copper. This work provides for the first time, information about the structure and conservation of transcriptional regulatory networks within a bacterial order.

2.3. Transcriptomic analyses of peaches during postharvest. The Chilean peach industry exports most of the country's fruit production, which requires cold storage for long distance shipments. Unfortunately, this procedure leads to a disorder known as chilling injury. Different varieties are affected in different manners. In addition, different postharvest treatments such as controlled atmosphere, temperature conditioning, and

inhibitors of ethylene have been utilized to decrease the damage produced by chilling injury, which adversely affects the peach industry. In order to identify genes that may play a role in this process we have analyzed four varieties that show different susceptibility to chilling injury. They were, in addition, exposed to different postharvest treatments and the extent of the damage evaluated. Using RNA Seq analyses, the lab of Dr. Orellana in collaboration with Dr. Reinaldo Campos, has found differences in the susceptibility to chilling injury among varieties and exposure to postharvest treatments were found. We are aiming to identify genes that are associated to chilling injury and genes that may protect fruit from this disorder.

2.4. Biotic and Abiotic stresses are related to the activation of transcription factors associated to stress in the endoplasmic reticulum. bZIP60 is a transcription factor from *Arabidopsis thaliana* that suffers an unconventional splicing mediated by IRE-1, a signaling molecule involved in the unfolding protein response (UPR). Dr. Orellana has found evidence that indicates that splicing of bZIP60 is required for the plants' response to pathogen attack. Interestingly, preliminary results suggest that the unspliced form of bZIP60 is also involved in the response to salt stress. Thus, our results suggest that bZIP60 is a transcription factor that acts in biotic or abiotic stress depending on its splicing.

2.5. Genetic engineering in *Prunus avium*. Genetic improvement of the sweet cherry (*Prunus avium L.*) by conventional breeding is a slow and difficult process because of the long generation time and heterozygous nature of this fruit crop. Genetic engineering techniques may facilitate the rapid transfer of valuable genes to well-established commercial *Prunus* cultivars, thus resulting in more efficient genetic improvement. Dr. Miyasaka is working on the generation of new varieties (i) of very early and very late production to extend the harvest season, and (ii) of low chilling requirement, to extend the area of cultivation north of Latitude 33°. For this aim we are studying bud dormancy and flowering gene regulation in the sweet cherry to select and clone candidate genes to be used as molecular markers to support conventional breeding and also to generate transgenic cherry trees.

2.6. Mechanisms governing cell stress in animal cells. The unfolded protein response (UPR) is a stress response evoked by the accumulation of misfolded proteins at the ER. Bax inhibitor-1 (BI1) is a conserved transmembrane protein belonging to the Lifeguard (Lfg) gene family known to modulate this response interacting with the Ire1a transducer at the ER membrane (Lisbona et al., 2009). In this period Dr. Glavic, working with the group of Dr. Claudio Hetz, ICBM, has shown that in addition to this function, BI1 control through the JNK pathway autophagy. In this way BI1 regulates a secondary process that helps to overcome the cellular stress generated by misfolded proteins (Castillo et al., 2011). Additionally we have started the functional characterization, using *D. melanogaster*, of TMBIM1/RECS1, another Lfg family member, as part of the stress response. Our results indicate that RECS1 protein is distributed in a late endosome-lysosomal compartment. Its expression is induced by the UPR and has proapoptotic activity, suggesting that RECS1 could work as a pivotal factor in the resolution of the UPR or apoptosis induction under stress conditions.

2.7. Molecular and Cellular Pathogenic Mechanisms in Niemann Pick type C Disease. Niemann Pick type C (NPC) disease is an inherited, fatal, neurodegenerative and hepatic disease. Although certain aspects of the disease are well known, such as the associated genetic alterations in the *NPC1* and *NPC2* genes, the accumulation of cholesterol in lysosomes of NPC cells, cell death of neurons and hepatocytes, as well as patient symptomatology, the cellular and molecular disease mechanisms associated with cell damage and death have been less explored. The general goal being pursued in Dr. Zanlungo's lab, is to determine whether the increase in oxidative damage contributes to the pathogenesis of Niemann Pick type C disease. Our recent results show that the levels of oxidative damage markers are elevated in livers and cerebella of the NPC

mouse model. Furthermore, we have shown, in a collaboration between the groups of Dr. Zanlungo and Dr. González, that copper transport is also disrupted in NPC disease (Vázquez et al., in press). As possible inducers of oxidative stress, we propose mitochondrial dysfunction, decreased antioxidant defenses and accumulation of cytotoxic oxysterols. Indeed, we have recently described alterations in vitamin E (α-tocopherol, α-TOH), the most potent endogenous antioxidant, in liver tissue and neurons from NPC mice. We found α-TOH buildup in the endosomal/lysosomal system that may result in a decreased bioavailability and impaired antioxidant function of vitamin E in NPC, contributing to the disease (Yévenes et al., in press). In addition, we are exploring the systemic consequences of Cu transport alterations in the NPC mice model analysing the consequences of a low-copper diet on cardiac hypertrophy. This research line is part of the PhD thesis of Graciela Argüello.
A related line of research carried out in Dr. Zanlungo's lab, involves the Relevance of Niemann Pick C2 protein in Cholesterol Gallstone Disease. Niemann-Pick C2 protein (NPC2) is a lysosomal soluble protein that is strongly expressed in the liver; it binds to cholesterol and is involved in intracellular cholesterol trafficking, allowing the exit of lysosomal cholesterol obtained via the endocytic pathway. Interestingly, this protein is also present in bile in a previously described cholesterol pro-nucleating fraction. Thus, this protein may play an important role in controlling hepatic cholesterol transport and metabolism and cholesterol solubility in bile. The aim of this work is to study the relevance of NPC2 protein expression in hepatic cholesterol metabolism, biliary lipid secretion and gallstone formation. In NPC2 KO mice we found reductions in biliary cholesterol secretion, gallbladder bile cholesterol saturation, and cholesterol crystal and gallstone formation when fed a diet that induces cholesterol gallstone formation (Balboa et al., 2011). The major conclusion of this work is that hepatic NPC2 expression is an important factor in the regulation of diet-derived cholesterol metabolism and disposal as well as in diet-induced cholesterol gallstone formation in mice.

2.8. Identification and characterization of genetic risk factors in neurological diseases. We have been analyzing data generated through genome-wide association studies (GWAS) in prevalent neurological disorders, including Alzheimer's disease (AD), autism and schizophrenia, in order to identify common genetic variants that may help us to understand the etiology of these diseases. During this year, and in particular for Alzheimer's disease, the lab of Dr. De Ferrari has performed a meta-analysis on four GWAS in AD (5,101 cases and 4,516 controls), using the inverse variance method implemented in PLINK and R. We have imputed 2.3 millions of single nucleotide polymorphisms (SNPs) from the HapMap CEU population to maximize information on linkage disequilibrium structure between the studies. We have found a strong genetic association in the PVRL2 gene (rs2075650, $P=1.94x10^{-86}$), surrounding the APOE locus. We likewise identified a novel genome-wide association signal on the MECOM locus (MECOM: MDS1 and EVI1 complex locus), which is located in chromosome 3q26 (rs9809961, $P=4.17x10^{-7}$), and detected marginally significant associations on PRKCQ (rs943451, $P=5.29x10^{-7}$) and MS4A3 (rs474951, $P=4.89x10^{-6}$). We are currently preparing a manuscript to be submitted during the next few months (see Bustos et al, 2011). Likewise the data is being screened in order to detect biological processes, molecular functions or cellular components, which might be overrepresented in AD. The ontological analyses have revealed thus far that genetic variants within synaptic, cytoskeletal and cellular adhesion categories are overrepresented in AD.

Aim 3. Mathematical modeling of gene networks in response to signals.

3.1. Response of the bacterial genome to stressful environmental conditions. During 2011 we pursued our research on biomining solutions in collaboration with the biotechnology company Biosigma S.A. In order to gain insight into the organization of the gene regulatory networks of the biomining bacterium *Acidithiobacillus ferrooxidans* and to provide a framework for further studies in bacterial growth under extreme conditions, the group of Dr. Maass identified 87 *A. ferrooxidans* transcription factors. This study

revealed that the *A. ferroxidans* genome contains several members of the ArsR and MerR families, which are involved in metal resistance and detoxification. Since *A. ferrooxidans* inhabits some of the most metal-rich environments known, the transcription factors identified here seem to be good candidates for functional studies in order to determine their physiological roles and to place them into transcriptional regulatory networks in response to heavy metal exposure (Hodar et al., 2011). Another axis of research was the study of the metabolic capabilities of some relevant biominig bacteria under metal stress and the bioinformatic analysis of a highthroughput metabolomic experiment in biomining. In a first article we apply stoichiometric models for central metabolism and iron oxidizing pathways in *Leptospirillum ferriphilum* and *A. ferroxidans*, in order to compare and understand their distinct metabolic responses (specifically, iron oxidation and energy maintenance). Our models include CO_2 fixation and oxygen reduction by a cytochrome oxidase and uphill and downhill pathways for ATP and NADH synthesis. Predicted flux distributions, when compared to experimental data from literature, showed important differences in iron oxidation stoichiometries of both microorganisms. This led to differential and improved models that take into account the predominance of *L. ferriphilum* over *A. ferrooxidans* at higher redox potential values. This tool is a means to explore and predict the metabolic behavior of these iron-oxidizing microorganisms in complex environments such as industrial bioleaching plants (Padilla et al., 2011). In a second article we developed the first experimentally validated stoichiometric model that is able to assess RISC oxidation quantitatively in *A. thiooxidans* (strain DSM 17318), the archetype of the sulfur oxidizing chemolithoautotroph. This model was based on literature and genomic analysis, considering a widespread combination of formerly proposed RISC oxidation models combined and evaluated experimentally (Bobadilla et al, 2011). Finally, in collaboration with the most important metabolomics center in Japan and Biosigma S.A. we obtained, for the first time, the metabolomic profile of two biomining bacteria involved in industrial bioleaching processes; Wenelen, *A. ferrooxidans* (strain DSM 16786), and Licanantay, *A. thiooxidans* (strain DSM 17318). Samples were taken at various growth phases, the former growing on iron and sulfur, and the latter on sulfur and chalcopyrite. In the case of growth on a solid substrate (sulfur and chalcopyrite), the analysis distinguishes between attached and sessile bacteria. Obtained data indicates an abundance of molecules that provide important insights on bioleaching related functions (Martínez et al., 2011).

In a related line of research, the PhD thesis of Andrés Aravena, co-advised by Dr. Maass and A. Siegel, partner of the Symbiose team from INRIA-Rennes, aims to produce bioinformatic and mathematical methods allowing the management, exploration and integration of large sets of heterogeneous omics data into networks of interactions allowing the production of biomarkers of their behavior. During 2011 we have worked on the reconstruction and validation of regulatory networks from transcriptomic data and genomic sequences. At this stage, we have proposed a method based on information theory techniques and ASP logic programming ideas producing two overlapping graphs. The first one links differentially expressed genes having positive “mutual information” and the second associates transcription factors with putative target genes. Finally, by using ASP programming, we have analyzed the compatibility of these two graphs in order to validate or complete a regulatory network. This technique has been applied to transcriptomic experiments for *A. ferrooxidans* in 7 different stress conditions (Aravena et al., 2011).

3.2. Dynamics of biological networks. The group of Dr. Maass has proposed a mathematical model accounting for the dynamics of uptake, efflux, storage and traffic of transition metals Cu, Zn, Mn and Fe in *Halobacterium NRC-1.* We prove, in a formal way, that the system presents a stable, stationary state. Additionally, we derive monotonicity conditions for the existence of global steady state responses, independently of the choice of the parameters. The simulations and rigorous mathematical proofs confirm that the system attains an equilibrium, and that the machinery determined by the trafficking and uptake mechanisms allow to adapt the internal state of the cell to changes in the external level of metal ions concentrations, suggesting that the

systems are robust under external perturbations. These evidences corroborate that the trafficking and detoxification mechanisms are necessary for cellular survival (Espinoza et al., 2011).
In a second area of work, we revisit a well known conjecture proposed by the biologist R. Thomas relating the structure of signed regulatory graphs with the asymptotic behavior of the associated dynamical system in the Boolean case. As a result, we propose an alternative proof of the necessary condition in the second Thomas conjecture in the Boolean case. A "cohomological idea" emerges behind the sign assignment in the regulatory graph. In the particular case of an isolated circuit (i.e., a regulatory graph consisting purely of a circuit) we prove that conditions in Thomas' conjectures are also sufficient. This work has been submitted to *Theoretical Computer Sciences*.
From a more mathematical point of view we continue working on the analysis of low complexity symbolic sequences, like genomics sequences. In particular we relate nilsequences and concepts of independence. An article in collaboration with scientists from U. Sci. Tech. China was accepted in *Ergodic Theory and Dynamical Systems*, the main journal in the area (Dong et al, 2011).

3.3. Computational platforms and development of new software. Our work was devoted to the consolidation of the bioinformatics visualization platform SalmonDB that aims to become a national reference to exploit genomic information regarding the Atlantic salmon (*Salmo salar*) in an easy way, to perform fast comparative genomic research with other salmon databases and fish reference genomes, and to incorporate genomic information provided from the Atlantic salmon genome sequencing project. Up to October 2011, this platform was visited more than 2,500 times per month. Two articles related to this platform were accepted in *DATABASE* (Di Genova et al & Guberman et al., 2011).

3.4. Mathematical models of cell differentiation programs. Through a collaborative effort with the group of Drs. Rodrigo Assar and David Sherman at Bordeaux University, France, the team led by Dr. Montecino begun defining mathematical models that can reflect early gene expression decisions of uncommitted mesenchymal cells during bone differentiation engagement. We believe that mathematical modeling may allow us to predict the behavior of mesenchymal stem cell populations when challenged to engage towards a specific cell lineage commitment. A reflection of the productive interaction between mathematicians and molecular biologists is provided by a manuscript submitted for publication in the journal *Byosystems* (Assar et al., submitted). More importantly, Dr. Assar has recently returned to Chile and joined the CGR as a post-doctoral fellow under the joint supervision of Drs. Maass and Montecino. The goal of this new research program during the next year will be to reinforce this field within our Center and in parallel to experimentally evaluate the predictions of these mathematical models.

3.5. Discriminative local subspaces in gene expression data for effective gene function prediction. In the last decade, the complete genomes of hundreds of organisms have been sequenced and thousands of genes have been identified. Despite these efforts, biology still faces the relevant challenge of discovering the biological processes that these genes carry out inside the cells. For example, in the model organism in plant molecular biology, *Arabidopsis thaliana*, more than half of its genes (52%) are not annotated in any biological process in the Gene Ontology (GO) database. In order to guide this discovery process, the group led by Dr. Gutiérrez has developed Discriminative Local Subspaces (DLS), a novel supervised machine learning method designed to analyze gene expression data and predict new genes involved in a biological process of interest. The general workflow of DLS can be seen in Figure 4. The main feature of DLS is its use of existing knowledge to find what we call *expression signatures.* An expression signature corresponds to a discriminative expression pattern with two main properties: i) it arises in a local subspace of the data (i.e. in a subset of genes and a subset of experimental conditions) and ii) it is highly discriminative (exclusive) to the genes associated to the biological

process of interest. The discriminative nature of expression signatures allows DLS to identify new genes potentially involved in the process of interest. Theses predicted genes are then exposed in the context of a function specific discriminative coexpression network that shows the relevant transcriptional connections among the genes associated to the biological process of interest. In comparison to the traditional Coexpression Networks (CNs), DLS stand out for gene function prediction mainly because in CNs the coexpression between genes is measured among all the available experimental conditions, restricting its discriminative capabilities. In contrast, DLS searches for coexpression patterns defined by subsets of conditions that are discriminative to the genes involved in the biological process. On the other side, DLS overcomes the representational limitations of SVMs and other black box supervised methods, as it offers a prediction system based on biologically meaningful concepts behind coexpression. The expression signatures and the generated functional networks expose the key experimental conditions and genes used to make each prediction, providing valuable insights to understand the predictions, the process of interest and to guide future experiments. Given the incomplete nature of gene functional knowledge, DLS must overcome a mayor challenge for defining an informative training set for supervised learning. To this end, it incorporates three main features: first, a novel way to define an informative negative training set of genes, based on the current knowledge in GO; second, an option to dynamically discard an expected percentage of potential false negatives during the training process, so that they do not obscure the search of suitable expression signatures; and third, an independent pre-processing method that takes advantage of the previous features to iteratively discover potential false negatives and refine the training set. To test the prediction performance of DLS, we did a systematic analysis over 101 biological processes from GO in an *Arabidopsis thaliana* dataset and compared its results with previous works based in traditional Coexpression Networks (CNs) and Support Vector Machines (SVMs). The results show that in overall, DLS obtains better prediction performances than SVMs and CNs in both a cross validation using annotations from year 2008 (Figure 5) and an enrichment analysis in new annotations from 2010 (Figure 6). This work shows that supervised learning and coexpression approaches can be effectively combined to make biologically meaningful predictions, opening new opportunities to extract biological hypothesis from expression data and cope the need to understand gene function and biological processes.

Objective 4. Center Aims.

The following sub-aims were not explicitly stated as scientific aims of the CGR but rather as strategic aims of national interest (and are linked to one or more of the scientific aims). We describe these achievements in the following paragraphs.

4.1. Towards an understanding of plant adaptation to extreme environmental conditions in the Atacama Desert (Collaboration with Dr. Claudio Latorre of the Institute of Ecology and Biodiversity). The Atacama Desert is one of the driest places on earth, the average yearly precipitation in the region is around 1 mm (Latorre *et al.*, *GSA Bull.*, 2002; Houston & Hartley, *Int. J. Climatol.*, 2003). Plants that grow here are exposed to extreme environmental conditions like drought, high diurnal temperature oscillations, high salt content in the soil, and a very high level of solar radiation. Despite these harsh conditions, the Atacama hosts a surprising diversity of plant life. It is not entirely understood how plants can adapt to this highly extreme environment. In this CGR project, our aim is to provide insight into the molecular mechanisms that are responsible for plant abiotic stress tolerance in the Atacama Desert. Plant species were collected along an altitudinal transect between 2470 and 4460 m above sea level, representing the prepuna, puna and steppe ecosystem of the Atacama. Transcriptomic analysis can reveal new molecular functions responsible for plant abiotic stress tolerance (Dassanayake *et al.*, *New Phytol.*, 2009; Chen *et al.*, *Mar. Genomics*, 2011). Therefore, we have isolated total RNA from selected plant species, and rRNA-depleted samples have been sent to Eastern Sequence and Informatics Hub, University of Cambridge, for high-throughput sequencing using Ion

Torrent and Illumina next generation sequencing platforms. Analysis of the transcriptomes by bioinformatic tools can help to identify new genes responsible for adaptation to extreme environments. Furthermore, comparison of the abiotic stress responsive genes between plants and bacteria from extreme environments may shed light on common mechanisms for abiotic stress tolerance in prokaryotes and eukaryotes. Phylogenetic and phylogeographic analysis of the plants may reveal the origin of abiotic stress responsive genes. It is known that plants in nature do not function as autonomous individuals, but accommodate diverse microbial communities (Hardoim *et al.*, *Trends Microbiol.*, 2008; Aly *et al.*; *Appl. Microbiol. Biotechnol.*, 2011). This interaction is beneficial for both of the partners: the microbes gain shelter, nutrition, and dissemination, and facilitate plant growth and survival at the same time. Several lines of evidence suggest that plant-associated microbiota have crucial importance in plant abiotic stress tolerance, especially under extreme environmental conditions such as alpine habitats, and arid or semiarid ecosystems (Yang *et al.*, *Trends Plant Sci.*, 2009; Singh *et al.*, *Plant Signal Behav.*, 2011). To understand the role of plant-microbe interactions in adaptation to extreme conditions in the Atacama Desert, we are investigating plant-associated fungal and bacterial communities. Since the traditional methods like 16S rDNA or ITS library construction and sequencing have several limitations (Hong *et al.*, *ISME J.*, 2009; Gazis *et al.*, *Mol. Ecol.*, 2011), we are applying high-throughput sequencing using Illumina technology to achieve the highest possible coverage. In summary, genetic analysis of plants grown in the Atacama Desert can contribute to our understanding of plant abiotic stress tolerance. These results may be of great importance for enhancement of crop tolerance to abiotic stresses.

4.2. Genomic adaptations and evolution of high-altitude dwelling vertebrates (Collaboration with Dr. Marco Méndez of the Dept. of Ecology, Faculty of Sciences, U. Chile). As we stated in our proposal, we have identified species of fish that inhabit unique environments in the high-altitude salt lakes in Northern Chile, the *Altiplano*. Most of these fish are of the genus *Orestias*, found in three South American countries: Bolivia, Perú and Chile. A founder species, probably *O. agassiz*, lived in a large continental lake encompassing the present distribution, which became fragmented as the Andes mountains rose about 30.000 years ago. This has lead to allopatric speciation of the *Orestias* fish as each lake or salt lake has its own species. Even today, speciation is ongoing as populations within a salt lake have become geographically isolated and differentiation can be seen in morphology and karyotype. We are interested in examining the behavior of genes and genomes in this rapidly speciating group of teleosts, as it is an unprecedented opportunity to witness evolution at the molecular level. We have selected two populations of *O. ascotanensis*, living in the *Salar de Ascotán* (21°34'25"S, 68°18'32"W), to carry out a low coverage, full genome sequence. We will compare this to the sequence of a related, but geographically distant sister species, *O. chungaraensis*, as well as to other teleosts which have been sequenced previously. We hope to find a link between habitat and genomic adaptation and to find genes that are responsible for the high level of environmental and developmental plasticity in this group.
Another group of animals that we have chosen to study in paralell are frogs of the species *Rhinella spinulosa.* This species lives in a diverse set of environments with a high range of temeperatures and seasonal variations and the different populations have adapted to their habitat with very specific reproductive strategies. How a species can achieve this physiological plasticity is unknown and through transcriptomic analysis we hope to unravel this mystery.

4.3. Admixture mapping of metabolic traits in a Chilean Hispanic/Mapuche population. Our goal is to understand the unique genetic characteristics of the Chilean population and identify variants that are linked to prevalent medical conditions in Chile. As a first step towards this goal, we decided to takle the challenge of whole genome sequencing of the DNA of selected Chileans, representative of the 3 major ethnic groups. This

proposal, was formally presented to the ethical committee of the Faculty of Medicine at the Pontificia Universidad Católica de Chile and the project was approved. DNA samples that met strict quality requirements were selected and sent to the newly created center for sequencing in Chile, OMICS-Solutions, as well as to a very competent international facility, Complete Genomics, Inc. (2071 Stierlin Court, Mountain View, CA 94043; USA). Due to the relevance of this aim for Chile, we decided to focus our first effort in one of the most relevant ancestry ethnic groups, the Mapuches (Huilliches). Because Associate Investigator Juan Francisco Miquel has longstanding experience in the study of risk factors for common metabolic diseases in this population, and because we have a clinical database coupled to metabolic studies and DNA bank of an isolated Huilliche community of the south of Chile (Huapi Island, Ranco Lake), we proceeded with whole genome sequence of 10 Huilliche individuals (undergoing sequencing at the time of this Annual Report submission). This community is unique in terms of isolation and high grade of consanguinity with almost 4 founder names, according to the ethnographic history of the community. The study of the ethnographic history of the island was performed in 1994 under the supervision of Dr. Miquel and in collaboration with the school of anthropology of the Austral University in Valdivia-Chile; a book with the history will be edited with the support of CRG in early 2012. By doing whole genome sequencing of these individuals and by comparing with the available information in HAPMAP and the 1000 genomes project, we will capture for the first time a significant number (probably >100) of private genetic variants present in the original population of Chile; these variants should be present and widely distributed in the Latin Mestizo Chilean population of biparental origin (Spaniard and Amerindian). With this information in our hands, which will be available during the first semester 2012, we plan to develop the following aims: (i) To define the prevalence of selected private Huilliche genetic variants in a general Chilean mestizo population (n~2000 individuals); (ii) To identify new variants of risk for Gallstone disease and gallbladder cancer in the Chilean population, two diseases highly prevalent and with major health burden in Chile. Cases and controls were well characterized and a clinical database and DNA bank is available in our labs; (iii) To find new genetic variants with potential functional role in prevalent disease could be study by our group at the molecular and cellular level.

4.4. Discovery of new genetic risk factors for cholesterol gallstone disease in the Chilean population. Dr. Miquel is in the final step of the study of 7 families with pediatric index cases of cholesterol gallstone disease (i.e. extreme index cases), a disease that normally affects almost exclusively adults. A systematic classic prospective family study, with a well-characterized phenotype and DNA samples of more than 18 members per family core, will be finished by December 2011. This study represents and unique opportunity to find new genetic variants for gallstone disease (*lithognic* genes) in Chile. With expert international collaborators we will develop classic linkage studies by using state-of-the-art technology: GWAS (whole genome association study) and/or exome sequencing. We have estimated the chance of finding new *lithogenic* genes by this strategy of 50-60%. Then, we can easily study the relevance of these new *lithogenic* genes in the general Chilean population.

4.5. Analysis of the A357D Acid Sphingomyelinase Chilean Gene Mutation in Niemann Pick type B. Niemann Pick type B (NPB) is one of the most frequent lysosomal storage diseases in Chile, with at least 40 confirmed cases (P. Mabe, unpublished results). The disease is produced by mutations in the acid sphingomyelinase gene that encodes for a lysosomal hydrolase that degrades sphingomyelin to ceramide and phosphocholine. Clinical signs include moderate to severe hepatosplenomegaly, hypercholesterolemia, hypertriglyceridemia, hemorrhagic diathesis and progressive interstitial pulmonary damage. NPB severely deteriorates the patient's quality of life, and can also lead to death during the first ten years of life. In the Chilean context, NPB is particularly intriguing from the clinical and molecular point of view since the A357D mutation was found to be unique to Chilean patients (Simonaro et.al., 2002; Mabe, unpublished results). From a total of 18 patients,

15 were A357D homozygotes, and 3 were compound heterozygotes for this mutation. Considering the high frequency in which the A357D mutation occurs in Chilean NPB patients, a founder effect is likely. Our interest focuses on revealing the frequency of this mutation in the Chilean population without NPB, so as to eventually measure the prevalence rate of this disease in Chile. Our preliminary data indicate that the frequency of heterozygous is of 2% and 7% in the Hispanic and Mapuche Chilean populations, respectively, a much higher prevalence than that estimated by the number of until now confirmed cases. In addition, we want to study the functional consequences of the mutation in the expression, activity and intracellular localization of the protein. This research line is part of the PhD thesis of Mariana Acuña, student of the PhD Medical Sciences Program from the Faculty of Medicine, UC.

4.6. Sequencing the Thompson seedless genome (Collaboration of the CGR with P. Hinrichsen, INIA-La Platina). Thompson seedless is one of the main fruits exported by Chile. One of the features of this variety is the size of the berry which requires exogenous applications of the hormone gibberelin (GA). In order to get varieties that are independent from GA applications, a breeding program using molecular markers is being developed. As a way to obtain polimorfisms linked to table grapes, a draft genome is being built up. This is the first genome from a table grape since the previously reported genome is from *Pinot noir*, a grapevine used for wine production. Thus, we are focusing on the identification of genes and polymorphisms related to table grape quality.

4.7. Genomics and Transcriptomics of flowering plants from the Atacama desert. (Collaboration of CGR with researchers from U. La Serena and CEAZA). Several plants species are dormant in the world´s driest dessert. However, once in a while, a few millimeters of rain lead to the growth and flowering of many of these species. Different amounts of water are required to induce the growth of different species. Therefore, questions such as what are the mechanisms involved in the differential sensing of water and how these plants are able to live and reproduce under these extreme conditions will be addressed by looking at gene expressed in seeds induced to germinate. Furthermore, given the number of different species growing under these stressful conditions, their genetic variation will be addressed by genomic analyses.

4.8. Using genomics to study host-pathogen interactions between Atlantic salmon and *Piscirickettsia salmonis*. We are studying the host-pathogen interactions between Atlantic salmon (*Salmo salar*) and *Piscirickettsia salmonis*, the causative agent of salmonid rickettsial septicaemia (SRS), the bacterial disease with the highest economic impact for the Chilean salmon culture industry. In doing so, we ranked forty full-sibling Atlantic salmon families according to accumulated mortality after challenging with *P. salmonis* and selected the families with the highest and the lowest resistance levels for gene expression analysis using 32K cGRASPS microarrays. We detected significant differences between the transcriptional responses of Atlantic salmon families with variable levels of resistant to the pathogen. We identified predictor genes of high and low resistance to infection, as well as, biological processes and metabolic pathways involved in the fish response to the pathogen, including changes in iron metabolism and energy production. Our results contribute to identify underlying mechanisms of disease resistance and can be used for stock-improvement programs. We have sequenced the genomes of virulent and non-virulent strains of *P. salmonis*. The comparative analysis of bacteria genomes will provide evidences of the virulence mechanisms of *P. salmonis* and reveal virulence factors and other gene products that may have application as vaccines or immunomodulatory candidates.

4.9. Bioinformatics platform. We have consolidated the bioinformatics platform Math^{omics} as a joint venture between the CGR and th CMM (Center for Mathematical Modeling). It has become a highy useful platform

providing state of the art bioinformatics and mathematical modeling tools, allowing us to face biotechnological problems from a Systems Biology point of view (www.mathomics.cl). During 2011 we have developed several bioinformatics pipelines in order to process data arising from projects in the areas of genomics, transcriptomics, metagenomics, metabolomics, massive probe design for molecular bioidentification, metabolic modeling and analysis of biological networks. The consolidation of this platform has stimulated us to start new collaborations with national and international groups. Among them we mention: 1) INIA: collaboration in Potato genome sequencing project (BGI-Shenzhen et al., 2011), analysis of mRNAseq experiments of *Prunnus persica*, development of the PotatoDB platform; 2) AquaINNOVO: development of bioinformatic techniques for the construction of SNP-chips for Atlantic Salmon and Tilapia; 3) School of Medicine, U. Chile: we will develop the bioinformatics tools of the project for genotyping chilean populations. The Math^{omics} platform was involved during 2011 in several sequencing projects: we started the sequencing project to produce a reference sequence for *Thompson seedless*, the main chilean export grape; we started the sequencing of six relevant biomining bacteria in order to develop a metagenomics program in this area (4 are ready); we were invited to participate in the potato sequencing project (BGI-Shenzhen et al., 2011), A. Maass is member of the scientific committee of the consortium for sequencing the Atlantic salmon and we are carrying out backup tests for this sequence; we were involved in the analysis of three hightroughput mRNAseq experiments: *Thompson seedless* grape, *Prunnus persica* and the algae *Nannochloropsis*.

ii. Synergy and collaboration among research lines

Interaction between Investigators and the research lines they carry out is intrinsic to the way this project was concieved. Although the three scientific aims appear dissimilar, they intersect due to the need to use the same approach to engage them: the use of advanced omic technologies and the requirement for powerful analytical tools that must be developed in-house. In addition, Aim 3 was designed to become the differentiating factor in our Center: massive biological data is meaningless unless it can be put into context and organized in mathematical terms; Aims 1 and 2 are beginning to benefit from this key strength of our group. Furthermore, Strategic Aim 4 is the responsibility of all CGR researchers, as we expect every investigator to become involved in the sequencing projects that will lead to the description of new genomes and transcriptomes. While we included each researcher assigned to only one of the three scientific aims, in fact, the great majority are working on projects that are related to more than one aim (this can be gauged by examining the participants named in each of the sub-aims in section i). As a further measure of the synergistic interplay between research lines, we can mention that 7/34 of the published ISI papers, 2/5 of the Non-ISI papers, and 25/138 of the meeting abstracts involved more than one Aim. Furthermore, of the 34 published papers, two were co-authored by two CGR investigators and one was co-authored by three of them. The values for the indicators that reflect synergistic activities within the CGR shall increase gradually as the projects we have initiated become mature.

A key aspect of our coordinated work has been the continuous interaction among the six Principal Investigators. We have had meetings practically every single week in 2011 as we reserved a fixed day and time in our schedules for this purpose. Advances in the different research lines, problems and obstacles, new ideas and proposals for activities were discussed at these meetings, which we will continue to have throughout the project's duration. They were fundamental in promoting the generation of interactions between researchers and the different research lines. We also took advantage of the *Interactomics* seminar series, where the Associated Investigators and all personell related with our labs participated. Finally, we had a two day retreat outside of Santiago, where all 14 researchers presented their plans for research within the context of the CGR project; details on this meeting and a photograph can be found in the Appendix.

In addition to the scientific results, we have also invested effort and funds on establishing a physical location for our meetings and administrative offices. We rented space in a building where the Director and administrators, journalist and other personnel have offices and in two large rooms we set up a bioinformatics cluster, where currently about 10 people work. This required special high-speed internet connections to be set up and the purchase of computing stations. We also have a large meeting room for the weekly investigator meetings and for seminars. The fact that we can meet in a neutral place outside of the universities and establish a presence as a Center has reinforced our view on the necessity to find ways in which to eventually bring our research groups together. We have planned to expand our use of the current building (additional floors will be rented) and alternatives for common lab space are being explored.

iii. Formation of advanced human capital directly related to the Center´s objectives

In our description of the progress in all of the objectives presented above, we stressed the contributions of PhD students and postdocs to the work as they are the principal authors of a great majority of the papers produced this year by the CGR. We strived to incorporate these young scientists into all of the research lines even though many of the topics were being initiated just this year with the birth of the project. We must say that a great majority of the students and postdocs recruited to our Center have obtained their own funding (fellowships from CONICYT and other agencies), and thus did not receive their stipends from the FONDAP program. This has allowed us to bring in more postdocs than originally anticipated, for example. We consider the emphasis on incorporation of young talent into the CGR as one of our principal accomplishments for 2011; a total of 84 students (43 of them PhD students) and 24 postdocs worked in our labs this year. Besides the formation of young scientists in our labs, we have also had a broader impact on training through the following activities (posters and announcements included in the Appendix):

1) We have created the Bioinformatics course within the new Master's Program in Medical Informatics given at University of Chile starting 2011. The Program has been launched during the 2nd semester of 2011 at the Faculty of Medicine and has numerous enrolled students.

2) We organized the first course on bioinformatics for the analysis of metabolomics data for graduate students in Santiago (see Appendix). We selected 25 students from different PhD programs in biology or biotechnology. It consisted of 10 sessions of theoretical and practical lessons.

3) Together with the Center for Mathematical Modeling and the SCIAN group, we organized the post-graduate course: “Mathematical and Bioinformatic Methods in Biological Data Analysis”, a course running from December 2, 2011 to January 14th, 2012 at the Faculty of Medicine, Universidad de Chile.

4) We coorganized a workshop on genetics, genomics and technical strategies for using the zebrafish as a model organism. We carried out this workshop in Porto Alegre, Brazil, together with the Universidade Catolica do Río Grande do Sul and we had 95 students enrolled from all over Latin America. It consisted of a week of theoretical lectures and practical sessions. Dr. Allende participated in another course as an invited instructor: “Experimental strategies for the molecular analysis of zebrafish development”, held at the Universidad Nacional Autónoma de México, Mexico City, September 26 to October 1, 2011.

iv. Collaborative networks both at the national and international level

While we concentrated on the initiation of the Center and organization of activities for 2011, we were still able to sustain an important number of exchanges and solidify the collaborative network that has been the signature of this group of researchers for many years. 36 exchanges with foreign colleagues were carried out, some of these new interactions established because of the new research lines bing implemented at the CGR.

As can be seen from our publications, much of our work is done in multidisciplinary international collaboration. Something we would like to work on in the near future is to establish alliances with Centers similar to ours around the world. We have had some interesting discussions with colleagues at the EMBL-Monterotondo and with the Centre de Regulació Genomica of Barcelona, Spain, for example. Ideally, we strive for agreements that could facilitate student and researcher exchanges.
Of critical importance to our project, the CGR has established a tight relationship with the "Centro Nacional de Genómica, Proteómca y Bioinformática" (CNGPB). This large equipment project, financed by CONICYT and managed by the Universidad de Chile, Universidad Católica de Chile, Universidad Andres Bello, Universidad de Talca and the Instituto Nacional de Investigaciones Agropecuarias was created to promote the use of genomics in research and development in Chile. Since its beginning, the CNGPB has consulted its choices and priorities with the CGR to make a better use of the platforms for next generation sequencing (NGS) and its applications that the CNGPB has implemented to reach its objective. This coordination has allowed more favorable negotiations with NGS platform providers in such way that with the funds originally assigned for two platforms has been expanded, making it possible to purchase the three main platforms being used worldwide. The CRG has contracted and received NGS services from CNGPB. Up to date the following services have been provided: runs for RNAseq in SOLID4 (life Technologies), runs for DNA fragment sequencing in 454 Junior (ROCHE) and runs in Ion Torrent (Life sequencing) using microchip 314 and 4 using microchip 316.

v. Dissemination and exploitation of results

Our main forum for dissemination of the results and work carried out at the CGR was the *Interactomics* seminar series. We chose the name to symbolize the expectation that it will lead to further interactions between CGR investigators and also with external colleagues. The seminars were held once a month from March to November and included presentations by graduate students, postdocs and invited professors; two of them foreign and one Chilean. We provide the announcement posters for all of these seminars in the Appendix. The locales for the seminars rotated among all of the participating institutions and they attracted from 75 to 150 attendants.
We also organized a workshop in Plant Genomics and Systems Biology (10-11 November, 2011) at the Pontificia Universidad Católica de Chile where several of our researchers and students presented their work (see also Appendix for poster and program).
We organized the Workshop "Estudios Genéticos Poblacionales en Chile: Taller de análisis sobre aspectos bioéticos y marco regulatorio" (Population genetic studies in Chile: workshop discussion of bioethical issues and regulatory norms). We organized for the first time in Chile this relevant workshop with the aim to discus with a panel of experts and with the scientific community what we now and what we still need to know and to develop in bioethical and legal aspects of population genetic studies in developed and developing countries like Chile. Two international and four national speakers were invited to present relevant topics and the audience (around 60 participants) was composed of a wide range of researchers and members of Ethics Committees of different Institutions, and authorities of health-related National Organizations (i.e. Ministry of Health, CONICYT, ISP, CORFO). The discussion was very fruitful and was a general agreement that in Chile we need to improve rules and laws to better regulate and guide genetic population studies in Chile. It was also clear that genetic studies in indigenous communities offer a greater challenge to be solved promptly (the program of the workshop made in October is included in the Appendix). A detailed report of this workshop and its findings is being prepared and will be published as a book by the CRG.
Our Center was very active in scientific meetings, as our students and researchers presented a total of 140 presentations at national, Latin American and international events.

vi. Outreach to society

Our Center decided to place a strong emphasis on communicating our science to the general public and, to that end, we hired a person in charge of communications for the CGR. Mrs. Karen Meyer is a journalist with several years of experience in print and television media, with a desire to specialize in scientific journalism. She became part of the CGR in July, 2011 and for these six months has become familiar with the people involved in the project, the science carried out in our labs and the topics which can be transformed into informative packets for the general public through mass media. One of our first activities in this regard, was to develop a radio program, which is already being aired on Radio Universidad de Chile, called "The Sounds of Science". This program consists of a first hand account by scientists at the CGR on how they carry out their work. For achieving a high story value, the program followed two groups of researchers to the field (the Northern desert of Chile) where they were collecting plant and animal samples for genomic analysis; the subsequent shows followed them into the labs where they carried out the experimental work. A combination of live sounds from these experiences and first hand accounts by the scientists gave the programs a very interesting and novel format.

During 2011, we worked with the Science and Technology Museum (Quinta Normal, Santiago) to enhance the Vegetalista exhibit. The Vegetalista exhibit aims to teach about plant biology and plant biotechnology. The target audience are middle and high school students and teachers. In collaboration with the Millennium Nucleus for Plant Functional Genomics we set up a module in the exhibit with 2 plasma televisions showing videos about plant biology and biotechnology, interviews with several Chilean scientists as well as educational videos about plant movement kindly donated by Roger Hangarter (Indiana University). We also implemented a computer with the 3D video game Vegetalista. This video game teaches players about Chilean geography and native flora. More information about the exhibit and some of the content presented at the museum can be found in the exhibits website www.vegetalista.cl

Our communications expert also developed a logo and a corporate image which will identify the Center in all of its materials and activities, both internally and externally; developed a web page (www.centrogenomica.cl; for now mostly in Spanish), letterhead, business cards, office materials and signage for identifying CGR installations. All of the activities of the Communications area of the CGR, together with samples of the material produced, are detailed in the Appendix.

References:

Henriquez, B., Hepp, M., Merino, P., Sepúlveda, H., van Wijnen, A. J., Lian, J. B., Stein, J. L., Stein, G. S., Montecino, M. (2011) C/EBPb binds to the P1 promoter region of the Runx2 gene and upregulates Runx2 transcription in osteoblastic cells. J. Cell. Physiol. 226: 3043-3052.

Grandy, R., Sepulveda, H., Aguilar, R., Pihan, P., Henriquez, B., Olate, J., Montecino, M. (2011) Ric-8B gene is highly expressed in proliferating pre-osteoblastic cells and down-regulated during osteoblast differentiation in a SWI/SNF- and C/EBPb-dependent manner. Mol. Cell. Biol. 31: 2997-3008.

Ugarte, G., Opazo, T., Leisewitz, F., van Zundert, B., Montecino, M. (2011) Runx1 and C/EBPb transcription factors directly up-regulate P2X3 gene transcription. J. Cell. Physiol. In press. Epub June 15.

van der Deen, M., Akech, J., Lapointe, D., Gupta, S., Young, D., Montecino, M., Galindo, M., Lian, J., Stein, J., Stein, G., van Wijnen, A. (2011) Genomic promoter occupancy of runt-related transcription factor Runx2 in osteosarcoma cells identifies genes involved in cell adhesion and motility. J. Biol. Chem. In press. Epub December 8.

Medina, R., Ghule, P., Cruzat, F., Montecino, M., Stein, J., van Wijnen, A., Stein, G. Epigenetic control of cell cycle-dependent histone gene expression to accommodate the abbreviated pluripotent cell cycle. Manuscript submitted.

Assar, R., Leisewitz, A. V., Garcia, A., Inestrosa, N. C., Montecino, M., Sherman, D. J. Reusing and composing models of cell fate regulation of human bone precursor cells. Manuscript submitted.

Aly AH, Debbab A, Proksch P (2011): Fungal endophytes: unique plant inhabitants with great promises. Appl. Microbiol. Biotechnol. 90(6): 1829-45.

Anderson-Carpenter LL, McLachlan JS, Jackson ST, Kuch M, Lumibao CY, Poinar HN (2011): Ancient DNA from lake sediments: bridging the gap between paleoecology and genetics. *BMC Evol. Biol.* 11:30. DOI:10.1186/1471-2148-11-30.

Chen S, Zhou R, Huang Y, Zhang M, Yang G, Zhong C, Shi S (2011): Transcriptome sequencing of a highly salt tolerant mangrove species Sonneratia alba using Illumina platform. Mar. Genomics 4(2): 129-36.

Dassanayake M, Haas JS, Bohnert HJ, Cheeseman JM (2009): Shedding light on an extremophile lifestyle through transcriptomics. *New Phytol.* 183(3): 764-75.

Gazis R, Rehner S, Chaverri P (2011): Species delimitation in fungal endophyte diversity studies and its implications in ecological and biogeographic inferences. *Mol. Ecol.* 20(14):3001-13. doi: 10.1111/j.1365-294X.2011.05110.x.

Hardoim PR, van Overbeek LS, Elsas JD (2008): Properties of bacterial endophytes and their proposed role in plant growth. *Trends Microbiol.* 16(10): 463-71.

Hong S, Bunge J, Leslin C, Jeon S, Epstein SS (2009): Polymerase chain reaction primers miss half of rRNA microbial diversity. *ISME J.* 3(12):1365-73.

Houston J, Hartley AJ (2003): The central Andean west-slope rainshadow and its potential contribution to the origin of hyper-aridity in the Atacama Desert. *Int. J. Climatol.* 23(12): 1453–1464.

Latorre C, Betancourt JL, Rylander KA., Quade J (2002): Vegetation invasions into absolute desert: A 45,000 yr rodent midden record from the Calama-Salar de Atacama Basins, northern Chile (lat 22°-24°S). *GSA Bull.* 114(3): 349-366.

Singh LP, Gill SS, Tuteja N (2011): Unraveling the role of fungal symbionts in plant abiotic stress tolerance. Plant Signal Behav. 6(2): 175-91.

Willerslev E, Hansen AJ, Binladen J, Brand TB, Gilbert MT, Shapiro B, Bunce M, Wiuf C, Gilichinsky DA, Cooper A (2003): Diverse plant and animal genetic records from Holocene and Pleistocene sediments. Science 300(5620): 791-5.

Yang J, Kloepper JW, Ryu CM (2009): Rhizosphere bacteria help plants tolerate abiotic stress. Trends Plant Sci. 14(1): 1-4.

Vázquez MC, del Pozo T, Robledo FA, Carrasco G, Pavez L, Olivares F, González M, Zanlungo S. Alteration of gene expression profile in Niemann-Pick type C mice correlates with tissue damage and oxidative stress. Plos One 2011 (in press). dx.plos.org/10.1371/journal.pone.0028777

Yévenes LF, Klein A, Castro JF, Marín T, Leal N, Leighton F, Alvarez AR, Zanlungo S. Lysosomal vitamin E accumulation in Niemann-Pick type C disease. Biochim Biophys Acta. 2011 (in press). 10.1016/j.bbadis.2011.11.009

Balboa E, Morales G, Aylwin P, Carrasco G, Amigo L, Castro J, Rigotti A, Zanlungo S. Niemann-Pick C2 Protein Expression Regulates Lithogenic Diet-Induced Gallstone Formation and Dietary Cholesterol Metabolism in Mice. Lipids. 2011 (in press). 10.1007/s11745-011-3625-2.

Lisbona F, Rojas-Rivera D, Thielen P, Zamorano S, Todd D, Martinon F, Glavic A, Kress C, Lin JH, Walter P, Reed JC, Glimcher LH, Hetz C. 2009. BAX inhibitor-1 is a negative regulator of the ER stress sensor IRE1alpha. Mol Cell, 33(6):679-91.

Gallardo VE, Liang J, Behra M, Elkahloun A, Villablanca EJ, Russo V, Allende ML, Burgess SM. 2010. Molecular dissection of the migrating posterior lateral line primordium during early development in zebrafish. BMC Dev Biol 10:120.

d'Alençon C, Peña O, Wittmann C, Gallardo VE, Jones R, Loosli F, Liebel U, Grabher C, Allende ML. 2010. A high-throughput chemically induced inflammation assay in zebrafish. BMC Biology 8:151.

Hernández PP, Undurraga C, Gallardo VE, Mackenzie NC, Allende ML, Reyes AE. 2011. Waterborne copper exposure influences gene expression, cell survival and embryonic development in zebrafish embryos. Biol Res 44:7-15.

Allende ML*, Calcaterra NB, Viana MR, Zolessi FR. 2011. First Meeting of the Latin American Zebrafish Network. Zebrafish 8:1-3.(*Corresponding author).

Valdivia LE, Young RM*, Hawkins TA, Stickney HL, Cavodeassi F, Schwarz Q, Pullin LM, Villegas R, Moro E, Argenton F, Allende ML* and Wilson SW*. 2011. Lef1-dependent Wnt/ß-catenin signalling drives the proliferative engine that maintains tissue homeostasis during lateral line development. Development 138:3931-3941. (*Corresponding authors).

Villegas R, Martin SM, Sagasti A, Allende ML. 2012. Dynamics of degeneration and regeneration in zebrafish peripheral axons reveals a requirement for extrinsic cell types. (Submitted).

Feijoo CG, Oñate, M., Milla, L. and Palma, V. 2011. Sonic Hedgehog (Shh)-Gli signaling controls neural progenitor cell division in the developing tectum in zebrafish. Eur. J. Neurosci. 33(4):589-98.

Rapacioli M, Joao Botelho, Gustavo Cerda, Santiago Duarte, Matías Elliot, Verónica Palma and Vladimir Flores. 2012. Sonic hedgehog (Shh)/Gli modulates the spatial organization of neuroepithelial cell proliferation in the developing chick optic tectum. (Submitted).

Martinez, C., Cornejo, Victor Hugo, Lois, Pablo and Palma V. 2012. Enhanced midbrain neurogenesis in three-dimensional neurospheres cultures: a functional study with Shh (Submitted).

Bergeron S, Milla LA, Villegas, R Shen M-C, Burgess S, Allende ML, Palma V, Karlstrom RO. 2008. Expression profiling identifies novel Hh/Gli regulated genes in developing zebrafish embryos. Genomics 91:165-177.

Milla LA, Claudio R. Cortés, Christian Hodar Q., Maritza Oñate, Veronica Cambiazo, Shawn Burgess and Veronica Palma. 2012. Yeast-based assay identifies novel Shh/Gli target genes in vertebrate development. BMC Genomics 13:2.

a. Describe unexpected difficulties encountered and indicate how they were dealt with.

While we did not encounter major difficulties during 2011, we can mention two things that were a hindrance to the rapid advance of some of the projects during this first year. As we were awarded the project at the end of December 2010, we did not have enough time to begin a strong effort in recruitment of postdocs and PhD students to the Center until after the summer holidays, in March 2011. Thus, many of the young scientists entered the project only at the middle of the year and a significant part of the work got underway at that point.
Secondly, we depended for several of our large sequencing efforts on the implementation of the equipment at the National Genomics and Proteomics Center. This Center was awarded a Large Equipment Grant from CONICYT, but due to slow importation procedures, and the natural length of time it takes to set up these complex machines, these only became available to us late in 2011. Therefore, several of our sequencing projects have not been completed as of this writing. Nonetheless, they will be ready in early 2012 and we will report on our progress in this front in next year's report.

2. RESULTS ACHIEVED PER RESEARCH LINE

Briefly describe the main results per research line achieved during the period.

Since we have described in detail the principal advances in our research in the previous section, here we will summarize the main achievements related to each research line.
Research Line 1: this line has been expanded from the topic of epigenetic control of cell differentiation to include neurogenesis, regeneration, stem cells and tissue repair, developmental pathways and genetic contribution to neurodegeneration. All of these sublines have shown progress and 14 papers were published within this line (in addition to 5 papers in combination with line 2). These papers reveal important advances in transcriptional control of bone differentiation, sensory cell development, signaling in stem cell maintenance and differentiation and in technical advances in tissue regeneration and organ repair.
Research Line 2. The original aim of analyzing the stress response has been widened, as we have not only examined stress mechanisms in plants and bacteria but also in animal cells. Our research on molecular mechanisms governing how molecular chaperones in diverse organisms are controlled to allow the cell to cope with environmental insults is advancing rapidly. Twelve papers associated with this research line were published (in addition to 5 in combination with line 1 and one with line 3).
Research Line 3. This line was designed to act as a support framework for all of the experimental projects ocurring within the CGR. True to this objective, of the three papers published associated with Line 3, two of them were in association with other lines (2 and 4) of the CGR. We also not that one of the papers (published in *Ergodynamic Theory and Dynamical Systems*) is one of the top journals in the field of dynamical systems.

Research Line 4. This is a strategic line of research for the Center and it takes advantage of the access we have to advanced sequencing technologies and our strength in data analysis for Omic projects. This line is only just beginning to bear fruit, as the projects got under way in late 2011. One very important paper for our Center, published in the journal *Nature*, describing the sequencing and analysis of the potato genome, serves as an illustration to what we aspire to achieve with this aim.
We also point out that, within Line 4, we have sought out many new lines of inquiry with experts in different fields where we believe we can make important contributions. Such is the case with the sequencing of the Chilean Human Genome, genomes of fish and plant species of the northern desert, genomes of pathogenic bacteria and of biomining microorganisms.

The following table, summarizes our total production in the different categories (deatiled in the annexed tables) per research line:

	Research Line					
	1	**2**	**3 (or 4)**	**1,2**	**1,3**	**2,3**
ISI Papers	14	12	2	5	0	1
Non-ISI papers	1	0	2	2	0	0
Book chapters	1	1	0	0	0	0
Congress presentations	61	37	15	6	3	16

V. SUGGESTIONS FROM PREVIOUS EVALUATION

Not applicable. This is the report for the first year of existence of the CGR.

VI. PRODUCTS GENERATED BY THE PROJECT

In what follows, complete the attached Excel spreadsheets taking into account the following:

REPORT ONLY PUBLISHED MATERIAL INCLUDING THOSE WITH AN OFFICIAL DOI POINTER (e.g., with EARLY ONLINE ACCESS).

EXCEPT FOR BOOKS, ALL BACKUP DOCUMENTS MUST BE PRESENTED IN DIGITAL FORMAT. DO NOT SEND PRINTED COPIES.

ONLY PUBLICATIONS THAT ACKNOWLEDGE THE FONDAP PROGRAM WILL BE CONSIDERED.

1. ISI Publications

- ✓ For each publication, if applicable, the principal author and the corresponding author must be indicated using the following terminology:
  - [1] For principal author (example: Toro[1], J.)
  - [2] For the corresponding author (example: Toro[2], J.)
  - [3] For principal and corresponding author (example: Toro[3], J.)
- ✓ Include a digital copy of each **PUBLISHED** paper.

2. Non ISI Publications

- ✓ For each publication, if applicable, the principal author and the corresponding author must be indicated using the following terminology:
  - [1] For principal author (example: Toro[1], J.)
  - [2] For the corresponding author (example: Toro[2], J.)
  - [3] For principal and corresponding author (example: Toro[3], J.)
- ✓ Include a digital copy of each **PUBLISHED** paper.

3. Books and book chapters

- ✓ Include a hard copy of every **PUBLISHED** book.
- ✓ Include a digital copy of the front page of the chapter in the case of a book chapter.

4. Patents

- ✓ Include all patents generated by the FONDAP Center.

5. Congress presentations

- ✓ Include abstracts of all presentations. Attach a digital copy of the front page of the congress/workshop book.

6. Organization of Scientific Meetings

- ✓ List all congresses, courses, conferences, symposia, or workshops organized by the FONDAP Center.
- ✓ Include abstracts of all presentations. Attach a digital copy of the front page of the congress/workshop book.

7. Collaborative Activities

- ✓ List the scientific visits of Center members to international institutions
- ✓ List the scientific visits of foreign researchers to the Center in Chile.

8. Postdoctoral Fellows

- ✓ List postdoctoral fellows working in the Center during the reported period regardless of their funding sources.
- ✓ Provide current affiliation and positions held by former postdoctoral fellows that left the Center during the reported period

9. Students

- ✓ List titles of theses framed in the project completed during the reported period. Attach an abstract and the subject index.
- ✓ List titles of theses in progress, framed in the project, during the reported period. Include digital copies of the corresponding thesis registrations.
- ✓ Provide current affiliation and positions held by former students that graduated during the reported period

10. Funding Sources

- ✓ List all funding sources including FONDAP.

VII. OTHER ACCOMPLISHMENTS

Report articles or notes published in the media (provide URL links, if available), awards, prizes, etc.

We provide a full report on Comunictions and Outreach activities in the Appendix.

Awards and Honors:
Dr. Rodrigo Gutiérrez has been awarded a prestigious Howard Hughes Medical Institute Investigator position. In a very difficult competition, carried out at the HHMI Center in Janelia Farms, Virginia, Dr. Gutiérrez became the only Chilean to obtain such an award in this round (held every five years). The official announcement will be made in January 2012.

Dr. Montecino was invited to give the prestigious lecture Hermann Niemeyer at the IUBMB/FASEB annual meeting in Seville, Spain.

Dr. Allende, together with Associate Investigator Dr. Egaña, were awarded a grant form the International Centre for Genetic Engineering and Biotechnology (ICGEB) which will start in 2012.

Earthquake donations. Dr. Miguel Allende was instrumental in the transfer to Chile of a very important shipment of equipment items donated by the National Institutes of Health (Bethesda, USA). Four Chilean Universities affected by the earthquake that struck the country in February 2010 were benefitted by this donation. The initiative was spurred by Chilean postdocs currently working at NIH, led by Dr. Pablo Moya, with collaboration of Dr. Allende, who acted as Chilean liason and coordinated distribution of equipment to affected institutions. Dr. Allende met in Washington with the Director of Intramural Research at NIH, Dr. Michael Gottesman and at the State Department with John Dickson, Director of the Public Diplomacy Office in Western Hemisphere Affairs, to encourage the action. After much bureaucratic problem solving, a task performed mainly by FONDECYT/FONDAP Director María Elena Boisier, the equipment finally arrived in Chile in March 2011 and was rapidly distributed to the different institutions. The arrival ceremony was widely covered by local media.

M Allende published two very highly downloaded articles (d'Alençon et al., 2010 and Gallardo et al., 2010) published in BMC journals (download counts four months after publication are provided in the Appendix).

VIII. SUGGESTIONS

What recommendations would you make to the FONDAP Program Office to improve the performance of the Center and the review process? Please describe.

We are very satisfied with the way in which the FONDAP Program Office handles requests, inquiries and communications. We understand that the main difficulties (restriction on the use of funds for the most part) are not the responsibility of this Office but rather are part of the general regulations that are required by the Chilean State for public spending.
We look forward to reading the comments of the peer reviewers to improve our performance in the coming years and to meet the expectations of the program in the best possible way.