**D-4-a Preliminary results related to validation:**

**D-4-a-i Low-frequency functionally active intronic & intergenic inherited**

**variants predisposing to cancer,,,,,we have experience w Tyrol cohort…..**

Emerging,,,,cut…. insights into the genetics of constitutional disease etiology demonstrate that germline polymorphisms are associated with a variety of diseases including Alzheimer’s, Parkinson’s, mental retardation, autism, schizophrenia\cite{19715442}and cancer \cite{19536264,18685109}. Relevant to this proposal our group recently performed a large scale profiling study for 2,000 individuals from the Tyrol Early Prostate Cancer Detection Program \cite{18321314,16829552}cohort. This cohort is part of a population-based prostate cancer-screening program started in 1993 and intended to evaluate the utility of intensive PSA screening in reducing prostate cancer specific death. By genotyping DNA extracted from peripheral blood samples, we annotated the cohort on more than 5,000 CNVs and 900,000 SNPs and then queried inherited low frequency deletions variants \cite{20059347} for their impact in driving prostate cancer \cite{20479773} and the more aggressive form of the disease \cite{10351184}. We reported on coding and non-coding functionally active risk variants. Among the top hits of the case-control study, an intronic variant in the Alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase C (MGAT4C) demonstrated transcript abundance association with genotype states both in prostate and in lymphoblastoid cells, significant increase in cell and migration upon overexpression in benign and cancer prostate cell lines, and significant decrease in proliferation upon knock down of MGAT4C expression with siRNA. In addition, we suggested that intergenic PCA risk variants affect gene regulation through modified transcription factor binding activity of the Activator Protein 1 (AP-1)\cite{20299548,21862627}. Altogether, we demonstrated that inherited variants may directly or indirectly modulate the transcriptome machinery of known oncogenic pathways in prostate cancer facilitating carcinogenesis.,,,,shrink 50 pvctent…..

#### D-4-a-ii In vitro characterization of SNVs within enhancer elements bound by AR and/or ER$α$

,,,,redundant w abo cut by half just focus on expt…..We and others have largely studied the genetics of prostate cancer individuals coupling serum levels and genomics data. Specifically, we characterized the impact of genetic variants relevant to the metabolism of Dihydrotestosterone [131](DHT), the most potent form of androgen, and investigated the incidence of common genomic rearrangements with respect to PSA levels and age at diagnosis [132]. The Tyrol Early Prostate Cancer Detection Program cohort we profiled for those studies is a well-characterized cohort with centralized data collection that ensures proper patients’ follow-up annotations and availability of well-preserved tissues and blood samples. The cohort currently includes more than 3,000 men. Genotype-transcript associations have been reported at large for multiple types of inherited variants [8, 9, 25, 26, 133]; however, experimental evidence of inherited variants allele-specific effect on enhancer activity are lacking. The challenge is to demonstrate how single variants or combinations can increment disease susceptibility or progression by perturbing the expression of a transcript, disrupting the function of a protein or affecting regulatory sequences. In order to study the potential role of inherited genetic variants within regulatory intergenic elements in the context of hormone dependent human tumors, we recently performed an unbiased computational search for AR/ER$α$ bound enhancers elements containing SNVs followed by *in vitro* characterization of two selected variants (Garritano et al, Oncotarget REF PMID: 25693204). ChIP-Seq ENCODE data analyses across 17 cell-lines with H3K4m1 and H3K4m3 data availability was performed including 136 transcription factors. Briefly, we applied the following filters to select regions of interest: i) chromatin signature of enhancer activity (H3K4m1, H3K4me1+H3K4me3), ii) binding by ER and AR and iii) presence of a SNP. Using dbsnp137 and AR [134] and/or ER$α$ (D Chakravarty, *et al REF to include*) binding data a total of 41 loci was identified (see Table 1 of Garritano et al, Oncotarget 2015). Based upon the minor allele frequencies of the SNPs of interest, two were selected for *in vitro* characterization (on 1q21.3, rs2242193 MAF=0.038 and 13q34, rs9521825 MAF=0.235), here referred to as Locus 1 and Locus 2. Selected loci were cloned in pGL4.26 plasmid (plasmid with alternative allele was also generated) and then validated and characterized *in vitro* by luciferase assay with and without DHT treatment in MCF7 cells. Both constructs reached high responsiveness to DHT treatment hinting at their strong enhancer role. Moreover, the SNP variant on 1q21.3, rs2242193, demonstrated a role in the transcription regulation (p=0.028, Student’s t-test) (**Figure XX).** As both loci were confirmed as transcriptionally responsive to DHT by luciferase assay, we next opted for ChIP assays with AR antibody (or with normal IgG as a control) using MCF7 cells that are heterozygous at rs2242193 in Locus 1, but homozygous for the reference allele at Locus 2. Using quantitative PCR (qPCR), we were able to detect AR binding to both selected loci in MCF7 cells transiently over-expressing AR (**Figure YYB**). Occupancy levels at KLK3, KLK2 and TMPRSS2 were measured as positive controls (**Figure YYA**). Moreover, to assess whether AR showed allele-specific DNA binding at rs2242193, we amplified AR-enriched Locus 1 region by standard PCR followed by double-strand direct DNA sequencing analysis. Quantification of the electropherograms showed that the A allele was significantly enriched (p<10-22, Fisher test) in chromatin fragments immunoprecipitated with antibody against AR compared to input genomic DNA (**Figure YYC**). Altogether, our results show that unbiased genome-wide search for polymorphic regulatory regions (PRRs) is an efficient methodology to discover new functional cis-elements relevant to hormone driven diseases and beyond by providing experimental evidence for selected variants mapping to regulatory regions.

,,,,,Conver cites…..

**D-4-a-iii ,,,,we have exper….Modeling mutations in cell lines using CRISPR CAS system:**

We have successfully used the CRISPR CAS system to generate mutations and deletions in genes. We have evidenced the presence of somatic mutation in the MAP3K7 gene in castrate resistant prostate cancer patients. In order to determine the functionality of the mutation we have used the CRISPR CAS system to generate the mutation in cell lines. We have successfully introduced cancer-specific MAP3K7 mutation in VCaP cells using the CRISPR-CAS system. Sequencing of cell lines confirmed mutation (Figure XXA). We studied the genomic influence of MAP3K7 mutation in evolution of castrate resistant prostate cancer. Another example is the deletion of the FANCA gene evidenced in 16% of localized prostate adenocarcinomas (11 of 69 cases) and 14% of advanced prostate cancers (4 of 29 cases). In some patients deletion of FANCA was associated with increased cisplatin sensitivity. We have used the CRISPR CAS system to generate FANCA deletion in prostate cancer cell line 22RV1. .Briefly, the CRISPR/Cas9 plasmid (Px459) was obtained from Addgene (Cambridge,MA). Using Ran et al(15) protocol we identified a FANCA CRISPR DNA target sequence using algorithms based on analysis in Hsu et al(16). The corresponding oligonucleotides were ordered (IDT Coralville, IA) and were cloned into Px459 vector. Sanger sequencing confirmed integration of the FANCA target site into the vector. CRISPR deletion of FANCA in 22 RV1 cells lead to increased cisplatin sensitivity (Figure XX2)

**D-4-a-iv Validation and functional evaluation of physiologic role of somatic**

**mutation predicted by FUN-seq bioinformatics pipeline.** ,,,,,we have done validationf for FunSeq but for somatic….,A mutation in the RET

promoter predicting a gain of an AP1 motif was determined using the in silico FUN-seq pipeline. Using a luciferase based reporter assay we studied the promoter activity of the WT and mutant RET promoter in the DU145 cell line. Luciferase activity confirmed

that the mutant promoter was 1.2-1.3 fold more active than the WT promoter (Fig XX3).

**Approach:**

**,,,,,why mention EDRN here n Tyrol up…..D-4-b-i Targeted genotyping:** We will determine if any or all 6 variants selected based on successful validation in Aim 3 are associated with cancer or cancer causing characteristics. We will achieve this by studying the specific variant in test cohorts. We will use both the Tyrol cohort (described above) and the Early Detection Research Network (EDRN) \cite{0000005} prostate cancer cohort with thousands of prostate cancer individuals as well as normal controls. The prostate cancer cohort includes men enrolled at three sites as part of the Prostate Cancer Clinical Validation Center that prospectively enrolls individuals at risk for prostate cancer at Beth Israel Deaconess Medical Center (Harvard), at the University of Michigan (Michigan) and at Weill Cornell Medical College (Cornell). Cases are defined as men diagnosed with prostate cancer and controls are men who have undergone prostate needle biopsy without any detectable prostate cancer and no prior history of prostate cancer.

We will first take the highest prioritized variants then subject them to validation. Overall we plan to start the validation pipeline with the top ~6 elements identified from the reporter assays (as described above). TaqMan assays for these 6 variants will be performed on 4,000 cases to see if the precise variants recur in a larger cohort. Then, we will follow up for detailed functional screening, to be discussed below. For controls, we will utilize deeply sequenced control cohorts (individuals with no cancer) that are already available, including deeply sequenced trios from the 1000 Genomes Project \cite{0000006}, 500 individuals with Complete Genomics sequencing also from 1000 Genomes \cite{0000007} and non-cancerous individual from the UK10K project \cite{0000008}. ,,,,use same controls as above…..Superior allelic discrimination is achieved in these assays as they utilize TaqMan minor groove-binding (MGB) probes. This technique generates a low signal to noise ratio and affords a greater flexibility. The Taqman probes are functionally tested to first ensure assay amplification and optimization for amplification conditions.

Methods: Genomic DNA will be extracted from the blood cellular-EDTA samples in a high-throughput fashion using the QIAamp 96 DNA Blood Kit (Qiagen). All DNAs will be evaluated by NanoDrop spectrophotometer (NanoDrop, Thermo Scientific) and gel electrophoresis (2% agarose). For TaqMan Real-Time Quantitative PCR, each DNA sample will be diluted to 10 ng/ml with nuclease-free water.

**D-4-b-ii Evaluation of functional consequence of variants**

Based on the reporter assays (described above), we will pick the top ~6 variants for functional follow up.

**D-4-b-ii-(1) Functional consequences**: Real-Time Quantitative PCR

Real-time quantitative PCR analysis of the genes downstream of the 6 selected variants will be performed on individuals that have been identified as recurrent for the variants and a similar sized group of non-recurrent individuals. ,,,,,look for perturbed…..

This analysis will inform us if a SNP (in promoter or enhancer regions) has any effect on transcription of the target gene. PCR will be followed by direct sequencing of the amplicon using an ABI 3730 DNA Sequence Analyzer on a subset of tumor-normal pairs to verify the individual promoter/enhancer mutations for further confirmation.

**D-4-b-ii-(2) Functional consequences: CRISPR/CAS system**

We will utilize the newly discovered CRISPR/CAS system \cite{0000009} to generate

endogenous mutations in target genes,,,,???... in a panel of prostate cancer cell lines (VCaP,

LnCaP, DU145 and PC3). This unique system will provide us an opportunity to

directly modulate endogenous genes and minimize artifacts due to the transfection

based reporter assays. Using CRISPR/CAS mediated genome-engineering method

\cite{23643243} we will directly generate mutations within promoter/enhancers of

target genes. Theoretically we will generate 6 individual SNPs in each cell line and will

study functional relevance of these changes compared to WT. In case of rare

mutations, which occur within both promoter and enhancer regions of the same

gene, we will develop cell lines having these combinatorial mutations.

Mutations within regulatory regions like promoter and enhancer regions might

contribute to one or more biological effects as described in the schematic (Fig. 12).

**D-4-b-ii-(3) Functional consequences:**

The mutant and WT cell lines generated using CRISPR/CAS system will be

monitored for a) phenotypic changes by confocal microscopy and actin staining to

determine effects of mutation on cytoskeletal reorganization b) Influence on

proliferation by MTT and CellTiter-Glo® Luminescent Cell Viability Assay (Promega)

c) Influence on invasive and migratory potential using, matrigel coated invasion and

boyden chambers in 24 well format d) senescence by Bgal staining e) apoptosis by

tunnel assay.

**D-4-b-ii-(4) Effect of the mutation on TF binding**

In vitro EMSAs will confirm specific binding to WT or mutant sequence by a

particular transcription factor. EMSA (electrophoretic mobility shift assay) is a common technique employed to study protein-DNA interaction. We will use the WT and the MT sequences to determine binding of the transcription factor predicted to be present at the site of mutation.,,,,,agree w ls…..

Chromatin immuno-precipitation (ChIP) assays for TFs overlapping the variant will be conducted to determine if the variant can distort TF binding in vivo. This would help validate the variants that are predicted to be motif breakers. Alternatively for the SNVs predicted to create a new motif, ChIP experiments will help validate binding.